John Salerno
Shanchieh Jay Yang
Dana Nau
Sun-Ki Chai (Eds.)

Social Computing, Behavioral-Cultural Modeling and Prediction

4th International Conference, SBP 2011
College Park, MD, USA, March 2011
Proceedings

Springer

Lecture Notes in Computer Science 6589

John Salerno   Shanchieh Jay Yang
Dana Nau   Sun-Ki Chai (Eds.)

Social Computing, Behavioral-Cultural Modeling and Prediction

4th International Conference, SBP 2011
College Park, MD, USA, March 29-31, 2011
Proceedings

Volume Editors

John Salerno
AFRL/RIEF, 525 Brooks Road, Rome, NY 13441, USA
E-mail: john.salerno@rl.af.mil

Shanchieh Jay Yang
Rochester Institute of Technology, Department of Computer Engineering
83 Lomb Memorial Drive, Bldg 09, Rochester, NY 14623-5603, USA
E-mail: sjyeec@rit.edu

Dana Nau
University of Maryland, Department of Computer Science
A. V. Williams Building, College Park, MD 20742, USA
E-mail: nau@cs.umd.edu

Sun-Ki Chai
University of Hawaii, Department of Sociology
2424 Maile Way 247, Honolulu, HI 96822, USA
E-mail: sunki@hawaii.edu

Preface

Welcome to the 2011 international conference on Social Computing, Behavioral–Cultural Modeling and Prediction (SBP11). The overall goal of the conference is to bring together a diverse set of researchers/disciplines to promote interaction and assimilation so as to understand social computing and behavior modeling. In 2008, the first year of SBP, we held a workshop and had 34 papers submitted; in 2010 we had grown to a conference and had 78 submissions. This year, our fourth, we continued to expand, joining forces with the International Conference on Cultural Dynamics (and amending our name accordingly) and we received a record 88 submissions. The submissions were from Asia, Oceania, Europe, and the Americas. We are extremely delighted that the technical program encompassed keynote speeches, invited talks, and high-quality contributions from multiple disciplines. We truly hope that our collaborative, exploratory research can advance the emerging field of social computing. This year we also continued offering four pre-conference tutorials, a cross-disciplinary round table session, a poster board session and a series of panels including a panel featuring program staff from federal agencies discussing potential research opportunities.

The accepted papers cover a wide range of interesting topics: (i) *social network analysis* such as social computation, complex networks, virtual organization, and information diffusion; (ii) *modeling* including cultural modeling, statistical modeling, predictive modeling, cognitive modeling, and validation process; (iii) *machine learning and data mining*, link prediction, Bayesian inference, information extraction, information aggregation, soft information, content analysis, tag recommendation, and Web monitoring; (iv) *social behaviors* such as large-scale agent-based simulations, group interaction and collaboration, interventions, human terrain, altruism, violent intent, and emergent behavior; (v) *public health* such as alcohol abuse, disease networks, pandemic influenza, and extreme events; (vi) *cultural aspects* including cultural consensus, coherence, psycho-cultural situation awareness, cultural patterns and representation, population beliefs, evolutionary economics, biological rationality, perceived trustworthiness, and relative preferences; and (vii) *effects and search*, for example, temporal effects, geospatial effects, coordinated search, and stochastic search. It is interesting that if we were to trace these keywords back to the papers, we would find natural groups of authors of different papers attacking similar problems.

While it may seem at first glance that the topics covered by the papers are too disparate to summarize in any succinct fashion, there are certain patterns that reflect trends in the field of social computing. One is the increasing participation of the human and social sciences in social computing, as well as the active collaboration between such fields and science and engineering fields. Disciplines represented at this conference include computer science, electrical

engineering, psychology, economics, sociology, and public health. A number of interdisciplinary and applied research institutions are also represented.

This conference cannot be run by only a few. We would like to first express our gratitude to all the authors for contributing an extensive range of research topics showcasing many interesting research activities and pressing issues. The regret is ours that due to the space limit, we could not include as many papers as we wished. We thank the Program Committee members for helping review and providing constructive comments and suggestions. Their objective reviews significantly improved the overall quality and content of the papers. We would like to thank our keynote and invited speakers for presenting their unique research and views. We deeply thank the members of the Organizing Committee for helping to run the workshop smoothly; from the call for papers, the website development and update, to proceedings production and registration.

March 2010   John Salerno
                                                          Shanchieh Jay Yang
                                                                   Dana Nau
                                                                Sun-Ki Chai

Organization

Conference Chairs	Dana Nau, Sun-Ki Chai
Program Chairs	John Salerno, Shanchieh Jay Yang
Steering Committee	Huan Liu, John Salerno, Sun-Ki Chai, Patricia Mabry, Dana Nau, VS Subrahmanian
Advisory Committee	Rebecca Goolsby, Terrance Lyons, Patricia L. Mabry, Fahmida Chowdhury, Jeff Johnston
Tutorial Chair	Anna Nagurney
Workshop Chairs	Fahmida N. Chowdhury, Bethany Deeds
Poster Session Chair	Lei Yu
Sponsorship Committee	Huan Liu, L. Patricia Mabry
Student Arrangement Chair	Patrick Roos
Publicity Chair	Inon Zuckerman
Web Master	Mark Y. Goh, Peter Kovac, Kobi Hsu, Rayhan Hasan

Technical Review Committee

Myriam Abramson, Naval Research Laboratory, USA

Donald Adjeroh, Western Virginia University, USA

Agarwal, University of Arkansas at Little Rock, USA

Denise Anthony, Dartmouth College, USA

Joe Antonik, AFRL (RI), USA

Gurdal Arslan, University of Hawaii, USA

Erik Augustson, NIH, USA

Chitta Baral, Arizona State University, USA

Joshua Behr, Old Dominion University, USA

Geoffrey Barbier, AFRL (RH), USA

Herbert Bell, AFRL (RH), USA

Lashon Booker, The MITRE Corporation, USA

Nathan Bos, John Hopkins University – APL, USA

Douglas Boulware, AFRL (RI), USA

Jiangzhou Chen, Virginia Tech, USA

Xueqi Cheng, CAS, P.R. China

Alvin Chin, Nokia Research Center

David Chin, University of Hawaii, USA

Jai Choi, Boeing, USA

Hasan Davulcu, Arizona State University, USA

Brian Dennis, Lockheed Martin, USA

Guozhu Dong, Wright State University, USA

Richard Fedors, AFRL (RI), USA

Laurie Fenstermacher, AFRL (RH), USA

William Ferng, Boeing, USA

Clay Fink, John Hopkins University – APL, USA

Anthony Ford, AFRL (RI), USA

Anna Haake, RIT, USA

Jessica Halpin, Not Available

Walter Hill, Not Available
Michael Hinman, AFRL (RI), USA
Jang Hyun Kim, University of Hawaii
   at Manoa, USA
Terresa Jackson, Not Available
Ruben Juarez, University of Hawaii,
   USA
Byeong-Ho Kang, University of
   Tasmania, Australia
Douglas Kelly, AFRL (RH), USA
Masahiro Kimura, Ryukoku
   University, Japan
Irwin King, Chinese University of
   Hong Kong
Minseok Kwon, RIT, USA
Alexander H. Levis, George Mason
   University, USA
Huan Liu, Arizona State University,
   USA
Lustrek, Jozef Stefan Institute,
   Slovenia
Patricia L. Mabry, National Institutes
   of Health, USA
K. Clay McKell, Not Available
Sai Motoru, Arizona State University,
   USA
Hiroshi Motoda, Osaka University and
   AOARD, Japan
Anna Nagurney, University of Mass
   Amherst, USA
Keisuke Nakao, University of Hawaii
   at Hilo, USA
Kouzou Ohara, Aoyama Gakuin
   University, Japan
Esa Rantanen, RIT, USA
Bonnie Riehl, AFRL (RH), USA

Kazumi Saito, University of Shizuoka,
   Japan
Antonio Sanfilippo, Pacific Northwest
   National Laboratory, USA
Hessam Sarjoughian, Arizona State
   University, USA
Arunabha Sen, Arizona State
   University, USA
Vincient Silenzio, Not Available
Adam Stotz, CUBRC, USA
Gary Strong, Johns Hopkins
   University, USA
George Tadda, AFRL (RI), USA
Lei Tang, Arizona State University,
   USA
Shusaku Tsumoto, Shimane
   University, Japan
Trevor VanMierlo, Not Available
Changzhou Wang, Boeing, USA
Zhijian Wang, Zhejiang University,
   China
Rik Warren, AFRL (RH), USA
Xintao Wu, University of North
   Carolina at Charlotte, USA
Ronald Yager, Iona College, USA
Laurence T. Yang, STFX, Canada
Shanchieh Jay Yang, RIT, USA
Michael Young, AFRL (RH), USA
Lei Yu, Binghamton University, USA
Philip Yu, University of Chicago, USA
Laura Zavala, Not Available
Daniel Zeng, University of Arizona,
   USA
Jianping Zhang, The MITRE
   Corporation, USA
Zhongfei Zhang, Not Available
Inon Zuckerman, UMD

Table of Contents

Using Models to Inform Policy: Insights from Modeling the Complexities of Global Polio Eradication

Kimberly M. Thompson

Kid Risk, Inc., Boston, MA, USA
`kimt@kidrisk.org`

Abstract. Drawing on over 20 years of experience modeling risks in complex systems, this talk will challenge SBP participants to develop models that provide timely and useful answers to critical policy questions when decision makers need them. The talk will include reflections on the opportunities and challenges associated with developing integrated models for complex problems and communicating their results effectively. Dr. Thompson will focus the talk largely on collaborative modeling related to global polio eradication and the application of system dynamics tools. After successful global eradication of wild polioviruses, live polioviruses will still present risks that could potentially lead to paralytic polio cases. This talk will present the insights of efforts to use integrated dynamic, probabilistic risk, decision, and economic models to address critical policy questions related to managing global polio risks. Using a dynamic disease transmission model combined with probabilistic model inputs that characterize uncertainty for a stratified world to account for variability, we find that global health leaders will face some difficult choices, but that they can take actions that will manage the risks effectively. The talk will emphasize the need for true collaboration between modelers and subject matter experts, and the importance of working with decision makers as partners to ensure the development of useful models that actually get used.

Toward Culturally Informed Option Awareness
for Influence Operations with S-CAT

Ken Murray[1], John Lowrance[1], Ken Sharpe[1], Doug Williams[2],
Keith Grembam[2], Kim Holloman[3], Clarke Speed[4], and Robert Tynes[5]

[1] SRI International, Menlo Park, CA, 94025 USA
{kenneth.murray,john.lowrance,kenneth.sharpe}@sri.com
[2] SET Corporation, Greenwood Village, CO, 80111 USA
{dwilliams,kgrembam}@setcorp.com
[3] SAIC, Sterling, VA, 20164 USA
hollomank@saic.com
[4] University of Washington, Seattle, WA 98195, USA
landogo@u.washington.edu
[5] SUNY at Albany, Albany New York 12222, USA
rt873361@albany.edu

Abstract. The Socio-Cultural Analysis Tool (S-CAT) is being developed to help decision makers better understand the plausible effects of actions taken in situations where the impact of culture is both significant and subtle. We describe S-CAT in the context of a hypothetical influence operation that serves as an illustrative use case. One of the many challenges in developing S-CAT involves providing transparency into the model. S-CAT does this by providing explanations of the analysis it provides. This paper describes how S-CAT can improve option-awareness during influence operations and discusses the explanation capabilities used by S-CAT to support transparency into the model.

Keywords: Modeling Culture, Effects-Based Modeling, Forecasting Effects.

1   Introduction

The United States and other countries are engaged in Influence Operations (IOs) in very diverse parts of the world. Successful outcomes for such operations require understanding the societies and cultures where they are conducted [1]. Misunderstandings due to cultural differences can yield unforeseen and potentially adverse consequences, delaying progress in the best cases and producing instability and tragedy in the worst cases [2, 3].

To improve effectiveness, IOs must be conducted with improved understanding of the human socio-cultural behavior (HSCB) of the host region in order to pursue plans that avoid otherwise unforeseen adverse consequences and exploit favorable opportunities. Option awareness is improved when selection among candidate plans is informed by an understanding of how they plausibly impact the attitudes, activities and circumstances of the host region. However, not every IO planning analyst can become a social scientist specializing in the host region.

To address this challenge, technology can gather, organize, and represent HSCB knowledge from diverse sources, including social scientists and regional experts, to capture culturally informed action-effect relationships that forecast the plausible consequences of candidate plans [3, 4]. We are developing the Socio-Cultural Analysis Tool (S-CAT) to help analysts and decision makers better understand the plausible effects of actions taken in situations where the impact of culture is both significant and subtle. Although its application is potentially much broader, S-CAT is specifically being developed to provide greater option awareness to planning analysts and decision makers in conducting IOs.

2   Example Scenario: Reducing Piracy Off the West-African Coast

The use of S-CAT is illustrated with the hypothetical IO scenario[1] in Fig. 1.

Following the international crackdown on piracy based in Sudan, Atlantic shipping operations report increasing piracy off the coast of Sierra Leone. Most raiding vessels seem to come from the Bonthe District. The IO coalition must decide how best to curtail this emerging piracy on the Bonthe coast. Four candidate plans are:

(1) Provide monetary aid to the Bonthe District Government to discourage piracy in the district.
(2) Provide monetary aid to the local tribal leader, Paramount Chief Sesay, to discourage piracy in the jurisdiction of his chiefdom.
(3) Provide equipment and supplies (e.g., fish-finding sonar, rice irrigation systems and fertilizer) to the fishing and rice-farming enterprises run by Pa Yangi-Yangi, a local entrepreneur and informal leader.
(4) Provide security to Pa Yangi-Yangi, because he appears to be someone the coalition can work with to reduce piracy.

Fig. 1. Hypothetical scenario for reducing West-African piracy

IOs often require selecting leaders or organizations from the host region and trying to leverage their local influence to achieve a goal. In the example, candidates include the district government, the local tribal leader, and a local entrepreneur and informal leader. Candidate plans are prepared for each. S-CAT contributes to option awareness during plan selection by:

- Forecasting the plausible effects of each alternative plan
- Comparing the forecasted effects of alternative plans with stated desired effects
- Revealing the cultural factors and lines of reasoning supporting forecasted effects

The following sections describe how S-CAT uses several types of knowledge to assess the candidate plans in Fig. 1 and how S-CAT is being developed to provide transparency into its model by explaining the results of its analysis.

---

[1] The people and events described are not real. The candidate plans are for purposes of illustration only and do not represent the real or intended actions of any government.

3   Improving Option Awareness by Assessing Alternative Plans

S-CAT supports two distinct users: a modeling analyst uses S-CAT to construct culturally informed models used to forecast the plausible effects of plans and a planning analyst uses S-CAT to construct and assess candidate plans.

3.1   Modeling in S-CAT

The modeling analyst requires access to cultural expertise but is otherwise lightly trained: S-CAT is being developed explicitly to enable use by IO analysts so that extensive training in knowledge representation and other esoteric modeling specialties is not required. The modeling analyst constructs two types of model components: the site model and the effects rules.

Sites. The site model captures elements of the area of interest that are reasoned about; this typically includes:

- Agents, including individual people, groups (e.g., tribes, populations) and organizations (e.g., governments, NGOs)
- Locations, including geopolitical regions (e.g., nations, districts, villages) and geographical regions (e.g., islands, rivers, plains)
- Activities, including economic enterprises (e.g., fishing, farming)

There is no restriction on the echelon or level of detail captured in the site model: international regions and nation states can coexist with individuals and villages. Each element is specified as instantiating one or more types (e.g., *Leader*, *Village*) selected from an existing, extensible hierarchy of element types. The site model includes relations over the elements (e.g., relation-type *resident-of* captures that a person resides in a region). The site also includes profiles that describe the cultural properties of elements. For this example, S-CAT uses two culture profiles, Affluence and Speed ([4] describes the capture and use of cultural profiles and their variables in S-CAT). Fig. 2 presents a portion of the site model developed for the example.

Site Element	Types	Outgoing Relations	Affluence	Speed
Al Hajii Ghenghe Sesay	Paramount Chief	has regional subordinate -- *Section Chief 1* has regional subordinate -- *Section Chief 2* has regional subordinate -- *Section Chief 3* resident of -- *Imperri Chiefdom*		
Kita	Village	has location -- *Imperri Chiefdom* has location -- *McCauley Island*		
Kita Fish Brothers	Secret Male Group	has leader -- *Jonathan (Pa Yangi-Yangi) Tucker*		
Kita Fishing	Fishing Industry Informal Commerce Small Enterprise	has performing group -- *Kita Fish Brothers*		

Fig. 2. Part of the site model for reducing West-African piracy

Effects rules. S-CAT includes two kinds of effects rules: direct and indirect. Direct-effects rules capture effects that plausibly occur directly from the execution of some type of action on some type of site element. Indirect-effects rules capture the effects that plausibly occur as a consequence of some other effect; they reveal how an impact on one element of the site model can propagate to other related elements.

The antecedent of a direct-effects rule is simply the action type, the element type, and optional cultural constraints on the site elements to which the rule applies. The antecedent of an indirect-effects rule includes typed "variable nodes" with relations and constraints on the values of effects and cultural variables. The consequences of all effects rules are plausible values for effects variables. This example uses variables capturing the political, military, economic, social, infrastructure, and information (PMESII) effects ([4] describes the use of PMESII effects variables in S-CAT, and [5] describes the *structured-argumentation* techniques supporting their capture). Fig. 3 presents a direct-effects rule and shows how the modeling analyst specifies a consequence PMESII effect. Fig. 4 presents an indirect-effects rule and its rationale.
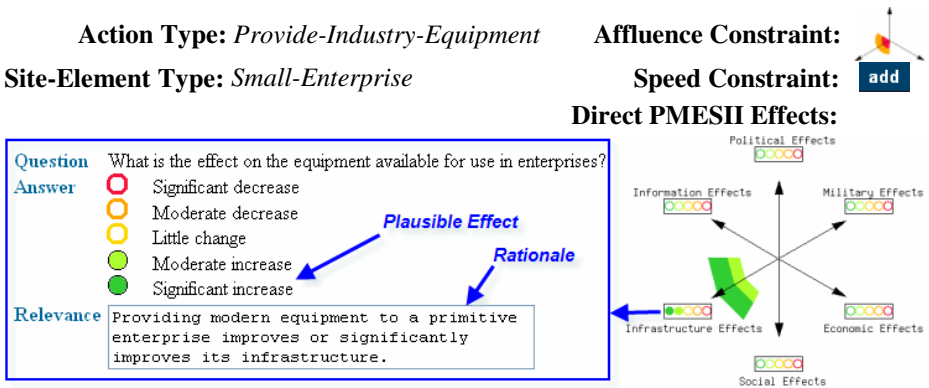
Fig. 3. Direct-Effects Rule: *Provide Industry Equipment to a Small Enterprise*

Rationale: *When an economic enterprise produces a surplus and its performers embrace allocative leadership, their leader controls that surplus.*

Fig. 4. Indirect-Effects Rule: *Allocative Leader Controls Enterprise Surplus*

3.2   Assessing Plans in S-CAT

After the modeling analyst has created the site model and a set of site-appropriate (culturally informed) effects rules, the planning analyst uses the model to define and assess candidate plans. This involves: creating a goal to capture the desired effects, creating plans to achieve the goal, creating forecasts of plausible effects for each plan, and then evaluating the plans' forecasted effects against the goal's target effects.

Goals. A goal is defined by selecting one or more elements from the site model, and for each creating a target PMESII profile that identifies the desired PMESII effects for that site element. Each target effect can be annotated with a weight to capture priorities. In the example, the two site elements selected are *Kita Piracy* and *Bonthe District Population*. The primary (and fully weighted) target effects include reducing the activity of piracy and reducing its profitability. The secondary target effects include reducing support for piracy among informal leaders and increasing the social stability of the population. Fig. 5 presents the goal defined for the example.

Intent	Site Element	Variable	Question	Target Effects	Weight
Primary Objective: reduce activity of piracy	*Kita Piracy*	PMESII > Economic Effects > INFORMAL ECONOMY	What is the effect on the strength of the informal economy?	Significant decrease	✪
Primary Objective: reduce economic gain from piracy	*Kita Piracy*	PMESII > Economic Effects > WEALTH	What is the effect on available wealth?	Significant decrease	✪
Secondary Objective: reduce informal leader support for piracy	*Kita Piracy*	PMESII > Social Effects > INFORMAL APPROVAL	What is the effect on approval by informal leaders?	Significant decrease	✪
Secondary Objective: increase social stability of the populace	*Bonthe District Populace*	PMESII > Social Effects > SOCIAL STABILITY	What is the effects on social stability?	Significant increase	✪

Fig. 5. A counter-piracy goal

Plans. A plan contains one or more actions. Each action is defined as the application of an action type, selected from an existing, extensible taxonomy of diplomatic, information, military, and economic (DIME) action types, to an element selected from the site model. In the example, there are four distinct plans created; the third plan from Fig. 1 is presented in Fig. 6. This plan aspires to leverage the influence of a local entrepreneur, Pa Yangi-Yangi. It includes providing industrial equipment (e.g., fish-finding sonar) to *Kita Fishing* and providing both industrial equipment (e.g., irrigation equipment, rice-milling machines) and industrial supplies (e.g., fertilizer) to *Kita Rice Farming*; both operations are controlled by Pa Yangi-Yangi.

Action	Target	Action Types
Provide Fish-Locating Sonars	Kita Fishing	Provide Industry Equipment
Provide Rice Milling and Irrigation Equipment	Kita Rice Farming	Provide Industry Equipment
Provide Fertilizer	Kita Rice Farming	Provide Industry Supplies

Fig. 6. A counter-piracy plan: *Invest in Kita Food Production*

Assessing Alternative Plans. For each candidate plan, the planning analyst uses S-CAT to create a forecast of plausible effects from executing the plan. This involves first running all the applicable direct-effects rules and then, iteratively, running all the applicable indirect-effects rules. Each forecast creates a distinct inferential context for the effects it establishes for the site elements and enables S-CAT to maintain multiple, independent forecasts (e.g., one for each candidate plan), where the effects in one forecast do not impact the effects in another forecast.

The applicable direct-effects initiate the forecast. A direct-effects rule is applicable for a plan if the rule's action-type matches or generalizes the action-type specified in one of the plan's actions, and the site element selected for that plan action instantiates the site-element type of the rule, and that site element satisfies the cultural constraints, if any, defined for the rule. In the example, the direct-effects rule in Fig. 4 applies to the first two actions of the plan in Fig. 6 (the Affluence-profile constraint in Fig. 4 requires the enterprise to use primitive tools and is satisfied by both *Kita Fishing* and *Kita Rice Farming*).

The applicable indirect-effect rules extend the forecast by iteratively propagating effects established by the forecast (e.g., by previously applied effects rules) to other related site elements. An indirect-effects rule is applicable when each of its variables bind to some site element such that the types for the variable are instantiated by the site element, the relations defined for the variables are satisfied by the site-element bindings, and each site element satisfies any cultural profile constraints defined for the variable as well as any effects constraints defined for the variable. Note that the effects constraints are tested using the forecast, whereas the other conditions are tested using the site model.

After forecasts are generated for the alternative plans, they are evaluated against the goal. S-CAT displays a table showing how well the effects of each plan's forecast satisfy the target effects of the goal (see Fig. 7). According to the model, investing in the local Paramount Chief or the district government has no impact on piracy. Among these four plans, only investing in the local food-production enterprises leads to a reduction in piracy. Furthermore, a quite surprising outcome is that providing Pa Yangi-Yangi protection might actually reduce social stability (because of the unforeseen consequence of putting him at greater risk from rival, militant leaders).

Label	Weight	Target Effects	Invest in Kita Food Production - Outcomes	Invest in Pmt. Chief Sesay - Outcomes	Invest in Bonthe Dist. Govt. - Outcomes	Protect Pa Yangi-Yangi - Outcomes
Kita Piracy INFORMAL ECONOMY		Significant decrease	Moderate Decrease Significant Decrease delta: 20	No effect forecast delta: 40	No effect forecast delta: 40	No effect forecast delta: 40
Kita Piracy WEALTH		Significant decrease	Moderate Decrease Significant Decrease delta: 20	No effect forecast delta: 40	No effect forecast delta: 40	No effect forecast delta: 40
Kita Piracy INFORMAL APPROVAL		Significant decrease	Significant Decrease delta: 0	No effect forecast delta: 40	No effect forecast delta: 40	No effect forecast delta: 40
Bonthe District Populace SOCIAL STABILITY		Significant increase	Significant Increase Moderate Increase delta: 20	No effect forecast delta: 40	Moderate Increase Little Change Moderate Decrease delta: 60	Significant Decrease delta: 80
Aggregate			delta: 16	delta: 40	delta: 44	delta: 49

Goal Satisfaction Scale Satisfied ▮▮▮▮ Unsatisfied

Fig. 7. Comparing forecasted effects among alternative candidate plans

4   Improving Option Awareness by Explaining Forecasted Effects

S-CAT is not intended to simply identify the best among a set of candidate plans. Rather, it forecasts specific effects and explains the forecasted effects in order to reveal the cultural factors and lines of reasoning that support each forecasted effect. This provides essential transparency into the HSCB model and distinguishes S-CAT from other HSCB tools using traditional agent-based simulation capabilities [6].

Identifying the cultural factors relevant to forecasted effects increases the understanding of the IO planning analysts and decision makers about how candidate plans impact and are impacted by the host region culture, thereby improving their option awareness. By revealing precisely those cultural factors in the model most relevant to a candidate plan as that plan is being evaluated for selection, S-CAT provides *just in time* cultural training to the IO planning analysts and decision makers.

Identifying the lines of reasoning used in a forecast has two primary benefits. It increases the understanding of the IO planning analysts and decision makers about why the forecasted effects may plausibly occur, thereby improving their option awareness. Furthermore, it increases their understanding of what is and what is not captured in the model and considered by S-CAT when computing forecasts, thereby improving their awareness of the model's scope and limitations.

Fig. 8 presents samples of the cultural factors and lines of reasoning revealed by S-CAT for the reduction in *Kita Piracy* forecasted for the plan of Fig. 6.

Fig. 8. Explaining the forecasted reduction in *Kita Piracy*

5   Summary and Discussion

S-CAT is a decision-support tool being developed to improve option awareness by providing culturally informed assessments of candidate plans. Each assessment forecasts plausible effects of a plan. Explanation capabilities reveal both the cultural factors and the lines of reasoning supporting each forecasted effect, providing transparency into the model and the automated reasoning used to assess the plans. Revealing the cultural factors enables users to better understand how the local culture impacts their candidate plans. Revealing the lines of reasoning enables users to better understand how the forecasted effects of their candidate plans are plausible. By identifying and explaining both the plausible effects and the underlying cultural factors relevant to candidate plans, S-CAT improves the option awareness of its users.

S-CAT does not obviate the need for cultural expertise about the region of interest. Rather, it allows lightly trained modeling analysts to capture cultural expertise provided from available sources (e.g., social scientists, regional experts) in culturally informed models. Once created, the models can be subsequently exploited anywhere and at anytime in support of IO planning. Thus, capturing cultural expertise within S-CAT enables the access to available sources of cultural expertise (during modeling) to be asynchronous and offsite from the application of that expertise (during planning).

References

1. Ng, K.Y., Ramaya, R., Teo, T.M.S., Wong, S.F.: Cultural Intelligence: Its Potential for Military Leadership Development. In: Proc. of IMTA 2005, Singapore, pp. 154–167 (2005)
2. Hearing: Army Transformation Hearing before the Committee on Armed Services, United States House of Representatives, Second Session, Washington, D.C., July 15-24 (2004)
3. Sharpe, K., Gremban, K., Holloman, K.: Evaluating the Impact of Culture on Planning and Executing Multinational Joint Force Stability, Security, Transition and Reconstruction Operations. In: Proceedings of KSCO 2007, Waltham, MA, pp. 76–81 (2007)
4. Murray, K., Lowrance, J., Sharpe, K., Williams, D., Grembam, K., Holloman, K., Speed, C.: Capturing Culture and Effects Variables Using Structured Argumentation. In: Schmorrow, D., Nicholson, D. (eds.) Advances in Cross-Cultural Decision Making, pp. 363–373. CRC Press, Boca Raton (2010)
5. Lowrance, J., Harrison, I., Rodriguez, A., Yeh, E., Boyce, T., Murdock, J., Murray, K.: Template-Based Structured Argumentation. In: Okada, A., Buckingham Shum, S.J., Sherborne, T. (eds.) Knowledge Cartography: Software Tools and Mapping Techniques, pp. 307–333. Springer, London (2008)
6. Russell, A., Clark, M., La Valley, R., Hardy, W., Zartman, I.W.: Evaluating Human, Social, Cultural, and Behavior (HSCB) Models for Operational Use. In: Schmorrow, D., Nicholson, D. (eds.) Advances in Cross-Cultural Decision Making, pp. 374–384. CRC Press, Boca Raton (2010)

Optimization-Based Influencing of Village Social Networks in a Counterinsurgency

Benjamin W.K. Hung[1], Stephan E. Kolitz[2], and Asuman Ozdaglar[3]

[1] United States Military Academy, Department of Mathematical Sciences,
West Point, New York 10996
Benjamin.Hung@usma.edu
[2] Charles Stark Draper Laboratory, Cambridge, Massachusetts 02139
Kolitz@draper.com
[3] Massachusetts Institute of Technology, Laboratory for Information and Decision Systems,
Cambridge, Massachusetts 02139
Asuman@mit.edu

Abstract. This paper considers the nonlethal targeting assignment problem in the counterinsurgency in Afghanistan, the problem of deciding on the people whom US forces should engage through outreach, negotiations, meetings, and other interactions in order to ultimately win the support of the population in their area of operations. We propose two models: 1) the Afghan COIN social influence model, to represent how attitudes of local leaders are affected by repeated interactions with other local leaders, insurgents, and counterinsurgents, and 2) the nonlethal targeting model, a nonlinear programming (NLP) optimization formulation that identifies a strategy for assigning k US agents to produce the greatest arithmetic mean of the expected long-term attitude of the population. We demonstrate in an experiment the merits of the optimization model in nonlethal targeting, which performs significantly better than both doctrine-based and random methods of assignment in a large network.

Keywords: social network, agent modeling, network optimization, counterinsurgency.

1 Problem Statement

Insurgent movements rely on the broad support of the local population in seeking to overthrow weak governments. The US military finds itself conducting counterinsurgency (COIN) operations on the side of these weak host nation governments. Thus, a counterinsurgency ultimately succeeds by gaining the support of the population, not by killing or capturing insurgents alone [1]. A key component of these counterinsurgency operations is deciding which among a broad class of non-lethal influencing activities ought to be undertaken to win the support of the local population. These activities range from engaging local leaders in constructive dialogue to addressing root causes of poor governance by providing aid and reconstruction assistance for the population.

   In the counterinsurgency operations in both Afghanistan and Iraq, units of battalions and companies are assigned vast amounts of territory and often charged

with winning the support of tens of thousands of people. These units are resource-constrained in terms of personnel, money, equipment, and time. These constraints force units to be very selective in targeting the number, type, frequency, and objective of the operations they conduct. Furthermore, because success or failure is contingent on human behavior, it is extremely difficult to predict how a given action will affect a group of people. The current methods of prioritizing and determining target value are qualitative and are based upon the commanders' and staffs' intuition.

Our objective is to develop a quantitative approach to address how units can determine the best non-lethal targeting assignment strategies and how the sentiments of the population might change as a result of them.

2   Technical Approach

The technical approach was developed to support a variety of uses, including: 1) providing decision support to commanders and their staffs when they make decisions on whom to target non-lethally in order to maximize popular support while operating within the unit's doctrinal and resource constraints, and 2) developing effective COIN Tactics, Techniques and Procedures (TTPs) for non-lethal engagement. We employ the following technical methods: social network analysis, agent modeling, and network optimization.

Social network analysis is an interdisciplinary field that models interrelated individuals or groups of people as nodes in a network [2]. For this application, we model the actors ("agents") in an Afghan counterinsurgency (e.g., local leaders, Taliban and US forces) as nodes and the relationships among them as links in a social network. *Agent modeling* provides analytic insight into the emergent behavior of a collection of agents (individuals or groups of individuals) based on a set of rules that characterize pair-wise interactions. In particular, each agent in the network has a scalar-valued attitude of favorability towards US forces, and the rules model probabilities of influencing each of its neighbors' attitudes. *Network optimization* here refers to formulating and solving the problem of assignment of activities by US forces in order to maximize the positive attitude of local leaders.

Our work is related to a large opinion dynamics literature that presents explanatory models of how individual opinions may change over time. See Abelson [3], DeGroot [4], Friedkin and Johnsen [5], Deffuant [6] [7], and Hegselmann and Krause [8]. This work is most closely related to and builds upon the work of Acemoglu, et al. which characterized beliefs in a social network in the midst of influential agents [9], and "stubborn" agents (i.e., agents which do not change belief) [10]. This work is also related to key person problem (KPP) literature, which uses quantitative methods to identify the *k*-best agents of diffusion of binary behaviors in a network. For example, see Borgatti [11] and Kempe, Kleinburg, and Tardos [12].

From previous work, we make three main contributions. First we apply existing opinion dynamics research to the context of the counterinsurgency in Afghanistan by introducing a methodology for determining approximations of interpersonal influences over Afghan social interaction networks based upon social science and anthropological research. We also provide a method to determine a quantitative value of nonlethal US targeting assignments and their predicted effect over a continuous scale on the local population in the counterinsurgency. Lastly, we present an optimization-based method of nonlethal target assignment selection.

The starting point for our Afghan COIN social influence model is a flexible and extensible Markovian model developed by Acemoglu, et al. [9]. The model represents the agents as nodes in the network graph. Values at each node represent the attitudes of the respective agent and the edges of the graph represent connections among the agents over which influence occurs. The level of influence an agent (node) exerts on other agents (nodes) is modeled by the degree of forcefulness of each, as described below.

We model an agent's attitude towards the counterinsurgents as a continuous random variable that takes on a scalar value at each interaction occurrence (over all the agents). Specifically, let $X_i(k)$ and $X_j(k)$ represent the value of node i and j, respectively, at interaction k, where both $X_i(k)$ and $X_j(k)$ are between -0.5 and +0.5. A negative (or positive) value means low (or high) favorability towards the counterinsurgents, and zero means neutral. This spectrum of attitudes is depicted in Figure 1. Extreme points along the scale denote a greater strength of attitude [7]. In our model, the ideologically motivated agents, the US counterinsurgents and the Taliban insurgents, are "immutable" agents because they possess immutable attitudes which remain at the extreme points and do not change over time.
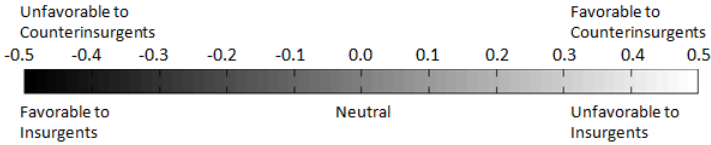
Fig. 1. The numeric and gray scale of the population attitudes

Figure 2 illustrates a simple example network of three villages as well as a US forceful agent and Taliban forceful agent. The shade of a node represents the attitude value using the scale above.
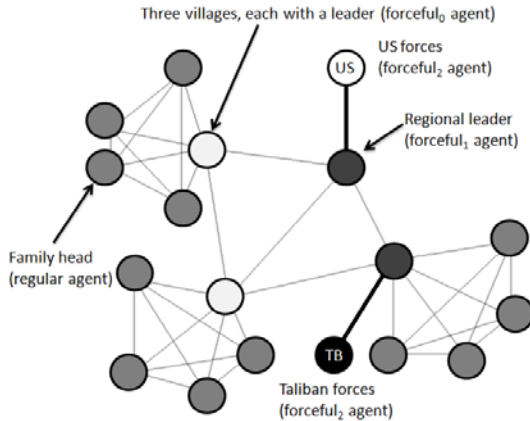
Fig. 2. Illustrative network model

The state of the system is the set of values (attitudes) for the nodes in the network. Changes in the state of the system are modeled using a Markov chain. Transitions occur when there is an interaction between two nodes. Interactions occur when two nodes are chosen sequentially (to model who approaches whom) out of an agent selection probability distribution across nodes. This meeting probability distribution could be uniform, (i.e., each agent is equally likely to be selected), or non-uniform, (i.e., certain agents are more likely to instigate interactions). The first node selected is denoted as i and the second node is denoted as j. The values for nodes i and j can change in response to the interaction between i and j. The transition from an interaction depends upon the forcefulness (described below) of the agents. In the interaction between two agents with values $X_i(k)$ and $X_j(k)$, the possible event outcomes are:

1. With probability β_{ij}, agents i and j reach a pair-wise consensus of attitudes:

$$X_i(k+1) = X_j(k+1) = \frac{X_i(k)+X_j(k)}{2}. \tag{1}$$

2. With probability α_{ij}, agent j exerts influence on agent i, where the parameter ε_{ij} measures the tendency for agent i to resist change to his/her attitude in response to the interaction with agent j:

$$\begin{cases} X_i(k+1) = \varepsilon_{ij} \cdot X_i(k) + (1 - \varepsilon_{ij}) \cdot X_j(k) \\ X_j(k+1) = X_j(k) \end{cases}. \tag{2}$$

3. With probability γ_{ij}, agents i and j do not agree and each retains their original attitude:

$$\begin{cases} X_i(k+1) = X_i(k) \\ X_j(k+1) = X_j(k) \end{cases}. \tag{3}$$

Note that $\beta_{ij} + \alpha_{ij} + \gamma_{ij} = 1$.

These interaction-type probabilities are parameters in the model and can depend on the relative forcefulness of the two agents. Forcefulness can take on one of four values[1]: regular (lowest level of forcefulness), forceful$_0$ (third highest level of forcefulness), forceful$_1$ (second highest level of forcefulness), and forceful$_2$ (highest level of forcefulness). In an exchange between two agents, the agent most likely to change in value is the less forceful of the two and the amount of change depends on the parameter ε_{ij}, which can be thought of as a measure of consistency of attitude.

The transitions can be symmetric, i.e., the order in which the agents interact (who approaches whom) does not matter, or can be different dependent upon the order of the interaction. The most forceful agents (forceful$_2$) always exert an influence on less forceful agents, although the strength of the influence is parameterized. At the other extreme, regular agents are always influenced by forceful agents. For cases in between, multiple outcomes are possible. Varying the parameters and probabilities provide significant modeling flexibility. See [13] for more details.

---

[1] Clearly, the model could have any number of forceful types, including 1 (as in the Acemoglu et al. paper [9]), or 3, as described here. We chose 3 based on a social science SME's initial guidance.

3   Optimization Formulation

While the pair-wise interactions between two agents in the social influence model are simple, the resultant behavior of the system with many agents in a large network is quite complex. However, we devised a method to compute the *expected long-term* attitudes for each agent analytically. See Section 3.3.4 in [13]. We found that:

$$A \cdot \mu_Y + B \cdot \mu_Z = 0 \qquad (4)$$

where μ_Y is the vector of expected long-term attitudes of agents $\in \mathcal{A}$ (mutable agents) at the interaction as $k \to \infty$, μ_Z is the vector of expected long-term attitudes of agents $\in S$ (immutable agents) at the interaction as $k \to \infty$, and the sub-matrices A and B are parameterized by data of meeting and interaction-type probabilities. The expected long-term attitudes of the mutable agents (immutable agent attitudes do not change) is

$$\mu_Y = A^{-1}(-B \cdot \mu_Z). \qquad (5)$$

This result is based upon a particular topology and parameterization. A natural question that follows is what topology produces the attitudes most favorable to the counterinsurgents? More specifically, how should the US agents form connections to other agents in the network that maximizes the favorable attitudes of the population?

   In order to answer this question we formulate the nonlethal targeting problem as a nonlinear program (NLP). Drawing from the general methodology of the classical static weapon-target assignment (WTA) problem, we seek to find the assignment of a fixed number of US agents to a fixed number of local leaders in a social network that maximizes the expected long-term attitudes of the population in favor of the US forces. While the complete derivation for this formulation is in Section 3.5 in [13], we explain the key components of the optimization formulation below.

   Decision variable. Given a fixed number of US agents with (possibly) different interaction-type probabilities, the decision one makes is which US agents are assigned to which non-US agents (local leaders or Taliban) in the network to connect with. The decision variable is therefore a binary matrix, where each entry is:

$$u_{ij} = \begin{cases} 1, & \text{if US Agent } i \text{ connects to agent } j \\ 0, & \text{otherwise} \end{cases}. \qquad (6)$$

Each US agent i can form a link with either 1) the mutable local leaders, or 2) the immutable Taliban leaders. A link formed between a US agent to any mutable agent (and the subsequent propagation of influence from the US agent to that agent) can be representative of any activity or communication in which the targeted local leader is frequently reinforced with pro-counterinsurgent attitudes. See Section 2.5.1 in [13]. A link formed between a US agent and any immutable Taliban agent alters the *meeting probabilities* with which the Taliban agent negatively influences others.

   Objective function. We maximize the weighted average of the expected long-term attitudes for all mutable agents in the network, where the unit commander can subjectively determine a weight (value) for each local leader's attitude based on a variety of factors including 1) the tactical importance of the village where the local leaders are from, and 2) political factors that may demand that one portion of the population be aligned earlier or in lieu of others.

Constraints. The constraints are provided by a reformulation of (4) above in terms of the decision variables, and a limit on the number of allowable connections for the i-th US agent. This limitation may be based off the leader's assessment of his or her ability (with limited resources) to reach a certain number of local leaders.

While this optimization formulation follows from the analytic expression for the expected long-term attitudes of the population (5), some of its properties make it difficult to solve exactly. Specifically, the formulation is both nonlinear as well as nonconvex, which requires heuristic methods to solve and often arrives only at local optima. Additionally, there are $O(n^2)$ variables and $O(n^2)$ constraints, where n is the total number of agents in the network, which means the problem is very large. In Section 3.5.6 of [13], we simplify the formulation which reduced the number of variables and constraints to $O(n)$.

4 Results

The model has been instantiated in a Monte Carlo Simulation. The starting values $\{X_i(0),$ for all $i\}$ are generated, based on *a priori* information or values chosen for analysis purposes. Then a pair of agents is selected using the agent selection probability distribution described previously, and the attitudes are updated according to the transition probabilities described above. The full process can be repeated many times with different starting pairs of nodes and realized transitions. Statistics on the node values can be calculated from results from multiple runs of the model. Example output is illustrated in Section 4.5.2 in [13].
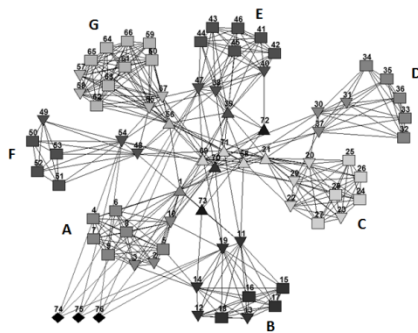
Fig. 3. Experiment network (with village labels)

The input data for the analysis presented here is based on social science aspects of Pashtun society and publicly available aggregate data on Pashtun districts (see [13] for details). A network of 73 local leaders is analyzed, with local leaders being those individuals within the population who by virtue of their authority, power, or position have influence over the attitudes of a group of people. Taliban and US agents are exogenous and exert influence on the network. The 76-node network shown in Figure 3 includes 3 Taliban agents, each with the ability to have connections to 4 local leaders. The network topology is similar that of the "islands" economic model of network formation developed by Jackson [2].

Results from *optimization-based* selection of the best agents to influence by US forces are compared both analytically as well as in simulation with two alternative methods. The first alternative method is to select connections randomly, choosing only from forceful$_0$ and forceful$_1$ agents. The second method was to select connections based on doctrine found in the US counterinsurgency field manuals including the following principles:

- "Identify leaders who influence the people at the local, regional, and national levels ([1]: 5-9).
- "Win over "passive or neutral people" ([1]: 5-22)
- "[Nonlethal targets include] people like community leaders and those insurgents who should be engaged through outreach, negotiation, meetings, and other interaction" ([1]: 5-30).
- "Start easy… don't go straight for the main insurgent stronghold …or focus efforts on villages that support the insurgents. Instead, start from secure areas and work gradually outwards. Do this by extending your influence through the locals' own networks" ([1]: A-5).

Each of the selection methods produced modified network topologies (between US agents and their targeting assignments). Compared to both the random- and doctrine-based selection methods, the optimization-based method produced higher mean expected long-term attitudes both analytically and in simulation in all cases. The full set of experimental results is described in Section 4.5.2 in [13]. The analytical evaluation included a test of the statistical significance of the performance of the optimization-based selection method over the random or doctrine-based selection methods. We tested the null hypothesis: the arithmetic mean of the expected long-term attitudes produced from optimization-based selection and random selection across the cases were drawn from identical continuous distributions with equal medians. We then conducted a pair-wise Wilcox Rank Sum (non-parametric) test on the arithmetic mean of the expected attitudes obtained. The test resulted in a p-value of 0.0041, and thus provided statistically significant evidence that the null hypothesis should be rejected.
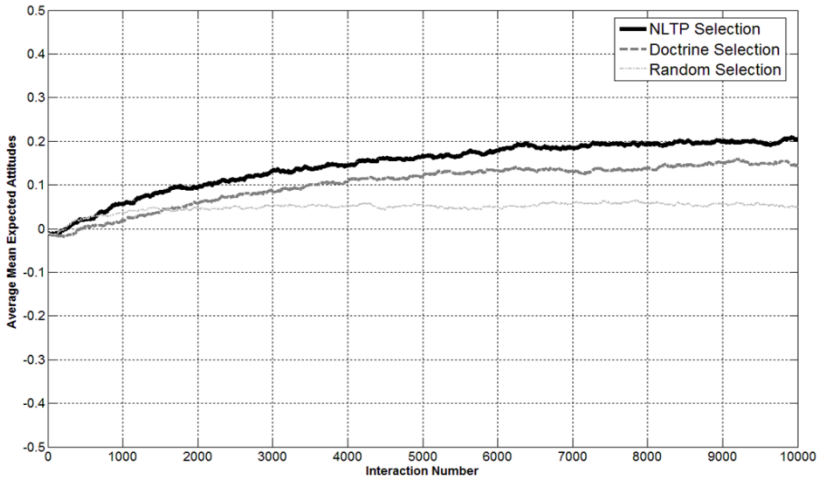
Fig. 4. Averaged Mean Expected Long-Term Attitudes

Simulations on the resulting network topologies produced plots of the arithmetic mean of the expected long-term attitude at each interaction, k, for $k = 1$ to 10000 as illustrated in Figure 4. Specifically, we conducted 50 realizations for each selection method within each case. Each realization produced an arithmetic mean of the expected long-term attitude (over all agents in the network) for each interaction, k. We then averaged over those realizations to produce the arithmetic mean (over all realizations) of the arithmetic mean (over all agents) of the expected long-term attitudes for each interaction, k. For simplicity, we will refer to these calculations as the *averaged mean expected long-term attitudes*. NLTP is the notation for the optimization-based method.

We observed that US agents were often assigned to the friendly forceful$_1$ local leaders who had a generally positive (pro-counterinsurgent) effect on attitudes of all village-level leaders connected to them. However, it is also interesting to note that US agents were also occasionally assigned to forceful$_0$ agents in lieu other forceful$_1$ agents (those with higher relative influence). This was likely due to two apparent principles: 1) the need to overcome Taliban influence in certain unfriendly villages, 2) the careful selection of agents to distribute US influence while not being redundant with other US influence efforts.

References

1. Field Manual 3-24, Counterinsurgency. Department of the Army, Washington DC (2006)
2. Jackson, M.: Social and Economic Networks. Princeton University Press, Princeton (2008)
3. Abelson, R.: Mathematical Models of the Distribution of Attitudes Under Controversy. In: Frederiksen, N., Gulliksen, H. (eds.) Contributions to Mathematical Psychology, pp. 141–160. Holt, Rinehart (1964)
4. DeGroot, M.: Reaching a Consensus. J. Amer. Statistical Assoc. 69(345), 118–121 (1974)
5. Friedkin, N., Johnsen, E.: Social Influence Networks and Opinion Change. Adv. Group Processes 16, 1–29 (1999)
6. Deffuant, G., Neau, D., Amblard, F.: Mixing Beliefs Among Interacting Agents. Adv. Complex Syst. 3, 87–98 (2000)
7. Deffuant, G., Amblard, F., Weisbuch, G., Faure, T.: How can extremism prevail? A study on the relative agreement interaction model. J. Art. Soc. Social Sim. 5(4) (2002)
8. Hegselmann, R., Krause, U.: Opinion Dynamics and Bounded Confidence Models: Analysis and Simulation. J. Art. Soc. Social Sim. 5(3) (2002)
9. Acemoglu, D., Ozdaglar, A., ParandehGheibi, A.: Spread of (Mis)Information in Social Networks. MIT LIDS report 2812, to appear in Games Econ. Behav. (2010)
10. Acemoglu, D., Como, G., Fagnani, F., Ozdaglar, A.: Opinion fluctuations and persistant disagreement in social networks. Submitted to Ann. Appl. Probab. (2010)
11. Borgatti, S.: Identifying sets of key players in a social network. Comp. Math. Org. Theory 12, 21–34 (2006)
12. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the Spread of Influence through a Social Network. SIGKDD (2003)
13. Hung, B.: Optimization-Based Selection of Influential Agents in a Rural Afghan Social Network. Master's Thesis, MIT, Operations Research Center, Thesis Supervisor, Asuman Ozdaglar (June 2010)

Identifying Health-Related Topics on Twitter
An Exploration of Tobacco-Related Tweets as a Test Topic

Kyle W. Prier[1], Matthew S. Smith[2],
Christophe Giraud-Carrier[2], and Carl L. Hanson[1]

[1] Department of Health Science
[2] Department of Computer Science
Brigham Young University, Provo UT 84602, USA
{kyle.prier,smitty,carl_hanson}@byu.edu,
cgc@cs.byu.edu

Abstract. Public health-related topics are difficult to identify in large conversational datasets like Twitter. This study examines how to model and discover public health topics and themes in tweets. Tobacco use is chosen as a test case to demonstrate the effectiveness of topic modeling via LDA across a large, representational dataset from the United States, as well as across a smaller subset that was seeded by tobacco-related queries. Topic modeling across the large dataset uncovers several public health-related topics, although tobacco is not detected by this method. However, topic modeling across the tobacco subset provides valuable insight about tobacco use in the United States. The methods used in this paper provide a possible toolset for public health researchers and practitioners to better understand public health problems through large datasets of conversational data.

Keywords: Data Mining, Public Health, Social Media, Social Networks, LDA, Tobacco Use, Topic Modeling.

1 Introduction

Over recent years, social network sites (SNS) like Facebook, Myspace, and Twitter have transformed the way individuals interact and communicate with each other across the world. These platforms are in turn creating new avenues for data acquisition and research. Such web-based applications share several common features [3], and while there are slight variations in actual implementations, each service enables users to (1) create a public profile, (2) define a list of other users with whom they share a connection, and (3) view and discover connections between other users within the system. Since SNS allow users to visualize and make public their social networks, this promotes the formation of new connections among users because of the social network platform [3,10]. Not only do SNS enable users to communicate with other users with whom they share explicit social connections, but with a wider audience of users with whom they would not have otherwise shared a social connection. Twitter, in particular, provides a medium whereby users can create and exchange user-generated content with a potentially larger audience than either Facebook or Myspace.

Twitter is a social network site that allows users to communicate with each other in real-time through short, concise messages (no longer than 140 characters), known as "tweets." A user's tweets are available to all of his/her "followers," i.e., all others who choose to subscribe to that user's profile. Twitter continues to grow in popularity. As of August 2010, a rough estimate of 54.5 million people used Twitter in the United States, with 63% of those users under the age of 35 years (45% were between the ages of 18 and 34) [17]. In addition, the US accounts for 51% of all Twitter users worldwide.

Because Twitter status updates, or tweets, are publicly available and easily accessed through Twitter's Application Programming Interface (API), Twitter offers a rich environment to observe social interaction and communication among Twitter users [15]. Recent studies have begun to use Twitter data to understand "real," or offline, health-related behaviors and trends. For example, Chew and Eysenbach analyzed tweets in an effort to determine the types and quality of information exchanged during the H1N1 outbreak [6]. The majority of Twitter posts in their study were news-related (46%), with only 7 of 400 posts containing misinformation. Scanfield et al. explored Twitter status updates related to antibiotics in an effort to discover evidence of misunderstanding and misuse of antibiotics [14]. Their research indicated that Twitter offers a platform for the sharing of health information and advice. In an attempt to evaluate health status, Culotta compared the frequency of influenza-related Twitter messages with influenza statistics from the Centers for Disease Control and Prevention (CDC) [7]. Findings revealed a .78 correlation with the CDC statistics suggesting that monitoring tweets might provide cost effective and quicker health status surveillance. These studies are encouraging. It is our contention that user-generated content on Twitter does indeed provide both public and relevant health-related data, whose identification and analysis adds value to researchers, and may allow them to tailor health interventions more effectively to their audiences.

In this paper, we address the problem of how to effectively identify and browse health-related topics within Twitter's large collection of conversational data. Although recent studies have implemented topic modeling to process Twitter data, these studies have focused on identifying high frequency topics to describe trends among Twitter users. Current topic modeling methods prove difficult to detect lower frequency topics that may be important to investigators. Such methods depend heavily on the frequency distribution of words to generate topic models. Public health-related topics and discussions use less frequent words and are therefore more difficult to identify using traditional topic modeling. Our focus, here, is on the following questions:

- How can topic modeling be used to most effectively identify relevant public health topics on Twitter?
- Which public health-related topics, specifically tobacco use, are discussed among Twitter users?
- What are common tobacco-related themes?

We choose to test our topic modeling effectiveness by focusing on tobacco use in the United States. Although we are interested in devising a method to identify

and better understand public health-related tweets in general, a test topic provides a useful indicator through which we can guide and gauge the effectiveness of our methodology. Tobacco use is a relevant public health topic to use as a test case, as it remains one of the major health concerns in the US. Tobacco use, primarily cigarette use, is one of the leading health indicators of 2010 as determined by the Federal Government [11]. Additionally, tobacco use is considered the most preventable cause of disease and has been attributed to over 14 million deaths in the United States since 1964 [16,1]. Also, there remain approximately 400,000 smokers and former smokers who die each year from smoking-related diseases, while 38,000 non-smokers die each year due to second-hand smoke [1,12].

In the following sections, we discuss our data sampling and analysis of tweets. We used Latent Dirichlet Allocation (LDA) to analyze terms and topics from the entire dataset as well as from a subset of tweets created by querying general, tobacco use-related terms. We highlight interesting topics and connections as well as limitations to both approaches. Finally, we discuss our conclusions regarding our research questions as well as possible areas of future study.

2 Methods and Results

In this section, we introduce our methods of sampling and collecting tweets. Additionally, we discuss our methods to analyze tweets through topic modeling. We used two distinct stages to demonstrate topic modeling effectiveness. First, we model a large dataset of raw tweets in order to uncover health-related issues. Secondly, we create a subset of tweets by querying a raw dataset with tobacco-related terms. We run topic modeling on this subset as well, and we report our findings.

2.1 Data Sampling and Collection

In order to obtain a representative sample of tweets within the United States, we chose a state from each of the nine Federal Census divisions through a random selection process. For our sample, we gathered tweets from the following states: Georgia, Idaho, Indiana, Kansas, Louisiana, Massachusetts, Mississippi, Oregon, and Pennsylvania.

Using the Twitter Search API, recent tweets for each state were gathered in two minute intervals over a 14-day period from October 6, 2010 through October 20, 2010. We used the Twitter Search API's "geocode" parameter to retrieve the most recent tweets from areas within each of the states, regardless of the content of the tweets. Since the "geocode" parameter requires a circle radius length, latitude and longitude, we gathered tweets from multiple specified circular areas that spanned each state's boundaries. To prepare the dataset for topic modeling, we remove all non-latin characters from the messages, replace all links within the dataset with the term "link," and only include users that publish at least two tweets. This process results in a dataset of 2,231,712 messages from 155,508 users. We refer to this dataset as the "comprehensive" dataset.

2.2   Comprehensive Dataset Analysis

We analyze the comprehensive dataset by performing LDA on it to produce a topic model. LDA is an unsupervised machine learning generative probabilistic model, which identifies latent topic information in large collections of data including text corpora. LDA represents each document within the corpora as a probability distribution over topics, while each topic is represented as a probability distribution over a number of words [2,9]. LDA enables us to browse words that are frequently found together, or that share a common connection, or topic. LDA helps identify additional terms within topics that may not be directly intuitive, but that are relevant.

We configure LDA to generate 250 topic distributions for single words (unigrams) as well as more extensive structural units (n-grams), while removing stopwords. We began at 100 topics and increased the topic number by increments of 50. At 250 topics, the model provided topics that contained unigrams and n-grams that exhibited sufficient cohesion. Additionally, we set LDA to run for 1,000 iterations. We suspected that this topic model would provide less-relevant topics relating to public health and specifically tobacco use, since such topics are generally less frequently discussed. The LDA model of the comprehensive dataset generally demonstrates topics related to various sundry conversational topics, news, and some spam as supported by Pear Analytics [5]. However, the model provides several topics, which contain health-related terms as shown in Table 1. The table includes a percent of the tweets that used any of the n-grams or unigrams within each topic. Although counting tweets that have any of the topic terms may be considered generous, we decided that within the study's context this measurement provides a relatively simple metric to evaluate the frequency of term usage.

Several themes are identifiable from the LDA output: physical activity, obesity, substance abuse, and healthcare. Through these topics we observe that those relating to obesity and weight loss primarily deal with advertising. Additionally, we observe that the terms relating to healthcare refer to current events and some political discourse. Such an analysis across a wide range of conversational data demonstrates the relative high frequency of these health-related topics. However, as we suspected, this analysis fails to detect lower frequency topics, specifically our test topic, tobacco use. In order to "bring out" tobacco use topics, we create a smaller, more focused dataset that we will refer to as the "tobacco subset."

2.3   Tobacco Subset Analysis

To build our "tobacco subset," we query the comprehensive dataset with a small set of tobacco-related terms: "smoking," "tobacco," "cigarette," "cigar," "hookah," and "hooka." We chose these terms because they are relatively unequivocal, and they specifically indicate tobacco use. We chose the term "hookah" because of an established trend, particularly among adolescents and college students, to engage in hookah, or waterpipe, usage [8]. The recent emergence of hookah bars provides additional evidence of the popularity of hookah smoking [13].

Table 1. Comprehensive Dataset Topics Relevant to Public Health. The last column is the percent of tweets that used any of the n-grams or unigrams within each topic.

Topic	Most Likely Topic Components (n-grams)	%
44	gps app, calories burned, felt alright, race report, weights workout, christus schumpert, workout named, chrissie wellington, started cycling, schwinn airdyne, core fitness, vff twitter acct, mc gold team, fordironman ran, fetcheveryone ive, logyourrun iphone, elite athletes, lester greene, big improvement, myrtle avenue	0.01
45	alzheimers disease, breast augmentation, compression garments, sej nuke panel, weekly newsletter, lab result, medical news, prescription medications, diagnostic imaging, accountable care, elder care, vaser lipo, lasting legacy, restless legs syndrome, joblessness remains, true recession, bariatric surgery, older applicants, internships attract, affordable dental	0.01
131	weight loss, diet pills, acai berry, healthy living, fat loss, weight loss diets, belly fat, alternative health, fat burning, pack abs, organic gardening, essential oils, container gardening, hcg diet, walnut creek, fatty acids, anti aging, muscle gain, perez hilton encourages	0.04
Topic	Most Likely Topic Components (unigrams)	%
18	high, smoke, shit, realwizkhalifa, weed, spitta, currensy, black, bro, roll, yellow, man, hit, wiz, sir, kush, alot, fuck, swag, blunt	13.63

Querying the comprehensive dataset with these terms, results in a subset of 1,963 tweets. The subset is approximately 0.1% of the comprehensive dataset. After running LDA on this subset, we found that there were insufficient data to determine relevant tobacco-related topics. We thus extended the tobacco subset to include tweets that were collected and preprocessed in the same manner as the comprehensive dataset. The only difference is that we used tweets collected from a 4-week period between October 4, 2010 and November 3, 2010. This resulted in a larger subset of 5,929,462 tweets, of which 4,962 were tobacco-related tweets (approximately 0.3%). Because of the limited size of the subset, we reduce the number of topics returned by LDA to 5, while we retain the output of both unigrams and n-grams.

The topics displayed in Table 2, contain several interesting themes relating to how Twitter users discuss tobacco-related topics. Topic 1 contains topics related not only to tobacco use, but also terms that relate to substance abuse including marijuana and crack cocaine. Topic 2 contains terms that relate to addiction recovery: quit smoking, stop smoking, quitting smoking, electronic cigarette, smoking addiction, quit smoking cigarettes, link quit smoking, and link holistic remedies. Topic 3 is less cohesive, and contains terms relating to addiction recovery: quit smoking, stop smoking, secondhand smoke, effective steps. Additionally, it contains words relating to tobacco promotion by clubs or bars: drink specials, free food, ladies night. Topic 4 contains terms related to both promotion by bars or clubs (ladies free, piedmont cir, hookahs great food, smoking room, halloween party, million people) and marijuana use (smoking weed, pot smoking). Topic 5 contains several terms that relate to anti-smoking

Table 2. Tobacco Subset Topics. The most likely topic components for each of the five topics generated by LDA. The last column is the percent of tweets that used any of the n-grams within each topic.

Topic	Most Likely Topic Components (n-grams)	%
1	smoking weed, smoking gun, smoking crack, stop smoking, cigarette burns, external cell phones, hooka bar, youre smoking, smoke cigars, smoking kush, hand smoke, im taking, smoking barrels, hookah house, hes smoking, ryder cup, dont understand, talking bout, im ready, twenty years people	0.16
2	quit smoking, stop smoking, cigar guy, smoking cigarettes, hookah bar, usa protect, quitting smoking, started smoking, electronic cigarette, cigars link, smoking addiction, cigar shop, quit smoking cigarettes, chronical green smoke, link quit smoking naturally, smoking pot, youtube video, link quit smoking, link holistic remedies, chronical protect	0.27
3	cigarette smoke, dont smoke, quit smoking, stop smoking, smoking pot, im gonna, hookah tonight, smoking ban, drink specials, free food, ladies night, electronic cigarettes, good times, smoking session, cigarette break, secondhand smoke, everythings real, effective steps, smoking cigs, smoking tonight	0.22
4	smoking weed, cont link, ladies free, piedmont cir, start smoking, hate smoking, hookahs great food, cigarette butts, thingswomenshouldstopdoing smoking, lol rt, sunday spot, cigarettes today, fletcher knebel smoking, pot smoking, film stars, external cell, fetishize holding, smoking room, halloween party, million people	0.25
5	smoke cigarettes, smoking hot, im smoking, smoking section, stopped smoking, chewing tobacco, smoking kills, chain smoking, smoking area, ban smoking, people die, ring ring hookah ring ring, love lafayette, link rt, damn cigarette, healthiest smoking products, theyre smoking, hate cigarettes, world series, hideout apartment	0.06

and addiction recovery themes: stopped smoking, smoking kills, chain smoking, ban smoking, people die, damn cigarette, hate cigarettes. In this case, topic modeling has helped to understand more fully how users are using tobacco-related tweets.

3   Discussion and Conclusion

In this study, we directly addressed the problem of how to effectively identify and browse health-related topics on Twitter. We focus on our test topic, tobacco use, throughout the study to explore the realistic application and effectiveness of LDA to learn more about health topics and behavior on Twitter. As expected, we determined that implementing LDA over a large dataset of tweets provides very few health topics. The health topics that LDA does produce during this first stage suggests the popularity of these topics in our dataset. The topic relating to weight loss solutions indicate a high frequency of advertisements in this area.

The topic relating to healthcare, Obama, and other current health issues indicate a current trend in political discourse related to health. Additionally, the high frequency of marijuana-related terms indicates a potentially significant behavior risk that can be detected through Twitter. While this method did not detect lower frequency topics, it may still provide public health researchers insight into popular health-related trends on Twitter. This suggests that there is potential research needed to test LDA as an effective method to identify health-related trends on Twitter. The first method can provide answers to research questions regarding what topics in general are most discussed among Twitter users. The second method we used to identify tobacco-related topics appears to be most promising to identify and understand public health topics. Although this method is less automated and requires us to choose terms related to tobacco use, the results indicated this method may be a valuable tool for public health researchers.

Based on the results of our topic model, Twitter has been identified as a potentially useful tool to better understand health-related topics, such as tobacco. Specifically, this method of Twitter analysis enables public health researchers to better monitor and survey health status in order to solve community health problems [4]. Our results suggest that chronic health behaviors, like tobacco use, can be identified and measured over shorter periods of time. However, our results do not verify the extent to which short term health events like disease outbreaks can or should be surveyed with this methodology, since tobacco use was used as a test topic. Also, we suspect that the demographics of Twitter users may affect the extent to which topics are discussed on Twitter. We suspect that different public health-related topics may be more or less frequently used on Twitter, and we recommend additional study in this area. Because LDA generates relevant topics relating to tobacco use, we are able to determine themes and also the manner in which tobacco is discussed. In this way, irrelevant conversations can be removed, while tweets related to health status can be isolated. Additionally, by identifying relevant tweets to monitor health status, public health professionals are able to create and implement health interventions more effectively. Researchers can collect almost limitless Twitter data in their areas that will provide practitioners with useful, up-to-date information necessary for understanding relevant public health issues and creating targeted interventions.

Finally, the results from the second method suggest that researchers can better understand how Twitter, a popular SNS, is used to promote both positive and negative health behaviors. For example, Topic 4 contains terms that indicate that establishments like bars, clubs, and restaurants use Twitter as a means to promote business as well as tobacco use. In contrast, Topic 2 contains words that relate to addiction recovery by promoting programs that could help individuals quit smoking.

The use of LDA in our study demonstrates its potential to extract valuable topics from extremely large datasets of conversational data. While the method proves a valuable outlet to automate the process of removing irrelevant information and to hone in on desired data, it still requires careful human intervention to select query terms for the construction of a relevant subset, and subsequent

analysis to determine themes. Research is required to further automate this process. In particular, new methods that can identify infrequent, but highly relevant topics (such as health) among huge datasets will provide value to public health researchers and practitioners, so they can better identify, understand, and help solve health challenges.

References

1. Armour, B.S., Woolery, T., Malarcher, A., Pechacek, T.F., Husten, C.: Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses. Morbidity and Mortality Weekly Report 54, 625–628 (2005)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. Journal of Machine Learning Research 3, 993–1022 (2003)
3. Boyd, D.M., Ellison, N.B.: Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication 13, 210–230 (2008)
4. Centers for Disease Control and Prevention, http://www.cdc.gov/od/ocphp/nphpsp/essentialphservices.htm
5. Pear Analytics, http://www.pearanalytics.com/blog/2009/twitter-study-reveals-interesting-results-40-percent-pointless-babble/
6. Chew, C.M., Eysenbach, G.: Pandemics in the Age of Twitter: Content Analysis of "tweets" During the, H1N1 Outbreak. Public Library of Science 5(11), e14118 (2010) (Paper presented 09/17/09 at Medicine 2.0, Naastricht, NL)
7. Culotta, A.: Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. In: Proceedings of the KDD Workshop on Social Media Analytics (2010)
8. Eissenberg, T., Ward, K.D., Smith-Simone, S., Maziak, W.: Waterpipe Tobacco Smoking on a U.S. College Campus: Prevalence and Correlates. Journal of Adolescent Health 42, 526–529 (2008)
9. Griffiths, T.L., Steyvers, M.: Finding Scientific Topics. Proceedings of the National Academy of Sciences 101, 5228–5235 (2004)
10. Haythornthwaite, C.: Social Networks and Internet Connectivity Effects. Information, Communication, & Society 8, 125–147 (2005)
11. Healthy People (2010), http://www.healthypeople.gov/lhi/
12. Mokdad, A.H., Marks, J.S., Stroup, D.F., Gerberding, J.L.: Actual Causes of Death in the United States. Journal of the American Medical Association 291, 1238–1245 (2004)
13. Primack, B.A., Aronson, J.D., Agarwal, A.A.: An Old Custom, a New Threat to Tobacco Control. American Journal of Public Health 96, 1339 (2006)
14. Scanfield, D., Scanfield, V., Larson, E.: Dissemination of Health Information through Social Networks: Twitter and Antibiotics. American Journal of Infection Control 38, 182–188 (2010)
15. Twitter API documentation, http://dev.twitter.com/doc
16. U.S. Department of Health and Human Services. The Health Consequences of Smoking: A Report for the Surgeon General. Report, USDHHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health (2004)
17. Quantcast, http://www.quantcast.com/twitter.com

Application of a Profile Similarity Methodology for Identifying Terrorist Groups That Use or Pursue CBRN Weapons

Ronald L. Breiger[1], Gary A. Ackerman[2], Victor Asal[3], David Melamed[1], H. Brinton Milward[1], R. Karl Rethemeyer[3], and Eric Schoon[1]

[1] University of Arizona
[2] START Center, University of Maryland
[3] University at Albany

Abstract. No single profile fits all CBRN-active groups, and therefore it is important to identify multiple profiles. In the analysis of terrorist organizations, linear and generalized regression modeling provide a set of tools to apply to data that is in the form of cases (named groups) by variables (traits and behaviors of the groups). We turn the conventional regression modeling "inside out" to reveal a network of relations among the cases on the basis of their attribute and behavioral similarity. We show that a network of profile similarity among the cases is built in to standard regression modeling, and that the exploitation of this aspect leads to new insights helpful in the identification of multiple profiles for actors. Our application builds on a study of 108 Islamic jihadist organizations that predicts use or pursuit of CBRN weapons.

Keywords: Networks, profile similarity, CBRN use or pursuit, terrorism.

1   Introduction

Use or pursuit of chemical, biological, radiological and nuclear weapons (CBRN) on the part of violent non-state actors (VNSA) has been a significant issue for policy makers, academics and the public alike for more than a decade. While concern over the dissemination of chemical and biological weapons predates their use during World War I, Ackerman [1] notes that the widespread contemporary attention to the use and pursuit of weapons capable of inflicting mass casualties was catalyzed in 1995 when the Aum Shinrikyo cult released sarin chemical nerve agent in the Tokyo subway system. Concerns over mass casualty attacks increased dramatically following the jihadist attacks on the World Trade Center on September 11, 2001 [2]. The unprecedented expansion of research and policy initiatives aimed at combating mass casualty attacks, coupled with a growing body of scholarly literature which hypothesizes that radical Islamist organizations—particularly those with a stated jihadist mission—are among the groups most likely to pursue unconventional weapons, has led to an increased focus on predicting the use and pursuit of CBRN among these ideologically motivated organizations, including the use of rigorous multiple regression methods [3,4]. The emphasis on predicting radical Islamist groups' involvement with CBRN has been

validated by credible evidence of their interest in these weapons [5]. An overarching question is where analytic attention should be focused [6].

We intend this paper as an illustrative demonstration of the kinds of analysis that are possible when standard regression models are viewed in a new light, one that illuminates the network of profile similarity among the cases that is assumed by the regression models even as it is typically invisible to the analysts.

In this paper we illustrate our analytical approach by using it to build on the analysis of Asal and Rethemeyer [3], as their work and the data they utilize stands as an exemplar of rigorous quantitative analysis of CBRN use and pursuit. We recognize nonetheless that a more solid database is necessary in order to advance research that can inform policy. Indeed, a primary objective of our research project is the substantial enhancement of open-source structured databases on CBRN activities by sharply increasing the number of CBRN events coded as well as the number of variables tracked for each event, by developing and applying stronger methods for validating events and assessing the reliability of knowledge of each coded occurrence, and by allowing for longitudinal analysis.

2   Data and Measures

Asal and Rethemeyer's analysis of CBRN use and pursuit encompassed 108 Islamic jihadist groups of which 14 were coded as having pursued or used CBRN weapons.[1] As the outcome was coded as binary, logistic regression was employed.

Two data sources on CBRN activities were leveraged. The Monterey WMD Terrorism Database, widely regarded as the most comprehensive unclassified data source on the use of CBRN agents by non-state actors, provided data on the use and attempted use of CBRN materials by VNSAs. The Monterey data were combined with data from the National Memorial Institute for the Prevention of Terrorism (MIPT) Terrorism Knowledge database (TKB), and additional data was collected by Asal and Rethemeyer. The data are complete for the years 1998 to 2005.

The predictor variables capture features of the organizations as well as of their environments. At the organizational level, *membership* is a four-category order-of-magnitude measure of group size ranging from under 100 members to over 10,000. *Age* is number of years from the group's inception to 2005, and *age-squared* is also included. *Eigenvector centrality* is applied to data (from the "related groups" module of the TKB; see [3] for further discussion) on the alliance connections that terrorist organizations have established among themselves. This measure of centrality recursively weights the number of alliance connections of each organization by the centrality of the groups with which it is connected. *Experience* is a dummy variable for groups known to have carried out at least four previous attacks (of any type). *Control of territory* is a dummy variable for groups that maintain control of extensive territory in their "host" country.

Turning now to environmental variables, *state sponsorship* refers to whether (1) or not (0) an organization is coded as being substantially backed by the resources of a

---

[1] Ten organizations considered and pursued CBRN terrorism to some degree but were unsuccessful in actually perpetrating an attack. Four other organizations succeeded at least once in deploying CBRN materials in an attack, none of which killed a large number of people [3].

state. The remaining environmental variables pertain to the country in which each organization has its home base. *Democracy* (POLITY2) is an assessment of regime type on a scale ranging from strongly autocratic (-10) to strongly democratic (+10). *Energy per capita*, energy consumption relative to total population, taps the level of economic development. To account for degree of economic embeddedness in the global West, trade data between the organization's home base country and the United States (*logged exports to US*) were employed. In an effort to measure a country's link to Western culture, a count of McDonald's restaurants in each home country (*McDonald's*) was obtained. *Civil strife* is the number of years of civil strife (whether isolated to the home country or internationalized) reported for the study period. Much related information about these variables is given in [3].

3   Profile Similarity

We begin with linear regression and then provide some generalization to logistic regression. Consider an $n \times p$ data matrix (denoted $\mathbf{X}$) whose rows represent each of the n cases (in our example, 108 terrorist organizations) and whose columns stand for the $p - 1$ predictor variables (plus an additional column of 1's, pertaining to the intercept in the regression equation; $p = 13$ in our example). Assuming a continuous outcome variable, $\mathbf{y}$ $(n \times 1)$, matrix notation for the fitted values of $\mathbf{y}$ (denoted $\hat{\mathbf{y}}$) in the linear regression model is

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} \tag{1}$$

where $\mathbf{b}$ is a $p \times 1$ vector of regression coefficients estimated by the ordinary least-squares criterion. We compute the singular value decomposition (SVD) of $\mathbf{X}$,

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}} \tag{2}$$

the point of which (for our purposes) is to produce (as the columns of $\mathbf{U}$) a set of orthogonal dimensions pertaining to the rows of $\mathbf{X}$ (the terrorist organizations), and (as columns of $\mathbf{V}$) a set of orthogonal dimensions for the columns of $\mathbf{X}$ (the predictor variables). (Superscript $\mathbf{T}$ denotes matrix transposition.) $\mathbf{S}$ is a diagonal matrix of weights (singular values) indicating relative importance of each dimension. We will be working only with $\mathbf{U}$. Substitution of (2) into (1) shows that eq. (1) is identical to

$$\left[\mathbf{U}\mathbf{U}^{\mathbf{T}}\right]\mathbf{y} = \hat{\mathbf{y}} \tag{3}$$

We will refer to the n x n matrix $\mathbf{U}\mathbf{U}^{\mathbf{T}}$ as the *projection matrix*, following standard usage.[2] Eq. (3) is an established result, widely known. We propose to use this result in an innovative way, however. The projection matrix is a network of profile similarity among the cases. In our application, the off-diagonal entries of this matrix indicate the degree of similarity among terrorist organizations i and j with respect to the attributes and behaviors specified in the regression model. We want to turn the regression model "inside out" to express its fitted values for the outcome variable as a function of this network among the cases.

---

[2] This square, symmetric, and idempotent matrix is known in mathematical statistics as the *projection matrix*, and in the study of regression diagnostics as the *hat matrix* (e.g., [7]).

Moving on to logistic regression, which is relevant to our application in that CBRN activity is coded as a binary variable, the logistic regression model implies

$$\log(\hat{p}/1-\hat{p}) = \mathbf{X}\hat{\beta} \tag{4}$$

where $\hat{\mathbf{p}}$ is the modeled estimate of the probability that each organization (in turn) engages in CBRN activity, and the vector on the right contains the logistic regression coefficients. We will define a diagonal matrix of weights, $\mathbf{W}$, where $w_{ii} = \hat{p}_i (1 - \hat{p}_i)$. We will compute the SVD of a weighted version of the design matrix $\mathbf{X}$:

$$\mathbf{W^{1/2}X} = \mathbf{U^* S^* V^{*T}} \tag{5}$$

where the superscript asterisks simply indicate that the respective quantities are different from those of eq. (2). Now, although logistic regression is nonlinear, we can transform logistic regression to a form paralleling eq. (3), as follows:

$$\left[\mathbf{U^* U^{*T}}\right]\left(\mathbf{W^{1/2} z}\right) = \left(\mathbf{W^{1/2} X}\,\hat{\beta}\right) \tag{6}$$

where $\mathbf{z}$ is a vector of pseudovalues that stand for the observed binary response variable [8].[3] The term on the left is the projection matrix for logistic regression [8]. The second term on the left is (a weighted version of) the response variable. The term on the right is (a weighted form of the logits of) the estimated response variable (see (4)).

Let's take a look inside the projection matrix in (6), with respect to Asal and Rethemeyer's logistic regression analysis of 108 terrorist groups.

4   Results

Fig. 1 is a dimensional representation of the cosines among all pairs of rows of matrix $\mathbf{U^*}$. This cosine matrix has entries $\mathbf{P}_{ij} / \sqrt{(\mathbf{P}_{ii} \times \mathbf{P}_{jj})}$ where $\mathbf{P} = \mathbf{U^* U^{*T}}$ is the projection matrix in (6). The dominant eigenvector tends to place organizations with CBRN activity (indicated by dark arrows) on the right, with 13 = al Qaeda at the extreme. Some other organizations are placed with them (e.g., 3 = Abu Sayyaf, 14 = al Qaeda in the Arabian Peninsula, AQAP) and therefore warrant further study. Probabilities of CBRN activity predicted by the logistic regression model are a function of scores on the first eigenvector of Fig. 1. Correlation of those scores with the cube root of the predicted probabilities is +.97 across the 108 cases.

The second eigenvector, though not dominant, splits the CBRN-active groups according to differences in their attributes and behaviors. At the top of dimension 2, groups 45, 51, 47, and 50 include three Kashmiri groups (51 = Lashkar-e-Taiba, LeT) and Lashkar-e-Jhangvi (50). Clustering of groups is important because an analyst wanting to identify the smallest set of groups that are "close" in their profiles to the largest number of CBRN-active organizations would do well to select groups from different regions of Fig. 1. For example, consider the top eight groups that each group is most similar to in the projection matrix in eq. (6). Group 51 (LeT) and group 48 (Jemaah Islamiya) are on opposite "sides" of dimension 2 of Fig. 1, and each is

---

[3] The pseudovalues are defined such that $\exp(\mathbf{z}) / (1 + \exp(\mathbf{z}))$ is an excellent approximation of the observed binary response, $\mathbf{y}$ (see **[8]**).
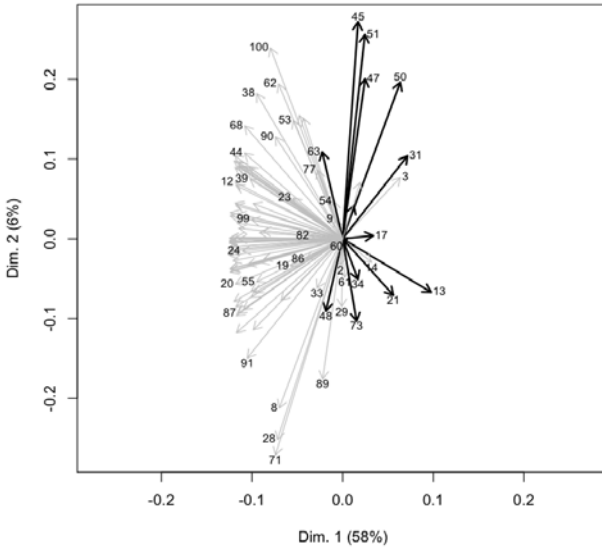
Fig. 1. Plot of first two eigenvectors of cosine matrix derived from $[\mathbf{U}^*\mathbf{U}^{*T}]$. Dark arrows point to organizations coded as using or pursuing CBRN. Organizations are named with numbers, some of which are omitted to enhance clarity.

closest to sets of four CBRN-active groups that do not overlap with each another. Likewise, group 31 (Hamas) is closest to four groups that do not overlap with either of the other two sets of four groups. In aggregate, these three groups (51, 48, 31) share strong profiles with 12 of the 14 CBRN-active groups to be found in our 108-group dataset. Clustering matters, because there is not a single profile that characterizes all CBRN-active groups. Clusters or families of groups do not "look like" other clusters.

The preceding discussion is important because it reveals clustering (and opportunities to identify distinctively different CBRN profiles) inside the projection matrix that is taken for granted in logistic regression modeling. In order to construct that projection matrix, however, knowledge of the dependent variable is required. (Notice from eq. 5 that matrix $\mathbf{W}$ requires the modeled probabilities). In addition to understanding these implications of the logistic regression model, we would like, in the spirit of exploratory analysis, to identify clustering without a reliance on having to know the results of this (or any other) regression model. Therefore, we will now search for clusters relevant to the projection matrix of eq. (3), which depends only on the predictor variables, not on the analyst's knowledge of the outcome variable.

With respect to matrix $\mathbf{U}$ of eq. (2), we estimated nine k-means cluster analysis models, yielding two through ten clusters. For each analysis we calculated the sums of squared errors. We then plotted the sum of squares by the number of clusters, yielding a figure that resembles a scree plot in its appearance and interpretation. Based on this plot we determined that a four-cluster solution provides an adequate tradeoff of parsimony vs. fit in defining the clusters for our analyses.

Partitioning the 108 Islamist groups into four clusters yields a 71-group cluster, a 22-group cluster, a 14-group cluster, and a 1-group cluster. The latter cluster (D) consists of an organization (the Islami Chhatra Shibir, ICS) that is an outlier with respect to several of the predictor variables and, consequently, does not have a high profile similarity with any of the other groups. The ICS was founded in 1941 in Bangladesh and is thus more than a quarter-century older than the next oldest group in the dataset.

To provide context for the following discussion, we note that the overall baseline probability in this dataset for pursuing or using CBRN weapons is .13 (= 14 groups out of 108 Islamist groups in the dataset).

The largest cluster, with 71 groups (A), contains only one organization coded as using or pursuing CBRN weapons. That group is the East Turkistan Liberation Organization (ETLO), a Uyghur group operating in central Asia.

Cluster B, consisting of 22 groups, contains 8 cases coded as pursuing or using CBRN weapons. This is a high concentration (.36) compared with the overall baseline probability of .13. Cluster C contains 14 organizations, of which 5 are coded as using or pursuing CBRN weapons, again a much higher conditional probability (.36) than the overall baseline. Thus, two clusters have conditional probabilities of CBRN activity that are close to 0, and two others have comparatively much higher likelihood.

The first column of Table 1 presents means and standard deviations for the variables in our analysis. The remaining columns provide variable means (and standard deviations for quantitative variables) by cluster.

In comparison to all groups in the dataset, those groups in cluster B are more connected based on eigenvector centrality ($t_{(106)}$ = 2.4, $p<.05$), are more likely to have at least four previous attacks ($t_{(106)}$ = 13.59, $p<.001$), and are less likely to control territory than the other groups ($t_{(106)}$ = 2.05, $p<.05$). The groups in cluster C are more likely than the average group in the dataset to be large ($t_{(106)}$ = 4.81, $p<.001$), older, ($t_{(106)}$ = 2.25, $p<.05$), state sponsored ($t_{(106)}$ = 3.52, $p<.001$), connected ($t_{(106)}$ = 3.34, $p<.01$), to have at least four previous attacks ($t_{(106)}$ = 3.28, $p<.01$) and have more years of civil strife ($t_{(106)}$ = 2.03, $p<.05$).

Clusters B and C differ from each other in some notable ways. Every single group in cluster C controls territory; not a single group in Cluster B controls territory. Cluster C organizations are on average older than those in B, and more likely to be coded as state-sponsored. Every group in cluster B has launched at least four attacks, whereas 64% of groups in cluster C have done so.

Organizations in cluster C include Ansar al-Islam, al-Qaeda, Hamas, and Hezbollah (all of whom are coded as having used or pursued CBRN), as well as the Abu Sayyaf Group, the Taliban, and the Mahdi Army (major terrorist groups but coded as having no CBRN activity). Groups in cluster B coded as having CBRN activity include four Pakistani groups (Lashkar-e-Taiba, Lashkar-e-Jhangvi, Jaish-e-Mohammad, Jamiat ul-Mujahedin) and an assortment of others based in Algeria, Russia, Iraq, and elsewhere.

The characteristics of clusters with propensities for CBRN activity that are considerably *below* the baseline are also helpful to an analyst wanting to predict CBRN activities among these organizations. Cluster A, for example, contains 71 organizations that, on average, have very low CBRN activity and that stand out for having the smallest-size memberships, youngest organizational age, lowest network connectivity, least experience in conducting attacks, and a very low likelihood of controlling territory, all in comparison to the other clusters.

Table 1. Means (and standard deviations) for variables by cluster

Variable	Overall (N=108)	Cluster A (N=71)	Cluster B (N=22)	Cluster C (N=14)	Cluster D (N=1)
Use or Pursuit of CBRN	.13	.02[***]	.36[***]	.36[**]	.00
Organization Membership	.62 (.81)	.38[***] (.68)	.77 (.62)	1.5[***] (.94)	2.00
Organizational Age	9.76 (11.07)	1.01[***] (8.50)	11.64 (11.06)	15.86[*] (10.29)	65
Organizational Age-Squared	216.67 (503.48)	123.00[**] (261.50)	252.09 (408.84)	349.71 (378.39)	4,225
State Sponsorship	.14	.08[*]	.14	.43[***]	.00
Eigenvector Centrality	7.45 (11.27)	4.23[***] (6.18)	12.48[*] (9.71)	16.42[**] (22.30)	.00
Four or More Attacks	.29	.00[***]	1.00[***]	.64[**]	.00
Control Territory	.13	.00[***]	.00[*]	1.00[1]	.00
Democracy (POLITY2)	-.94 (7.64)	-1.76 (.91)	1.77 (7.16)	-1.50 (7.67)	6.00
Energy Per Capita	1.61 (1.62)	1.61 (1.63)	1.97 (1.81)	1.09 (1.16)	.11
Logged Exports to the US	20.48 (2.98)	20.18 (3.34)	21.36 (1.51)	20.59 (2.72)	21.24
Number of McDonald's	40.98 (108.43)	28.35 (8.08)	73.95 (196.56)	56.14 (79.42)	.00
Years in Civil Strife	3.31 (3.24)	2.90 (3.03)	3.77 (3.56)	4.93[*] (3.45)	.00

Note: t-tests compare groups within each cluster to all others. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

5 Concluding Remarks

Threat assessment is not productive if it is conceived as a one-size-fits-all endeavor. While good coefficient estimates for the effects of predictor variables on terrorist be- haviors will continue to be vitally important contributions, we also need methods for discovering how the same variable might be both present and absent in distinctively different profiles of terrorist behavior, depending on how that variable combines with others in different contexts. A good illustration in this paper is the control of territory. Every group in cluster C controlled extensive territory; no group in cluster B did. Yet both clusters were 2.8 times as likely as chance to contain groups engaged in use or pursuit of CBRN weapons. In this paper we have shown how a network of profile similarity among the cases (potentially exhibiting different profiles in different sec- tors of profile space) is implicit in both linear and nonlinear regression modeling, and we have suggested the benefits of turning standard regression models "inside out" in

order to exploit analysis of profile similarity among the cases. Similar in spirit (though different in operational procedures) is work of Ragin [9] as applied in a recent RAND study of the relative effectiveness of multiple profiles of counterinsurgency activities across a range of cases [10].

As we related at the front of our paper, a primary objective of our research project is the substantial enhancement of open-source structured databases on CBRN activities. This expanded data context will be key to the further development of our profile similarity methodology. The methods and knowledge necessary to assess CBRN threat are evolving simultaneously with VNSAs' ability to acquire these resources. This makes the development of methods that are capable of comprehensively analyzing the rapidly expanding body of data and knowledge essential for preventative efforts. It is our hope that the new approach reported herein, that we are continuing to develop, will stand as a meaningful contribution toward this effort.

References

1. Ackerman, G.: Defining Knowledge Gaps within CBRN Terrorism Research. In: Ranstorp, M., Normark, M. (eds.) Unconventional Weapons and International Terrorism: Challenges and New Approaches, pp. 13–25. Routledge, London (2009)
2. Ackerman, G.: The Future of Jihadists and WMD. In: Ackerman, G., Tamsett, J. (eds.) Jihadists and Weapons of Mass Destruction, pp. 359–400. CRC Press, Boca Raton (2009)
3. Asal, V.H., Rethemeyer, R.K.: Islamist Use and Pursuit of CBRN Terrorism. In: Ackerman, G., Tamsett, J. (eds.) Jihadists and Weapons of Mass Destruction, pp. 335–358. CRC Press, Boca Raton (2009)
4. Ivanova, K., Sandler, T.: CBRN Attack Perpetrators: An Empirical Study. Foreign Policy Analysis 3, 273–294 (2007)
5. Mowatt-Larssen, R.: Al Qaeda Weapons of Mass Destruction Threat: Hype Or Reality? Harvard Kennedy School Belfer Center for International Affairs, Cambridge (2010)
6. Raab, J., Milward, H.B.: Dark Networks as Problems. J. Public Adm. Res. Theory 13, 413–439 (2003)
7. Belsley, D.A., Kuh, E., Welsch, R.E.: Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley, Hoboken (2004)
8. Pregibon, D.: Logistic Regression Diagnostics. The Annals of Statistics 9, 705–724 (1981)
9. Ragin, C.C.: Redesigning Social Inquiry: Fuzzy Sets and Beyond. University of Chicago Press, Chicago (2008)
10. Paul, C., Clarke, C.G., Grill, B.: Victory has a Thousand Fathers: Sources of Success in Counterinsurgency. RAND Corporation document number MG-964-OSD (2010)

Cognitive Aspects of Power in a Two-Level Game

Ion Juvina[1], Christian Lebiere[1], Jolie Martin[2], and Cleotilde Gonzalez[2]

[1] Carnegie-Mellon University, Department of Psychology,
5000 Forbes Ave., Pittsburgh, PA 15213, USA
[2] Carnegie-Mellon University, Department of Social and Decision Sciences, Dynamic Decision
Making Laboratory, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
{ijuvina,cl,jolie,coty}@cmu.edu

Abstract. The Intergroup Prisoner's Dilemma with Intragroup Power Dynamics (IPD^2) is a new game paradigm for studying human behavior in conflict situations. IPD^2 adds the concept of intragroup power to an intergroup version of the standard Iterated Prisoner's Dilemma game. We conducted an exploratory laboratory study in which individual human participants played the game against computer strategies of various complexities. We also developed a cognitive model of human decision making in this game. The model was run in place of the human participant under the same conditions as in the laboratory study. Results from the human study and the model simulations are presented and discussed, emphasizing the value of including intragroup power in game theoretic models of conflict.

Keywords: Iterated Prisoner's Dilemma, Intergroup Prisoner's Dilemma, Intragroup Power, Cognitive modeling.

1 Introduction

Erev and Roth have argued for the necessity of a Cognitive Game Theory that focuses on players' thought processes and that develops simple general models that can be appropriately adapted to specific circumstances, as opposed to building or estimating specific models for each game of interest [1]. In line with this approach, Lebiere, Wallach, and West [2] developed a cognitive model of the Prisoner's Dilemma (PD) that generalized to other games from Rapoport et al.'s taxonomy of 2X2 games [3]. This model leverages basic cognitive abilities such as memory by making decisions based on records of previous rounds stored in long-term memory. This record includes only directly experienced information such one's own move, the other player's move, and the payoff. The decision is accomplished by a set of rules that, given each possible action, retrieves the most likely outcome from memory and selects the move with the highest payoff. The model predictions originate from and are strongly constrained by learning mechanisms occurring at the sub-symbolic level of the ACT-R cognitive architecture [4]. The current work builds upon and extends this model. We use abstract representations of conflict as is common in the field of Game Theory [5], but we are interested in the actual (rather than normative) aspects of human behavior that explains how people make strategic decisions given their subjective experiences and cognitive constraints [6].

In understanding human behavior in real world situations, it is also necessary to capture the complexities of their interactions. The dynamics of many intergroup conflicts can usefully be represented by a two-level game [7]. At the intragroup level, various factions (parties) pursue their interests by trying to influence the policies of the group. At the intergroup level, group leaders seek to maximize their gain with respect to other groups while also satisfying their constituents. For example, domestic and international politics are usually entangled: international pressure leads to domestic policy shifts and domestic politics impact the success of international negotiations [7]. A basic two-level conflict game extensively studied is the Intergroup Prisoner's Dilemma [8]. In this game, two levels of conflict (intragroup and intergroup) are considered simultaneously. The intragroup level consists of an *n*-person Prisoner's Dilemma (PD) game while the intergroup level is a regular PD game. A variant of this game, the Intergroup Prisoner's Dilemma - Maximizing Difference, was designed to study what motivates individual self-sacrificial behavior in intergroup conflicts [9].

A characteristic of social real-world interactions that the two-level games currently do not represent is *power*. Here we introduce *intragroup power* to an Intergroup Prisoner's Dilemma game as a determinant of each actor's ability to maximize long-term payoffs through between-group cooperation or competition. The game we introduce here, the Intergroup Prisoner's Dilemma with Intragroup Power Dynamics (IPD^2), intends to reproduce the two-level interactions in which players are simultaneously engaged in an intragroup power struggle and an intergroup conflict. We introduce a *power* variable that represents both *outcome power* and *social power*. Outcome power (power to) is the ability of an actor to bring about outcomes, while social power (power over) is the ability of an actor to influence other actors [10]. The outcomes are reflected in payoff for the group. Oftentimes in the real world, power does not equate with payoff. Free riders get payoff without having power. Correspondingly, power does not guarantee profitable decision-making. However, a certain level of power can be a prerequisite for achieving significant amounts of payoff, and in return, payoff can cause power consolidation or shift. A leader who brings positive outcomes for the group can consolidate her position of power, whereas negative outcomes might shift power away from the responsible leader. We introduced these kinds of dynamics of payoff and power in the IPD^2.

2   Description of the IPD^2 Game

IPD^2 is an extension of the well-known Iterated Prisoner's Dilemma paradigm. In the Iterated Prisoner's Dilemma, two players, "Player1" and "Player2," each have two actions that can be referred to as "cooperate" (C) or "defect" (D). The players choose their actions simultaneously and repeatedly. The two players receive their payoffs after each round, which are calculated according to a payoff matrix setting up a conflict between short-term and long-term payoffs (see example of the payoff matrix used in Table 1).

In IPD^2, two groups play an Iterated Prisoner's Dilemma game. Each group is composed of two players. Within a group, each player chooses individually whether to cooperate or defect, but only the choice of the player with the greatest power within the group counts as the group's choice. This is equivalent to saying that the two

Table 1. Payoff matrix used in Prisoner's Dilemma. Each cell shows values X/Y with X being the payoff to Player 1 and Y the payoff to Player 2 for corresponding row and column actions.

		Player2	
		C	D
Player1	C	1/1	-4/4
	D	4/-4	-1/-1

players simultaneously vote for the choice of the group and the vote of the player with more power bears a heavier weight. By analogy with the political arena, one player is the majority and the other is the minority. The majority imposes its decisions over the minority. In what follows, the player with the higher value of power in a group will be referred to as the *majority* and the player with the lower power in a group will be referred to as the *minority*.

A player's power is a quantity assigned at the start of the game and increased or decreased after each round of the game depending on the outcome of the interaction. The sum of power within a group remains constant throughout the game. If the two members of a group made the same decision (both played C or both played D), their powers do not change after the round other than for random variation. If they made different decisions, their powers change in a way that depends on the outcome of the inter-group game, as follows:

For the player in majority i,

$$\text{Power(i)}_t = \text{Power(i)}_{t-1} + \text{Group payoff}_t / s \qquad (1)$$

For the player in minority j,

$$\text{Power(j)}_t = \text{Power(j)}_{t-1} - \text{Group payoff}_t / s \qquad (2)$$

where Power_t is the current power at round t, Power_{t-1} is the power from the previous round, Group payoff_t is the current group payoff in round t from the inter-group Prisoner's Dilemma game, and s is a scaling factor (set to 100 in this study). The indices i and j refer to the majority and minority players, respectively.

The total payoff to the group in each round is shared between the two group mates in direct proportion to their relative power levels as follows:

$$\text{Payoff}_t = \text{Payoff}_{t-1} + \text{Power}_t * \text{Group payoff}_t / s \qquad (3)$$

where Payoff_t is the cumulative individual payoff after round t, Power_t is the current power at round t and Group payoff_t is that obtained from the inter-group Prisoner's Dilemma game matrix.

On a given round, individual power and payoff increases or decreases depending on the group payoff, the power status, and whether or not there is implicit consensus of choice between the two players on a group. The key feature is that in the absence of consensus, positive group payoffs will result in an increase in power for the majority while negative group payoffs will result in a decrease of power for the majority.

The players make simultaneous decisions and they receive feedback after each round. The feedback is presented in a tabular format as shown in Table 2. In this example, the human participant was randomly assigned to Group-2. The three computer strategies

were given non-informative labels. The choice of the majority player (i.e., the choice that counts as the group choice) is shown in bold font and gray background in Table 2. The cumulative payoff (payoff total) was shown in red when negative and in blue when positive (both are shown in black here).

Table 2. Example of feedback presented to the human participants after each round in the IPD^2 game

Group	Player	Choice	Power	Group payoff	Player payoff	Payoff total
Group-1	P1-1	**B**	**0.525**	1	0.005	-0.003
	P1-2	B	0.475		0.005	-0.007
Group-2	P2-1	A	0.265	1	0.003	0.087
	Human	**B**	**0.735**		0.007	0.143

The participants choose between A and B. The labels A and B are randomly assigned to Cooperate and Defect for each participant at the beginning of the experimental session. In the example shown in Table 2, A was assigned to Defect and B to Cooperate. The labels keep their meaning throughout the session, across rounds and games.

3   A Cognitive Model of IPD^2

A cognitive model of human behavior in IPD^2 was developed to help understand and explain the dynamics of power and the two-level interactions. This model was inspired by the Instance-Based Learning Theory (IBLT) [11] and implemented in the ACT-R cognitive architecture [4].

For each round of the game, the model checks whether it has previously encountered the same situation, and if so, what action (decision) it took in that situation. The situation is characterized by the choices of all the players and the group choices in the previous round. For example, let us assume that the current state of the game is the one represented in Table 2, where the human player is substituted by the model. We see that in the previous round the model played B, its group mate played A, and the two players on the opposite group played B (where A and B were randomly assigned to cooperate and defect, respectively). The group choice was B for both groups. This information constitutes the situation components of the instance. These attributes will act as the retrieval cues: the model will attempt to remember whether an identical situation has previously been encountered. The model's memory stores as an instance each situation that is encountered at each round of the game and the associated decision taken by the model in that situation.

For a given situation, the model can have up to two matching instances, one with A and the other one with B as the model's next decision. If two instances match the

current situation, the model will only retrieve the most active one. Activation of a memory element in ACT-R is a function of its frequency (how often it was used), recency (how recently it was used), and noise.

Once a decision is made, the model receives positive or negative payoff. The sign of the payoff (positive or negative) determines the course of action that is undertaken by the model. When the payoff is positive, the activation of the memory element that was used to make the decision increases. This makes it more likely that the same instance will be retrieved when the same situation reoccurs. When the payoff is negative, activation of the retrieved instance increases, and the model additionally creates a new instance by saving the current situation together with the alternative decision. Thus, after receiving a negative payoff, the model has two instances with the same situation and opposite decisions in its memory. When a similar situation occurs, the model will have two decision alternatives to choose from. In time, the more lucrative decision will be activated and the less lucrative decision will decay. Retrieving and creating instances makes them more active and more available for retrieval when needed. A decay process causes the existing instances to be forgotten if they are not frequently and recently used. In short, negative feedback causes the model to "consider" the alternative action, whereas positive feedback causes the model to persist in choosing the lucrative instance.

4 Human Data and Model Simulation Results

We conducted an exploratory laboratory study with 68 participants (average age 24, 19 females) recruited from the Carnegie Mellon University community with the aid of a web ad. Each human participant played IPD^2 together with three computer strategies: one as group mate and two as opponents. The three computer strategies were selected from a set of five strategies: always-cooperate, always-defect, tit-for-tat, and two other strategies designed specially for the IPD^2 game. Ten game types were constructed by matching human participants to different subsets of computer strategies. A Latin square design was used to counterbalance the order of game types. All participants played ten games, with each participant being assigned to one of ten ordering conditions as determined by the Latin square design. Each game was played for 50 rounds, thus each participant played a total of 500 rounds. No information was conveyed regarding the number of rounds per game, the identity of the other players, and the strategies the other players used. Participants were asked to try to maximize their payoffs. Due to space limitations we do not present more details of this study here. For more details on the human study see [12].

The model was able to learn the game and achieve human levels of performance in terms of payoff and power (Figure 1). Note that the learning curve for payoff is different than the typical learning curve found in single-agent non-interactive tasks where a steep increase in performance is followed by a plateau. The learning curve found here starts with a decline and ends with a steep increase in the second half of the game. At the beginning of the game, the model and presumably the human participants do not have enough instances in memory to identify and carry out a profitable course of action. As the game progresses, more instances are accumulated, effective ones become more active, and ineffective ones are gradually forgotten.
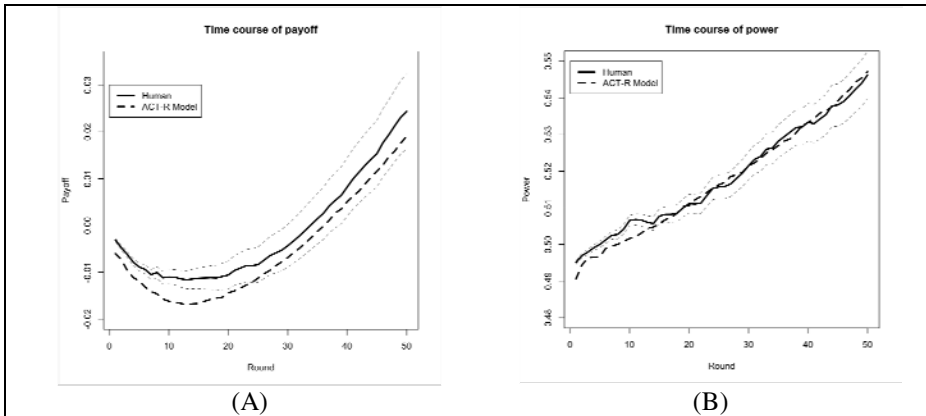
Fig. 1. Time course of payoff (A) and power (B). The dotted-line interval around the human data represents standard error of the mean.
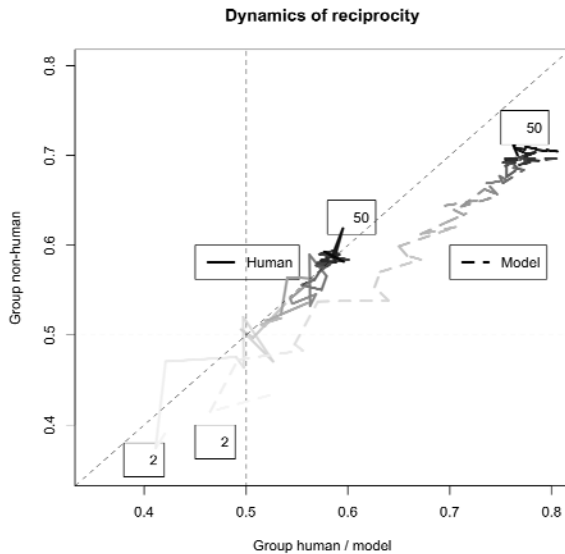
Fig. 2. The dynamics of reciprocity between the two groups in the IPD^2 game averaged across all game types and participants (humans or model simulations). The solid line represents the human data and the dashed line model simulations.

The model and the human participants chose to cooperate more often than to defect and cooperate more when they were in the majority than in the minority. However, the model shows overall much higher levels of cooperation than the human participants. More interestingly, the model also caused an increase in the level of cooperation in the opposing group as compared to the human participants.

Figure 2 shows the dynamics of reciprocity between the two groups playing the IPD^2 game. The amount of cooperation in the group including a human participant

(or model) is represented on the X-axis. The amount of cooperation in the opposing group (comprised of two computer strategies) is represented on the Y-axis. The horizontal and vertical dashed lines mark an equal amount of cooperation and defection in each group (i.e., proportion of cooperation = 0.5). The diagonal dashed line is the reciprocity line where the two groups have equal proportions of cooperation. The gray spectrum marks the progression of a game. White denotes the start (round 2, marked with number 2) and black the end (round 50, marked with number 50) of the game. The group including a human participant achieves perfect reciprocity with the opposing group (on average). In contrast, the model systematically departs from the reciprocity line (the diagonal in Figure 2). This bias of the model brings about a higher level of cooperation in the opposing group as well.

5   Discussion

We have introduced IPD^2 – a new game paradigm to study the impact of intragroup power on intergroup conflict and cooperation. We argue that IPD^2 offers a richer opportunity than the classical Iterated Prisoner's Dilemma for modeling real-world situations in which intra- and inter-group interactions are entangled [7]. In IPD^2, mutual defection results in negative payoffs that over time cause changes in the power structure of the two groups, thus creating the chance to break free from that pattern and possibly evolve toward mutual cooperation. This is analogous to the political sphere where there is competition for power and leaders who bring negative outcomes to their groups can only maintain their position of power for a limited time.

The pattern of cooperation in IPD^2 oscillates around the line of perfect reciprocity (Figure 2). This result corroborates what has been found in other studies regarding the human tendency to reciprocate. The human brain is specialized at detecting and discriminating against non-reciprocators. Negative emotions triggered by unreciprocated attempts at cooperation bias subsequent decision making toward defection [13]. In our study, human participants tend to stay on the line of perfect reciprocity (on average); they do not "walk the extra mile" that would increase the overall mutual cooperation in the game. Most likely, they fear that their attempts at cooperation will not be reciprocated. In contrast, the ACT-R model does not capture this fear, which allows it to continue cooperating even when its cooperation is temporarily unreciprocated. As a consequence of this sustained cooperation, the model causes the level of mutual cooperation in the game to increase. This is merely a prediction of our model, because we do not yet know how the model would behave if it were interacting with humans. We are currently exploring ways to improve the model's fit to the human data by adding an emotional component to the existing cognitive mechanisms of ACT-R.

References

1. Erev, I., Roth, A.E.: Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. Am. Econ. Rev. 88(4), 848–881 (1998)
2. Lebiere, C., Wallach, D., West, R.L.: A Memory-Based Account of the Prisoner's Dilemma and Other 2x2 Games. In: Proceedings of the 3rd International Conference on Cognitive Modeling, Groningen, Netherlands, pp. 185–193 (March 2000)
3. Rapoport, A., Guyer, M.J., Gordon, D.G.: The 2X2 Game. The University of Michigan Press, Ann Arbor (1976)
4. Anderson, J.R., Lebiere, C.: The Atomic Components of Thought. Erlbaum, Mahwah (1998)
5. Rapoport, A.: Game theory as a theory of conflict resolution. D. Reidel Publishing Company, Boston (1974)
6. Camerer, C.F.: Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton (2003)
7. Putnam, R.D.: Diplomacy and Domestic Politics: The Logic of Two-Level Games. Int. Organ 42(3), 427–460 (1988)
8. Bornstein, G.: Intergroup Conflict: Individual, Group and Collective interests. Pers. Soc. Psychol. Rev. 7, 129–145 (2003)
9. Halevy, N., Bornstein, G., Sagiv, L.: "In-Group Love" and "Out-Group Hate" as Motives for Individual Participation in Intergroup Conflict. Psychol. Sci. 19(4), 405–411 (2008)
10. Dowding, K.: Power (Concepts in Social Thought). University of Minnesota Press, Minneapolis (1996)
11. Gonzalez, C., Lerch, J.F., Lebiere, C.: Instance-Based Learning in Dynamic Decision Making. Cogn. Sci. 27, 591–635 (2003)
12. Juvina, I., Lebiere, C., Martin, J., Gonzalez, C.: Intergroup Prisoner's Dilemma with Intragroup Power Dynamics. Games (in press)
13. Rilling, J.K., Goldsmith, D.R., Glenn, A.L., Jairam, M.R., Elfenbein, H.A., Dagenais, J.E., et al.: The Neural Correlates of the Affective Response to Unreciprocated Cooperation. Neuropsychologia 46, 1256–1266 (2008)

Predicting Demonstrations' Violence Level Using Qualitative Reasoning

Natalie Fridman, Tomer Zilberstein, and Gal A. Kaminka⋆

The MAVERICK Group
Computer Science Department
Bar Ilan University, Israel

Abstract. In this paper we describe a method for modeling social behavior of large groups, and apply it to the problem of predicting potential violence during demonstrations. We use qualitative reasoning techniques which to our knowledge have never been applied to modeling crowd behaviors, nor in particular to demonstrations. Such modeling may not only contribute to the police decision making process, but can also provide a great opportunity to test existing theories in social science. We incrementally present and compare three qualitative models, based on social science theories. The results show that while two of these models fail to predict the outcomes of real-world events reported and analyzed in the literature, one model is successful. We believe that this demonstrates the efficacy of qualitative reasoning in the development and testing of social sciences theories.

Keywords: Demonstrations, Social Simulation, Qualitative reasoning.

1   Introduction

A violent demonstration, resulting in casualties among its participants, police forces and innocent bystanders, is unfortunately not a rare phenomena. This paper deals with improving the police decision making process, by providing useful predictions as to the potential outcomes of demonstrations, given the specific settings. The hope is to decrease the number of casualties by preventing violence.

In general, there are several technologies that can be used to generate predictions. Agent based simulations [7] require detailed individual cognitive modeling, and furthermore, modeling at the individual participant level is too fine a resolution for useful predictions. Numerical simulation [12] models at an appropriate resolution (global group behavior), but unfortunately requires complete and precise domain information, which is not available here. There exists significant literature on the factors that impact violence during demonstrations, but it mostly reports on partial, macro-level qualitative descriptions of the influencing factors.

In this paper we describe a novel application of Qualitative Reasoning (QR) [10,3] to modeling and reasoning about potential violence level during demonstrations. QR is a sub-area of AI, which enables reasoning with partial or imprecise numeric information. Using QR, it is possible to draw useful conclusions even with only qualitative representation of data and order values (such as little/medium/large). Thus such modeling

provides an opportunity to test existing social science theories regarding the influencing factors on the violence level during the demonstrations.

Based on social science research, which provides qualitative information regarding the factors influencing the violence level in demonstrations, we incrementally present and compare three qualitative models of demonstrations. The first two models are based on an extensive research report initiated by Israeli police [2]. The third is our extension of the second model based on sociological consultation. We evaluated the models on four real-life scenarios. The results show that the first two models make incorrect predictions, but the BIU model makes good predictions on the examined test cases.

2   Related Work

Usage of computer simulation is considered to be a leading approach for modeling and reasoning regarding different social phenomena [6]. There are several micro and macro level techniques that enable such modeling, e.g., usage of agent based simulation, cellular automata and system dynamics.

Agent-based simulation is a micro-level approach where by social behaviors are simulated by simulating each individual, and their interactions. By applying agents as an "intelligent" entity we have the ability to model complicated social interactions. Such simulations have been successfully used in modeling crowd behaviors [5,7], economic phenomena [16], and more. However, it is a bottom-up approach in the sense that to receive a macro-level behavior we must model the micro-level interactions which necessitates detailed individual modeling, and when number of agents is scaled up it may provide significant computational barriers. Furthermore, there are domains such as predicting the likelihood of violence that modeling at the individual participant level may be too high a resolution and even unnecessary.

System dynamics approach [6] is a macro level approach in the sense that it models an entire system. It uses defined stocks, flows and feedback loops to model system behavior. The models are basically sets of differential equations that describe changes in the system. In our domain, such accurate and full definitions are not available.

Qualitative Reasoning (QR) is another macro level approach, allowing modeling and reasoning with partial and imprecise information. It has been used to allow for commonsense reasoning in physics [10,3]. However, it has also been applied to other branches of science: ecology [13], social science [8], politics [4] etc. However, our use of QR to model and predict the violence level during demonstrations is novel.

Fuzzy Cognitive Maps (FCM) [9] is also a macro level approach which enables causal reasoning using fuzzy directed graphs. Similarly to QR, FCM enables imprecise and qualitative representation of the model. However, the output of FCM is a recommendation on a single action or goal, while QR returns the set of all possible behaviors that the model may manifest.

3   Qualitative Reasoning and Simulation

Qualitative simulation enables reasoning about possible system behaviors that can emerge from an initial world state. The simulation takes as input the initial state of

the world which contains a structural description of the model and produces a state transition graph. A final state graph captures the set of all possible behaviors that the model may manifest. It consists of a set of states and the transitions between them (state-transitions). Each state is a possible unique behavior that the model develops, it contains a unique set of values and inequality statements (quantities) which describe the current behavior of the system. State transitions transform one state into another, by specifying the changes in values and in inequality statements. Each state may contain multiple transitions which enables multiple possible developments of the current state. A sequence of states connected by state transitions where each state is the immediate successor of the one before, is called a behavior path.

Each state is composed of a set of quantities. Quantity is the lowest resolution representation for continuous parameters and it is composed of a pair of values: magnitude and derivative. The magnitude represents the amount of quantity and derivative represents the direction of change. The set of possible values is described by Quantity Space (QS) which is a finite and ordered set of landmark values. Changes in the system are defined explicitly by causal relationships. There are two types of casual relationship between quantities, direct $(I+, I-)$ and indirect $(P+, P-)$ influence. Each influence may be positive $(I+, P+)$ or negative $(I-, P-)$ meaning the derivative of the target quantity increases or decreases accordingly.

In each cycle and on each quantity, all influences (direct and indirect) are combined. When positive and negative influences are combined ambiguities may occur. The qualitative simulation considers all the possible combinations thus, when qualitative description is incomplete, it provides a non deterministic prediction.

4   Modeling Violence Level in Demonstration

Knowledge regarding demonstrations is not accurate nor complete. There are many micro-theories in social science regarding the influencing factors on the violence level: Each such theory focuses on a small sub-set of factors. Integrating all of them into a single unified model is real challenge. The Israeli police initiated a comprehensive study to address this challenge, resulting in a report [2] that provides a collection of factors and their influence on the violence level and also on each other. Their goal was to classify and analyze different kinds of demonstrations in order to propose appropriate methods for the police force in dealing with the mass. They studied 102 crowd events (in particular demonstrations) during the years 2000–2003 and interviews with 87 policemen and police officers. They analyzed a variety of factors that may affect violent behavior, as well as relevant literature. This report is a qualitative collection of factors which provide a challenge to the reasoning process. We use this report as a source of knowledge based on which we developed our models and by using qualitative simulation we provide an ability for reasoning regarding potential violence level.

Base Model. The first (*Base*) model was developed based on the report's literature review [2]. It was proposed there as a first attempt at building a baseline, purely based on literature review. According to the Base model the most influential factors on the violence level during the demonstration are (1) the crowd's a-priori hostility towards

the police; (2) willingness to pay the personal price (such as willingness to be arrested); (3) low chance for punishment for violent actions (e.g., a belief that police will not respond strongly); (4) group cohesiveness; (5) previous history of violence. All of these increases the level of violence.

Police Model. The second model is an extension of the Base model. Karmeli and Ravid-Yamin [2] significantly expanded the Base model, based on their interviews with police officers and their investigation into 102 demonstrations. In addition to the factors from the Base model, the Police model adds 12 more variables, roughly divided into several groups. *Environmental factors* include weather, time of day, location sensitivity (e.g., for religious reasons), and time of year sensitivity (e.g., Christmas). *Participant factors* include the number of participants, the existence of violent core among the participants, the existence of group leader, and the cohesiveness of the group (e.g., if they all come from a particular ethnic minority). *Procedural factors* include a request for demonstration license, the purpose of the event (emotional or rational), the timing and strength of police intervention.

The research results showed significant relations between these variables and also their impact on the event outcome (the violence level). For example, political or social demonstrations that express protest or support for leader or cause usually end with low level of violence. However, demonstrations with nationalistic flavor that intend to express emotions (letting off steam) are characterized by much more violent outcomes. The research results also showed a relationship between existence of license and united identity: it was found that some united identities tend to apply for a license before the protest while others do not. It was also found that the time of the day has impact on the violence level; more violent demonstrations occur at night than during the day [2]. A graphic representation of the qualitative model is presented in Figure 1. It shows three entities (Population, Nature and Police) and 18 quantities: 6 are of the Base model, and additional 12 listed above.
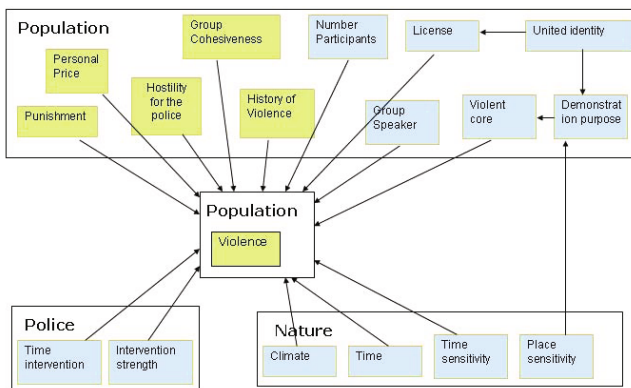
Fig. 1. Police Model: Structure

BIU Model. The third model is our own novel extension of the Police model. Based on interviews with social and cognitive scientists, as well as additional literature surveys, we added four additional variables, and updated 19 influences (relations) among the variables. The added factors are: (1) maintenance of order; (2) participants anonymity (indicates whether the participants believed they could be recognized); (3) participants' visual cohesiveness (such as similarly-colored clothes among football fans); and (4) the presence of light. The resulting QR model is shown in Figure 2.

We provide here several examples for updated influences. First, we updated the influence of police's intervention strength, thus instead of direct impact on violence level as in the Police model, it impacts the participants' belief that they may be punished, and their hostility for the police. In BIU model, high intervention strength increases participants' hostility for the police and increases the participants' chance for punishment. However, low intervention strength just decreases the participants' chance for punishment without a change in hostility for the police factor. Another example is that existence of group speaker and existence of license increase the maintenance of order, which decreases the violence level. In contrast, in the Police model, license and group speaker variables had a direct influence on the violence level. Moreover, for the variable *number participants*, we no longer allow direct influence on the violence level as in Police model, but instead have it influence the participants' anonymity level ("the more participants around me the less recognizable I am"). Another example of addition to the BIU model is: participants visual cohesiveness has an impact on group cohesiveness, it actually increases the sense of belonging to the same group.
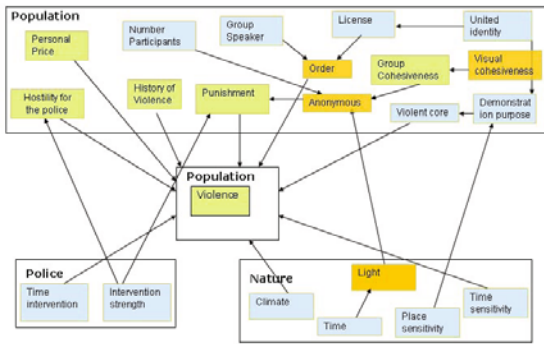
Fig. 2. BIU Model: Structure

5   Models Evaluation

We compare these models on four real-life scenarios. These are cases for which we had a detailed description along with an analysis [2]. The first three are well known events which where extensively analyzed and described [2,11,15,14], and ended with tragic outcomes. The last event was a calm one-hour protest (which we video-taped and

analyzed), involving about 100 people, which ended without any confrontation. The results show that the Base model and Police model failed to provide correct prediction on one or more of the test-cases, while the BIU model provides good results.

The first event is the Heysel Stadium Disaster which occurred in 1985 [11]. It was the 1985 European cup final, Liverpool vs. Juventus which was a very tragic and violent event with many casualties. According to Lewis [11] who analyzed this event, one of the reasons for this violent outcome is the police's lack of intervention to prevent the developing violence.

The second event is the Los Angeles Riots which occurred in 1991. This was also a very violent event with many casualties, with 55 killed people and over 2000 injured. Useem [15] who analyzed this event, argued that the police were not properly organized and did not react on time with appropriate force to prevent the eruption, but for the six hours from the beginning of the event, police did little to prevent it which allowed the violent core to grow.

The third event is the London Riot Disaster which occurred in 1990 [14]. As opposed to the previous two events, here the police used enormous force against the protests without distinguishing between anarchists and peaceful marchers. The marchers, with nowhere to go to escape had to fight back. What started as a peaceful protest turned to a very violent event with many casualties.

The last event is a Petach Tikva protest occurred in 2009. In 2009, several children from the Ethiopian community in Petach Tikva were not assigned to schools by time summer vacation was over. This was due, according to some opinions, to racial prejudice against these children. Consequently, approximately 100 people got together in the city square to protest. The incident began calmly and also ended without any violence.

For the evaluation of the QR models, we implemented the models in GARP, a QR engine which enables building and simulating qualitative models and was successfully used in many domains [13,1]. GARP takes as input an initial setting of the world state (partial state information is acceptable) and produces a simulation state-transition graph. Each sequence of states, following transitions from the initial state and ending with a different outcome state is a possible system trajectory—a possible sequence of qualitative state changes that may occur given the initial state, and the qualitative dynamics specified. The end state in of each such path is where the system dynamics allow no further evolution (i.e., the system is stable). Taking the value of the outcome variables (in our case, violence level) in these final states allow categorical predictions.

However, it is not enough to know whether a demonstration might be violent; in a sufficiently complex model, all three possible values will have at least one state transition path leading to them. Instead, our goal is also estimate the likelihood of different outcomes. Such knowledge may provide a sufficient addition to the decision making process of the police force. To do this, we use the received state-graph as an input and based on this developed graph we calculate the likelihoods of different outcomes as follows: we count the number of behavior paths that lead to a specific violence level and divide it by the total number of paths.

To initialize the test cases, we utilized the information appearing in their descriptions in the literature. We initialized only the quantities for which we had explicit information; qualitative simulation can work with such partial information.

Results. First we want to explore whether the models correctly predict the results of the test cases. For the Heysel Stadium Disaster, the Base model and the Police models predict 100% high violence and 0% low and no violence, and the BIU model predict 96% high violence, 3% low violence and 1% no violence. For the Los Angeles Riots, the Base model and the Police models predict similar outcomes: 66% high violence and 34% no violence. The BIU model predicts for the same event 99% high violence and 1% low violence. For the London Riot Disaster Base model predict 66% high violence and 34% no violence, the Police model predicts 80% high violence and 20% no violence. The BIU model for the same event predicts 57% high violence, 30% low violence and 13% no violence. For the Petach Tikva Protest which was a calm demonstration, in contrast to the other events, the Base model predicts 100% of no violence, Police model predict **66%** high violence and 34% no violence, and the BIU model predicts 5% high violence, 78% low violence and 17% of no violence.

The results demonstrate that Base model and BIU model made a successful prediction in all examined test cases. However, The Police model provide a poor results in prediction of Petach Tikva Protest where the demonstration as it occurred in real life was a peaceful protest while Police model predicts 66% of high violence outcome.

In the following experiment we want to demonstrate the use of QR for hypothetical changes to the police intervention strength. Lewis [11] who analyzed the Heysel stadium disaster concluded that police did too little to prevent the rioting. Similar conclusion was also presented by Useem [15] who analyzed the LA riots. However, Stott and Drury [14], who analyzed the London riots, concluded that police used too much force, and that this was one factors in the tragical outcome. Based on these conclusions, if the police would act differently the events could end differently. Thus, we want to examine the presented model's prediction in *what if* scenarios.

Table 1 presents the experiment results. As before the first column corresponds to the examined test case. The second column corresponds to recommended police intervention strength. Then we present the models predictions for each possible outcome: no violence, low violence and high violence. Below of each experiment, we presented whether the recommended reaction changed the model's prediction.

Table 1. Experiments results: changed police intervention strength

Exp.	Recommended Change	Model Outcome	Basic Model	Police Model	BIU Model
Exp1	Increase strength [11]	High violence	100%	66%	83%
		Low violence	0	0	6%
		No violence	0	34%	10%
	Change/No-Change		No-Change	**Change**	**Change**
Exp2	Increase strength [15]	High violence	66%	66%	87%
		Low violence	0	0	3%
		No violence	34%	34%	10%
	Change/No-Change		No-Change	No-Change	**Change**
Exp3	Decrease strength [14]	High violence	66%	80%	19%
		Low violence	0	0	45%
		No violence	34%	20%	36%
	Change/No-Change		No-Change	No-Change	**Change**

The results demonstrate that Basic model and Police model failed in providing correct prediction while the BIU model provided good results. The failure of the Base model is not surprising, since the Base model not accounts for the factor of police intervention strength therefore there are no change in the model's predictions.

6   Summary and Future Work

In this paper we described a method for modeling and reasoning about social behavior of large groups, and applied it to the problem of predicting potential violence during demonstrations. We used qualitative reasoning (QR) techniques, which to our knowledge have never been applied for modeling crowd behaviors, nor in particular demonstrations. Based on social science research, we incrementally presented and compared three QR models for predicting the level of violence in demonstrations: A Base model, Police model and BIU model. We evaluated these models on four real life test cases scenarios. The results show that BIU model makes good predictions on the examined test cases even where the others fail.

In our future work we plan to expand our model to account for bidirectional influences (feedback loops). For example, in the BIU model the "hostility for the police" quantity increases the violence level. However, increasing the violence level has no impact on hostility. We believe that such expansion is necessary to provide a more accurate prediction. Also, in our future work, we plan to provide a statistical analysis of the developed state-graph and enables reasoning regarding the developed process and not only regarding the final outcome. The third direction is to expand our evaluation process by test our model on additional real-life test cases.

References

1. Bredeweg, B., Salles, P.: Mediating conceptual knowledge using qualitative reasoning. In: Jrgensen, S., Chon, T.-S., Recknagel, F.E. (eds.) Handbook of Ecological Modelling and Informatics, pp. 351–398. Wit Press, Southampton (2009)
2. Carmeli, A., Ravid-Yamin, I.: Research report on the subject of crowd events and public order. Technical report, Ministry of public security, Bureau of the chief scientist (2006)
3. Forbus, K.D.: Qualitative reasoning. In: CRC Handbook of Computer Science and Eng. CRC Press, Boca Raton (1996)
4. Forbus, K.D., Kuehne, S.E.: Towards a qualitative model of everyday political reasoning. In: Proceedings of the Nineteenth International Qualitative Reasoning Workshop (2005)
5. Fridman, N., Kaminka, G.A.: Towards a cognitive model of crowd behavior based on social comparison theory. In: AAAI 2007 (2007)
6. Gilbert, N., Troitzsch, K.G.: Simulation for the social scientist. Open University Press, Stony Stratford (2005)
7. Jager, W., Popping, R., van de Sande, H.: Clustering and fighting in two-party crowds: simulating the approach-avoidance conflict. Journal of Artificial Societies and Social Simulation 4(3) (2001)
8. Kamps, J., Péli, G.: Qualitative reasoning beyond the physics domain: The density dependence theory of organizational ecology. In: Proceedings of QR 1995 (1995)
9. Kosko, B.: Fuzzy cognitive maps: advances in theory, methodologies, tools and applications (studies in fuzziness and soft computing). Springer, Heidelberg (2010)

10. Kuipers, B.: Qualitative reasoning: modeling and simulation with incomplete knowledge. Automatica 25(4) (1989)
11. Lewis, J.M.: A value-added analysis of the heysel stadium soccer riot. Current Psychology (1989)
12. Patrick, S., Dorman, P.M., Marsh, R.L.: Simulating correctional disturbances: the application of organization control theory to correctional organizations via computer simulation. Journal of Artificial Societies and Social Simulation 2(1) (1999)
13. Salles, P., Bredeweg, B.: Modelling population and community dynamics with qualitative reasoning. Ecological Modelling 195, 114–128 (2006)
14. Stott, C., Drury, J.: Crowds, context and identity: Dynamic categorization process in the poll tax riot. Human Relations (2000)
15. Useem, B.: The state and collective disorders: The los angeles riot/protest of april 1992. Social Forces 76 (1997)
16. Wander, J.: Modelling consumer behavior. PhD thesis, University of Groningen (2000)

Abstraction of an Affective-Cognitive Decision Making Model Based on Simulated Behaviour and Perception Chains

Alexei Sharpanskykh and Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{sharp,treur}@few.vu.nl

Abstract. Employing rich internal agent models of actors in large-scale socio-technical systems often results in scalability issues. The problem addressed in this paper is how to improve computational properties of a complex internal agent model, while preserving its behavioral properties. The problem is addressed for the case of an existing affective-cognitive decision making model instantiated for an emergency scenario. For this internal decision model an abstracted behavioral agent model is obtained, which ensures a substantial increase of the computational efficiency at the cost of approximately 1% behavioural error. The abstraction technique used can be applied to a wide range of internal agent models with loops, for example, involving mutual affective-cognitive interactions.

Keywords: model abstraction, agent systems, cognitive modeling, decision making.

1 Introduction

Large-scale socio-technical systems are characterized by high structural and behavioral complexities. Modeling and analysis of such systems is a challenging task. On the one hand, models of such systems should account for complexity of dynamics of humans and technical systems. On the other hand, the models should ensure acceptable computational efficiency. Models for large-scale socio-technical systems have been developed in several areas. In the area of Social Physics socio-technical systems are often modeled from a lattice gas perspective by representing the system's actors by particles interacting through forces and fields [5]. Although such models are highly scalable, they ignore (complex) internal dynamics underlying the decision making of actors, and, thus, cannot be used in cases for which rich cognitive and affective representations are required (e.g., reasoning, human decision making). System Dynamics models of socio-technical systems abstract from single events, entities and actors and take an aggregate view on the system dynamics (e.g., [9]). On the one hand, such models are highly computationally efficient. On the other hand, it may be difficult to map the actual system structures and processes to the abstract aggregated model variables. By doing so, the link to the behavior of actors and their interactions is lost, so that the level of analysis is reduced. Agent-based modeling

approaches take into account the local perspective of separate actors and their specific behavior and interactions, and models them as interacting agents in a multi-agent system. To ensure computational efficiency on a large scale, models of agents are often kept overly simple. The plausibility of such models has been often criticized [4, 8]. The generic agent-based affective-cognitive decision making model from [13] is among few exceptions with rich cognitive representations based on findings from Cognitive Science, Neurology, and Social Science. However, because of its high complexity, this agent model has limited scalability. The problem addressed in this paper is how to improve the computational properties of the agent model from [13], while preserving its behavioral properties. This is illustrated in this paper for a variant of the model instantiated for an emergency case. To address this problem, a loop abstraction technique from [14] has been used. The abstracted agent model obtained ensures a more than twice increase of the computational efficiency of large-scale multi-agent systems based on the original model. This is achieved at the cost of 1% behavioural error, which is, however, not critical for most applications.

To specify dynamic properties of a system, the order-sorted predicate logic-based language called LEADSTO is used [1]. Dynamics in LEADSTO is represented as evolution of states over time. A state is characterised by a set of properties that do or do not hold at a certain point in time. To specify state properties for system components, ontologies are used which are defined by a number of sorts, sorted constants, variables, functions and predicates. For a given ontology Ont, state properties are specified as propositions that can be made in the form of conjunctions from (negations of) ground atoms. LEADSTO enables modeling of direct temporal dependencies between two state properties in successive states, also called *dynamic properties*. The format is defined as follows. Let α and β be state properties of the form 'conjunction of atoms or negations of atoms', and e, f, g, h non-negative real numbers indicating timing parameters. In the LEADSTO language the notation $\alpha \twoheadrightarrow_{e, f, g, h} \beta$ means: if state property α holds for a certain time interval with duration g, then after some delay (between e and f) state property β will hold for a certain time interval of length h. When the timing parameters are chosen in a uniform manner, it is written $\alpha \twoheadrightarrow \beta$. To indicate the type of a state property sometimes prefixes input(A), internal(A) and output(A), will be used, where A is the name of an agent. Consider an example dynamic property with e = f = 0 and g = h = Δt:

   input(A)|observation_result(fire) $\twoheadrightarrow$ output(A)| performed(runs_away_from_fire)

Informally, this example expresses that if agent A observes fire at t, lasting until t+Δt, then at t+Δt A will run away from the fire.

The paper is organized as follows. In Section 2 the model from [13] instantiated for an emergency scenario is presented. The model abstraction is considered in Section 3. In Section 4 behavioral and computational properties of the abstracted model obtained are considered. Section 5 concludes the paper.

2 An Affective-Cognitive Decision Making Model

In this section the affective-cognitive decision making model from [13] is introduced and illustrated as instantiated for an emergency scenario. In the scenario a group of

agents considers two options (paths) to move out of a burning building. An option is a (partially) ordered sequence of actions (i.e., a plan) to satisfy the agent's goals. The goal of each agent in the scenario is to be outside of the building. Options are represented internally in agents using the neurological theory of simulated behaviour and perception chains proposed by Hesslow [6], as shown in [13]. By a simulated chain the agent simulates mentally the complete sequence of preparations of actions constituting an option, including perception of their effects. In the scenario an option is specified by a sequence of (intermediate) locations to be reached with an exit as the last location.

The simulated sensory representation states relate to emotions, which provide either positive or negative reinforcement of the simulated actions, and thus guide the agent's decision making process. By evaluating sensory consequences of actions in simulated behavioural chains using cognitive structures from the OCC model [10], different types of emotions can be distinguished. In the example two types of emotions - fear and hope – are distinguished, which are often considered in the emergency domain. According to [10], the intensity of fear induced by an event depends on the degree to which the event is undesirable and on the likelihood of the event. The intensity of hope induced by an event depends on the degree to which the event is desirable and on the likelihood of the event. Thus, both emotions are generated based on the evaluation of a distance between the effect states for the actions from an option and the agent's goal state. In the evacuation case the goal is to be outside of the building, thus the evaluation functions for both emotions can be defined over the distance between the agent's location and the nearest reachable exit. In particular, the evaluation functions were specified as follows.

for hope: $\text{eval}_1(g, \text{effect}(a)) = e^{-\omega|\text{effect}(a) - g|}$;
for fear: $\text{eval}_2(g, \text{effect}(a)) = 1 - e^{-\omega | \text{effect}(a) - g |}$,

where effect(a) is the agent's location after the execution of action a; g is the aimed location (i.e., the nearest exit); ω is a parameter.

In literature [2,3] it is recognized that emotions emerge and develop in dynamics of reciprocal relations between cognitive and body states of a human. These relations, omitted in the OCC model, are modelled from a neurological perspective using Damasio's principles of 'as-if body loops' and somatic marking [2,3]. Formally, the 'as if body loop' for hope for option o is specified for each agent A_i as follows:

srs(g, V1) & srs(effect(a), V2) & hope(o, V3) &
connection_between_strength(preparation_for(a), srs(effect(a)), V4) &   (1)
→ srs(eval_for(effect(a), bhope), V4*eval$_t$(V1, V2))

$\wedge_{i=1..n}$ srs(eval_for(effect(a$_i$), bhope), Z$_i$) & hope(o, U) & preparation_for(bhope, W)   (2)
→ preparation_for(bhope, (W+γ(h(f(Z$_1$, ..., Z$_n$), U) - W)Δt),

where f(Z$_1$, ..., Z$_n$) is the arithmetic mean function, and
h(V1, V2) = β (1-(1- V1)(1- V2)) + (1-β) V1 V2.

preparation_for(bhope, V) → srs(bhope, V)   (3)

srs(bhope, V2) & srs(G(bhope), V1) & hope(o, V3)   (4)
→ hope(o, V3+ γ1(g(V1, V2) – V3)Δt)

where, for agents with an overestimation bias $\beta \geq 0$:

$$g(V1, V2) = \alpha V1 + (1-\alpha)V2 + \beta(1-\alpha V1-(1-\alpha)V2) \tag{5}$$

and for agents with an underestimation bias $\beta < 0$:

$$g(V1, V2) = \alpha V1 + (1-\alpha)V2 + \beta(\alpha V1+(1-\alpha)V2) \tag{6}$$

Here α indicates the importance of the group's opinion (or influence) for the agent; G(bhope) is the aggregated group preparation to the emotional response (body state) for hope. The influence of the group on the individual decision making is modelled based on *the mirroring function* of preparation neurons in humans; cf. [7], [11]. This mirroring function in social decision making is realised in two forms: (1) by *mirroring of emotions*, which indicates how emotional responses in different agents about a decision option mutually affect each other, and (2) by *mirroring of intentions* or *action preparations* of individuals for a decision option.

It is assumed that the preparation states of an agent for the actions constituting options and for emotional responses for the options are expressed in body states that can be observed by other agents from the group. The contagion strength of the interaction from agent A_2 to agent A_1 for a preparation state p is defined as follows:

$$\gamma_{pA_2A_1} = \varepsilon_{pA_2} \cdot \alpha_{pA_2A_1} \cdot \delta_{pA_1} \tag{7}$$

Here ε_{pA_2} is the personal characteristic expressiveness of the sender (agent A_2) for p, δ_{pA_1} is the personal characteristic openness of the receiver (agent A_1) for p, and $\alpha_{pA_2A_1}$ is the interaction characteristic channel strength for p from sender A_2 to receiver A_1.

By aggregating such input, an agent A_i perceives the group's joint evaluation of each option, which comprises the following dynamic properties.

(a) The aggregated group preparation to (i.e., the externally observable intention to perform) each action p constituting the option. Formally:

$$\wedge_{j\neq i} \text{ internal}(A_j)|\text{preparation\_for}(p, V_j) \rightarrow \text{internal}(A_i) \mid \text{srs}(G(p), \Sigma_{j\neq i} \ \gamma_{pA_jA_i} V_j / \Sigma_{j\neq i} \ \gamma_{pA_jA_i}) \tag{8}$$

(b) The aggregated group preparation to an emotional response (body state) be for each option. For each emotional response be a separate preparation state is introduced. Formally:

$$\wedge_{j\neq i} \text{ internal}(A_j)|\text{preparation\_for}(be, V_j) \rightarrow \text{internal}(A_i) \mid \text{srs}(G(be), \Sigma_{j\neq l} \ \gamma_{beA_jA_i} V_j / \Sigma_{j\neq i} \ \gamma_{beA_jA_i}) \tag{9}$$

According to the Somatic Marker Hypothesis [3], each represented decision option induces a feeling which is used to mark the option. For example, a strongly positive somatic marker linked to a particular option occurs as a strongly positive feeling for that option. Through these connections emotions influence the agent's readiness to choose the option. From a neurological perspective, the impact of a sensory representation state to an action preparation state via the connection between them in a behavioural chain will depend on how the consequences of the action are felt emotionally. Thus, the preparation state for the first action from an option is affected by the sensory representations of the option, of the perceived group preparation for the action and of the emotion felt towards the option. Formally, for the emergency example:

srs(o, V1) & hope(o, V2) & fear(o, V3) & srs(G(a1), V4) & preparation_for(a1, V5)
→ preparation_for(a1, V5 + γ2(h(V1, V2, 1-V3, V4) – V5)Δt),   (10)

where h(V1,V2,1-V3,V4) = β1(1-(1-V1)(1-V2)V3(1-V4)) + (1-β1) V1 V2 (1-V3) V4.

Preparation states for subsequent actions a in the behavioural chain are specified by:

srs(effect(a), V1) & hope(o, V2) & fear(o, V3) & srs(G(a), V4) & preparation_for(a, V5)
→ preparation_for(a, V5 + γ2(h(V1,V2,1-V3,V4) – V5) Δt)   (11)

3   Abstraction of the Internal Agent Model

In this section the process of abstraction of the affective-cognitive internal agent model from Section 2 is described, making use of the procedure from [14], which allows elimination of the 'as if body loops' from the model. To apply this loop elimination procedure, the following representation of a loop is assumed:

$$\text{has\_value}(u, V_1) \wedge \text{has\_value}(p, V_2) \rightarrow \text{has\_value}(p, V_2 + e(V_1, V_2) \Delta t) \qquad (12)$$

Here u is the name of an input variable, p of the loop variable, and $e(V_1, V_2)$ is a function combining the input value with the current value for p. Such a representation of an 'as if loop' can be obtained from (2)-(4):

$$\wedge_{i=1..n} \text{srs(eval\_for(effect}(a_i), \text{bhope}), Z_i) \text{ \& preparation\_for(bhope, W) \&}$$
$$\text{srs(G(bhope), V1) \& hope(o, U)} \rightarrow_{3,3,1,1} \text{hope(o, U+ } \gamma1(g(V1,V2) - U)\Delta t), \qquad (13)$$

where V2 = (W+γ(h(f(Z_1, …, Z_n), U) - W) Δt

Here srs(eval_for(effect(a_i), bhope), V_i) and srs(G(bhope), V1) are inputs to the loop.

Loop abstraction is based on identifying dependencies of equilibrium states for loops. An equilibrium state for an input value V_1 in (12) is a value V_2 for p such that $e(V_1, V_2) = 0$. A specification of how V_2 depends on V_1 is a function g s.t. $e(V_1, g(V_1)) = 0$. When a specification of g is obtained, (12) can be transformed into:

$$\text{has\_value}(u, V1) \rightarrow_{D,D,1,1} \text{has\_value}(p, g(V1)) \qquad (14)$$

where D is chosen as a timing parameter for the process of approximating the equilibrium value up to some accuracy level. In an equilibrium state for the loop for hope (13) U=V3, W=V2, and

$$W = h(f(Z_1, …, Z_n), U) \qquad\qquad U = g(V1, h(f(Z_1, …, Z_n), U))$$

Thus, for an overestimating agent with g(V1, V2) as defined in (5):

$$U = \alpha V1 + (1-\alpha)V2 + \beta(1-\alpha V1 - (1-\alpha)V2),$$

where V2 = β1 (1-(1- f(Z_1, …, Z_n))(1- U)) + (1-β1) f(Z_1, …, Z_n) U.

By rearranging terms:

$$U = \beta + \alpha(1-\beta)V1 + (1-\alpha)(1-\beta)V2, \qquad \text{where } V2 = \beta1 * f(Z_1, …, Z_n) + (\beta1 + (1-2\beta1)f(Z_1, …, Z_n))U$$

Thus, in the stable state for the loop the value of the hope state depends only on the loop's input states:

$$U = (\beta + \alpha(1-\beta)V1 + (1-\alpha)(1-\beta)*\beta1*f(Z_1, ..., Z_n))/ (1- (1-\alpha)(1-\beta)(\beta1+ (1-2\beta1)f(Z_1, ..., Z_n)) \qquad (15)$$

Similarly, for an underestimating agent with g(V1, V2) as defined in (6):

$$U=(\alpha(1+\beta)V1+(1-\alpha)(1+\beta)\beta1*f(Z_1, ..., Z_n))/ (1- (1-\alpha)(1+\beta)(\beta1+ (1-2\beta1)f(Z_1, ..., Z_n)) \qquad (16)$$

Thus, by the loop abstraction the properties (2)-(4) are replaced in the agent model specification by the property

$$\wedge_{i=1..n} \text{ srs(eval\_for(effect(a}_i\text{), bhope), } Z_i\text{) \& srs(G(bhope), V1)} \rightarrow_{3,3,1,1} \text{hope(o, U),} \qquad (17)$$

where U is calculated by (15) when $\beta \geq 0$ or by (16) when $\beta < 0$. The 'as if body loop' for fear is treated in the same manner.

Given stable emotional states, one can determine stable preparation states for actions in the behavioral chain which are influenced by these emotional states. The stable preparation state for action a is determined based on (10) and (11) as follows:

$$V5 = h(V1,V2,V3,V4)$$

Thus, (10) and (11) are replaced by (18) and (19) correspondingly:

srs(o, V1) & hope(o, V2) & fear(o, V3) & srs(G(a1), V4)
$\rightarrow$ preparation_for(a1, h(V1, V2, 1-V3, V4)) $\qquad (18)$

srs(effect(a), V1) & hope(o, V2) & fear(o, V3) & srs(G(a), V4)
$\rightarrow$ preparation_for(a, h(V1,V2,1-V3,V4)) $\qquad (19)$

Thus, the abstracted model specification comprises the properties (1) and (17) for hope and for fear; and (7)-(9), (18), (19).

Using the obtained abstracted model specification, the stable states of the whole multi-agent system can be determined analytically as follows. A stable emotional state (i.e., hope, fear) U_o = { $U_{o,Ai}$| A_i is an agent from the system } of the system towards option o is identified by solving the system of linear equations based on (15) and (16):

$$U_{o,Ai} = k_{1,o,U,Ai}+k_{2,o,U,Ai} \, g(\{U_{o,Aj, Aj \neq Ai}\}), \qquad (20)$$

where
$k_{1,o,U,Ai} = (\beta + (1-\alpha)(1-\beta)*\beta1*f(Z_1, ..., Z_n))/(1- (1-\alpha)(1-\beta)(\beta1+ (1-2\beta1) f(Z_1, ..., Z_n))$, when $\beta \geq 0$;
$k_{1,o,U,Ai} = (1-\alpha)(1+\beta)\beta1*f(Z_1, ..., Z_n)/(1- (1-\alpha)(1+\beta)(\beta1+ (1-2\beta1)f(Z_1, ..., Z_n))$, when $\beta < 0$;
$k_{2,o,U,Ai} = \alpha(1-\beta)/(1- (1-\alpha)(1-\beta)(\beta1+ (1-2\beta1)f(Z_1, ..., Z_n))$, when $\beta \geq 0$;
$k_{2,o,U,Ai} = \alpha(1+\beta)/(1-(1-\alpha)(1+\beta)(\beta1+(1-2\beta1)f(Z_1, ..., Z_n))$, when $\beta < 0$;
$g(\{U_{o,Aj, Aj \neq Ai}\}) = \Sigma_{j \neq I} \, \gamma_{UA_jA_i} U_{o,Aj}/ \Sigma_{j \neq i} \, \gamma_{UA_jA_i}$.

A stable preparation state for an action of the system can be determined in a similar manner.

4   Evaluation

In this section the results of a computational complexity evaluation of the abstracted behavioural agent model from Section 3 in comparison to the more complex internal model from Section 2 are presented. For the behavioural evaluation 1000 simulation trials

of both the original and abstracted models were performed. In each trial the values of the individual parameters were taken from the interval with uniformly distributed values

[0.2, 0.8] for α;   [0.1, 0.9] for $\beta1$;   [-0.5, 0.5] for β;
[0, 1] for ε_{pA_i}, for $\alpha_{pA_iA_j}$, and for δ_{pA_j} for all i, j, i≠j).

Furthermore, the agent's beliefs about accessibility of locations were initialized randomly from interval [0, 1]. The behavioural error for simulation trial i for the set of all agents AG was calculated as the averaged normalized root mean squared error:

$$NRMSE^i = \sum_{a \in AG} NRMSE_a^i / |AG|, \quad NRMSE_a^i = (\sum_{t=1..end\_time} (v_{a,t}^i - w_{a,t}^i)^2 / end\_time)^{1/2} \cdot 100\%,$$

Here $v_{a,t}^i$ is the degree of agent's a preparation to choose the first option estimated using the original model from Section 2, $w_{a,t}^i$ is the degree of agent's a preparation to choose the first option estimated using the abstracted model from Section 3; end_time is the simulation time (100 time points). The maximum and average (over the simulation trials) NRMSE obtained for 10, 100, 500 and 1000 simulation trials are provided in Table 1. As can be seen from the table the NRMSE is low and almost insensitive to the amount of agents used in the simulation.

Table 1. The maximum and average normalized root mean squared errors (NRMSE) for the abstracted agent model in comparison to the original agent model

# of agents	10	100	500	1000
Maximum NRMSE	*1.65%*	*1.5%*	*1.47%*	*1.54%*
Average NRMSE	*0.95%*	*0.89%*	*0.98%*	*0.96%*

The computational complexity evaluation is based on the estimation of the simulation time for the original and abstracted agent models. As can be seen in Table 2, the simulation time difference between the models is not very significant for a small number of agents, but becomes apparent for larger-scale multi-agent systems. Thus, the abstracted model ensures an increase of the computational efficiency at least from twice up to 3 to 4 times (and probably more) for large-scale multi-agent systems in comparison to the original model. This is achieved at the cost of approximately 1% behavioural error, which is not critical for most applications.

Table 2. Simulation time (in seconds) of the original and abstracted models

# of agents	10	100	500	1000	2000
Original model	*0.0173*	*0.59*	*12.98*	*62.12*	*281.43*
Abstracted model	*0.0157*	*0.25*	*5.28*	*20.96*	*78.18*
Gain factor	*1.1*	*2.4*	*2.5*	*3.0*	*3.6*

5 Conclusions

It is often argued that to ensure a high plausibility of models of socio-technical systems rich internal models of actors are required; cf. [4], [7]. However, such models

usually involve complex interactions between cognitive and affective processes, and therefore do not scale well. In this paper it is demonstrated how an abstracted behavioral agent model can be obtained from the complex decision making model from [13]. The type of abstraction used for internal agent models has similarities to an approach known in computational biochemistry as quasi steady state approximation, to address stiff differential equation models (in which some subprocesses run very fast compared to other subprocesses that run more slowly); e.g. [12], [15].

The abstracted agent model ensures an increase of the computational efficiency from more than twice for smaller numbers upto at least 3 to 4 times for larger numbers, at the cost of approximately 1% behavioural error. The abstraction mechanism employed can be used for a wide range of agent models with internal loops limited only by the requirements identified in [14]. Further abstraction steps can be made by dynamic clustering of agents in groups considered as higher-order agents; this will be addressed in future research.

References

1. Bosse, T., Jonker, C.M., van der Meij, L., Treur, J.: A Language and Environment for Analysis of Dynamics by Simulation. Int. J. of AI Tools 16, 435–464 (2007)
2. Damasio, A.: The Feeling of What Happens. In: Body and Emotion in the Making of Consciousness. Harcourt Brace, New York (1999)
3. Damasio, A.: The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. Philosophical Transactions of the Royal Society: Biological Sciences 351, 1413–1420 (1996)
4. Edmonds, B., Moss, S.: From KISS to KIDS – an 'anti-simplistic' modelling approach. In: Davidsson, P., et al. (eds.) MABS 2004. LNCS (LNAI), vol. 3415, pp. 130–144. Springer, Heidelberg (2005)
5. Helbing, D., Farkas, I.J., Vicsek, T.: Freezing by heating in a driven mesoscopic system. Phys. Rev. Lett. 84, 1240–1243 (2000)
6. Hesslow, G.: Conscious thought as simulation of behaviour and perception. Trends in Cog. Sci. 6, 242–247 (2002)
7. Iacoboni, M.: Mirroring People: the New Science of How We Connect with Others. Farrar, Straus & Giroux, New York (2008)
8. Kuligowski, E.D., Gwynne, S.M.V.: The Need for Behavioral Theory in Evacuation Modeling. In: Proc. of the 4th International Conference on Pedestrian and Evacuation Dynamics, PED 2008 (2008)
9. Leveson, N.: A new accident model for engineering safer systems. Safety Science 42, 237–270 (2004)
10. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
11. Rizzolatti, G., Sinigaglia, C.: Mirrors in the Brain: How Our Minds Share Actions and Emotions. Oxford Univ. Press, Oxford (2008)

12. Schauer, M., Heinrich, R.: Quasi-steady-state approximation in the mathematical modeling of biochemical reaction networks. J. Math. Biosci. 65, 155–170 (1983)
13. Sharpanskykh, A., Treur, J.: Adaptive Modelling of Social Decision Making by Agents Integrating Simulated Behaviour and Perception Chains. In: Pan, J.-S., Chen, S.-M., Kowalczyk, R. (eds.) ICCCI 2010. LNCS (LNAI), vol. 6421, pp. 284–295. Springer, Heidelberg (2010)
14. Sharpanskykh, A., Treur, J.: Behavioural Abstraction of Agent Models Addressing Mutual Interaction of Cognitive and Affective Processes. In: Yao, Y., Sun, R., Poggio, T., Liu, J., Zhong, N., Huang, J. (eds.) BI 2010. LNCS (LNAI), vol. 6334, pp. 67–77. Springer, Heidelberg (2010)
15. Stiefenhofer, M.: Quasi-steady-state approximation for chemical reaction networks. J. Math. Biol. 36, 593–609 (1998)

Medicare Auctions:
A Case Study of Market Design in Washington, DC

Peter Cramton

University of Maryland
`pcramton@gmail.com`

Abstract. One sensible way to reduce healthcare costs is to harness market forces, where practical, to nurture competition and innovation. Lower prices and improved services should follow. However, the switch to market pricing is not an easy one. Medicare's experience with medical supplies illustrates the challenges and offers some important lessons. The key lesson is that government programs can benefit from introducing market methods, but doing so requires good market design—something that may not come naturally to the implementing agency, especially in light of political forces and organizational inertia.

An auction design for Medicare Durable Medical Equipment is presented. The design addresses the flaws in the current program. Bids are binding commitments. Each bid binds the bidder to particular performance obligations depending on the auction outcome. The bids are made credible through a rigorous qualification one month before the auction. Each bidder provides a financial guarantee in the form of a bid bond or a deposit in proportion to the bidder's capacity. Capacity is objectively estimated based on the bidder's supply in recent years. Each winner provides a performance guarantee in proportion to the winner's estimated volume won. The auction establishes a market clearing price for each product in each service area. The price paid to all suppliers is the clearing price that balances supply and demand. These prices are found in a simultaneous descending clock auction, a simple price discovery process that allows both substitutions across items and complementarities. Competition in the auction comes from new entry or the expansion of existing suppliers into new product categories and service areas. After the auction, the winners compete for Medicare beneficiaries by offering quality products and services. Thus, beneficiary choice is used to further strengthen incentives to provide high quality products and services.

A Longitudinal View of the Relationship
Between Social Marginalization and Obesity

Andrea Apolloni[1], Achla Marathe[2], and Zhengzheng Pan[2]

[1] Institut des Systèmes Complexes Rhône-Alpes(IXXI),and Laboratoire de Physique,
École Normale Supérieure de Lyon 5, rue de Vercors, 69007 Lyon, France
andrea.apolloni@ens-lyon.fr
[2] Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics
Institute, Virginia Tech, Blacksburg, VA 24061, USA
{amarathe,zzpan}@vbi.vt.edu

Abstract. We use 3 Waves of the Add Health data collected between
1994 and 2002 to conduct a longitudinal study of the relationship be-
tween social marginalization and the weight status of adolescents and
young adults. Past studies have shown that overweight and obese chil-
dren are socially marginalized. This research tests (1) if this is true when
we account for the sample size of each group, (2) does this phenomenon
hold over time and (3) is it obesity or social marginalization that precedes
in time. Our results show that when the sample size for each group is
considered, the share of friendship is conforming to the size of the group.
This conformity seems to increase over time as the population becomes
more obese. Finally, we find that obesity precedes social marginalization
which lends credence to the notion that obesity causes social marginal-
ization and not vice versa.

Keywords: Obesity, friendship, social marginalization, causality, Add
Health.

1 Introduction

Obesity has become a global epidemic. According to the World Health Organi-
zation (WHO), there are one billion overweight adults and at least 300 million
of them are obese. The national health care cost projections show that by 2018,
21% of the health care spending will be on weight related medical bills, costing
the nation $344 billion [12]. Studies have shown a positive relationship between
childhood weight and adult weight, diseases, and mortality [5]. In the last three
decades, the prevalence of obesity has at least tripled among children in the US,
which has become a major public health concern.

The forces driving the obesity epidemic are multiple and complex. A com-
plicated interplay of behavioral, environmental, social and economic factors is
associated with weight gain. For example, social marginalization and peer re-
jection tend to increase sedentary behavior, reduce active leisure activities and
encourage eating out of boredom. Adolescents and young adults often form peer

groups around shared behaviors such as watching TV, eating out, playing sports or video games etc. which often has either a direct or indirect effect on the weight status [13]. This homophily in behavior is likely to result in homophily in terms of the weight status. At the same time, researchers have shown that obese people form clusters with other obese people and this homophily in BMI (Body Mass Index) can cause homophily in behavior [3,4,7]. This clearly implies that the relationship between BMI, the behavior that leads to high BMI and social marginalization is complex and intertwined.

Previous studies have shown that overweight children and adolescents are often not favored in their social networks resulting in their social marginalization [11]. In this research, social marginalization is measured by the number of friendship links an individual has. These links are split by the BMI status and adjusted for the sample size in each category to account for the fact that fewer friendship links with overweight and obese may occur because there are less of them in the data. We also consider the role of race and gender in assessing individuals' tolerance of the friends' weight status [10,2].

This research uses longitudinal data on friendship network from the National Longitudinal Study of Adolescents Health (Add Health). It compares the average number of friends a group (identified by race, gender and BMI) has to the sample size of the group to capture the bias in friendship links. This comparison is performed on individuals that are present across the 3 Waves to get a longitudinal view of the bias. The friendship data and BMI status for the 3 Waves is further used to extricate the causality relationship between social marginalization and obesity. The aim is to determine if social marginalization causes obesity or if obesity causes social marginalization among adolescents and young adults.

2   Data Description

Information about individual weight status and friendship network is obtained from the Add Health data [9]. This national longitudinal survey on adolescents started in 1994-95 and followed the same cohort from adolescence into young adulthood. Wave I and II data were collected in 1994-1996 and Wave III data were collected in 2001-2002. This study uses longitudinal data on 20,502 survey respondents who were present in all 3 Waves.

The survey collects data on respondents' physical, educational, and economical status, together with their friendship network. In each Wave, each respondent was asked to nominate 5 male and 5 female friends. Each nomination creates a directional link in the friendship network between the ego and the nominated person.

In this study, any two individuals are called "friends" when each individual nominates the other as his/her friend. This mutual-nomination measure of friendship differs from other notable research on this topic where friends are measured as the number of friendship nominations or the in-degree measure [11]. The reason for choosing mutual friends as the measure, as opposed to in-degree, is that

studies have found that adolescents nominated by overweight and obese individuals as friends are less likely to reciprocate the nomination as compared to the friends of the normal weight adolescents. Out-degree measures the self-reported friendship ties which is known to provide inaccurate number of friends [1]. This means both in-degree and out-degree can provide a biased view of the friendships since reciprocity is often missing among the friends of the obese individuals. Therefore we resort to bi-directional nominations to get a more realistic view of the friends.

3   Methodology

3.1   Weight Assessement

We use Body Mass Index (BMI) as the indicator of an individual's weight status. BMI (in kg/m^2) is defined as the weight in kilograms divided by the square of height in meters. We categorize an individual's BMI status into 4 classes i.e. 1 = underweight (BMI < 18.5); 2 = normal ($18.5 \leq$ BMI ≤ 24.9); 3 = overweight ($25 \leq$ BMI ≤ 29.9); and 4 = obese (BMI ≥ 30) [6].

(a) BMI status   (b) Friendship by BMI status
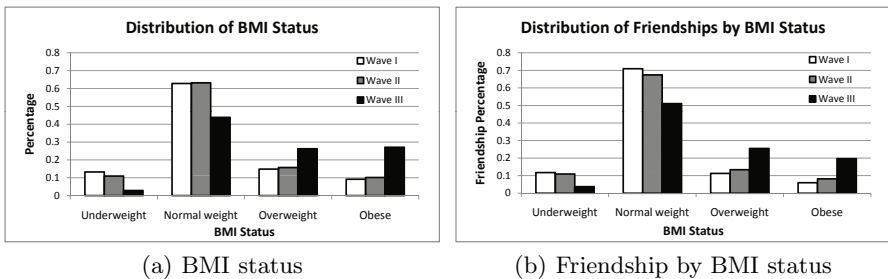
Fig. 1. Distribution of BMI status and friendships for Wave I, II, and III

Figure 1 shows the distribution of BMI status and friendships by BMI status for all 3 Waves. The comparison between Figure 1(a) and Figure 1(b) reveals the following findings. First, the normal weight people make the most friends in all 3 Waves. They are also the biggest group in all 3 Waves. Their share of friends is higher than the percentage of their BMI group, which suggests that they are indeed popular, although this difference declines over time. Second, underweight group has the most conforming share of friends since Figure 1(a) and Figure 1(b) have almost the same fractions for this group. In other words, they are neither favored nor discriminated against in the social networks. Third, overweight and obese individuals have less friends than the percentage of their sample. However, note that in Wave III, as there are more overweight individuals in the population, they also make more friends and the share of friends is roughly

equal to the percentage of the group. This shows that social marginalization of the overweight adolescents decreases as the percentage of overweight population increases.[1]

3.2   The Friendship Matrix

Next, we build a friendship matrix deduced from the mutual friendship nominations as given in the Add Health data. In this matrix, each row shows the distribution of friends across various classes for the row index, where the index represents a group.

For example, suppose we have a row for the underweight group and this group makes 100 friends total. Then the numbers in the row imply that 25 of these

	1	2	3	4
1	0.25	0.50	0.15	0.10

friends are underweight, or of BMI status 1; 50 are of normal weight; 15 are overweight; and 10 are obese.

To further understand how each BMI status class performs given its sample size in the Add Health population, we take the friendship data and split it by race, gender and BMI status. We use a 3-digit system to categorize individuals. The first of the 3-digits shows the race of the individual which ranges from 1 to 5: 1 = White; 2 = Black; 3 = Asian; 4 = Native; 5 = Other. The second digit shows the gender and takes value 1 for male and 2 for female. The third digit is the person's BMI status which ranges from 1 to 4 as mentioned earlier: 1 = Underweight; 2 = Normal; 3 = Overweight; 4 = Obese. For instance, a category of 314 refers to the group of obese Asian boys, and 422 refers to the group of normal weight Native girls.

A complete friendship matrix has 40 rows and 40 columns for each of the 40 race-gender-weight categories, which is very hard to read. Therefore, to summarize the popularity of a race-gender-weight group, we use the average of each column. For instance, the following friendship matrix of 3 categories shows that the normal weight white boys (112) are the most popular with an average percentage of friends of 0.667 and the overweight white boys (113) are the least popular with the column average of 0.15. In the middle, the average percentage of friends made with underweight white boys (111) is 0.183.

However, this friendship matrix can provide a biased view if we do not account for the size of each group in the population. In order to understand this relationship better, we calculate the distribution of race-gender-weight categories as a reference. If there is no discrimination, then the average share of friendship of each category should be equal to the distribution of that category. For instance, assume that in the example shown above, there are 20% boys in category

	111	112	113
111	0.35	0.60	0.05
112	0.15	0.75	0.10
113	0.05	0.65	0.30

	111	112	113
avg	0.183	0.667	0.15
distri	0.20	0.60	0.20

---

[1] Note that Wave I and II data are not too different, since the data for the first two Waves were collected in quick succession (i.e. between 1994-1996) which probably did not allow enough time for variables such as BMI and number of friends to change significantly.

111, 60% in category 112, and 20% in category 113, then we see that both under-weight and overweight white boys are not favored as friends (i.e., avg < distri), the overweight boys, even more so.

Figures 2 and 3 show the average share of friends and the size of the group in each category split by race, gender and BMI for Waves I and III. Since Wave I and II plots are very similar, we only show the plot for Wave I here.
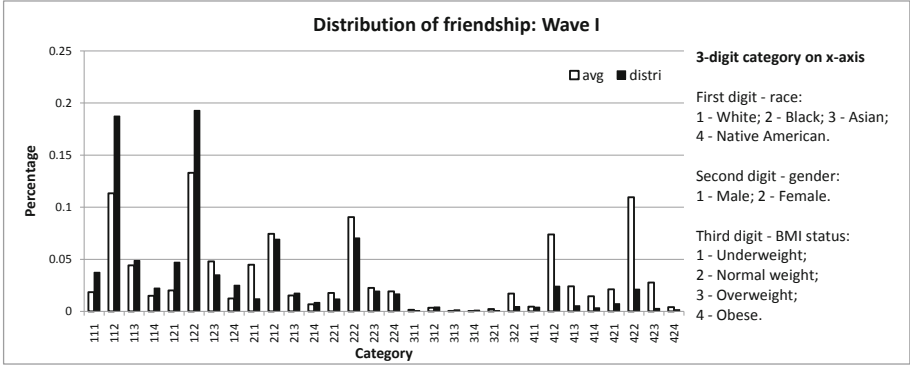
Fig. 2. Distribution versus average percentage of friends by each category split by race, gender and BMI status for Wave I

The plots show that (1) the fraction of Asians and Native Americans is small among the Add Health respondents, which causes low to zero friendship share in some cases. (2) In Wave I, Native Americans have a higher average number of friends than the distribution. But in Wave III, this difference drops. (3) In Wave I and III, almost all Whites' averages are less than the distribution which means they were not favored as friends. (4) In Wave III, all categories of Black except obese (i.e. 214 and 224) are slightly favored as friends. (5) The worst performer in terms of the number of friends is the white obese male category in Wave III where the average number of friends is only one third of the distribution. The obese group does fairly well in other categories in Wave III. This highlights the social stigma attached to the higher weight status among whites compared to other races.

Overall, these results show that whites are the least popular as friends among all the races. White obese males face the highest risk of social marginalization among all the categories. Native females of normal weight seem to be very pop-ular as friends. These results give us insight on how the race, gender and BMI status are associated with the number of friends a person has. However, it is still not clear as to whether the lack of friends is causing the person to be overweight or it is the overweight that causes the person to not be able to make friends. We discuss this issue in the next section.
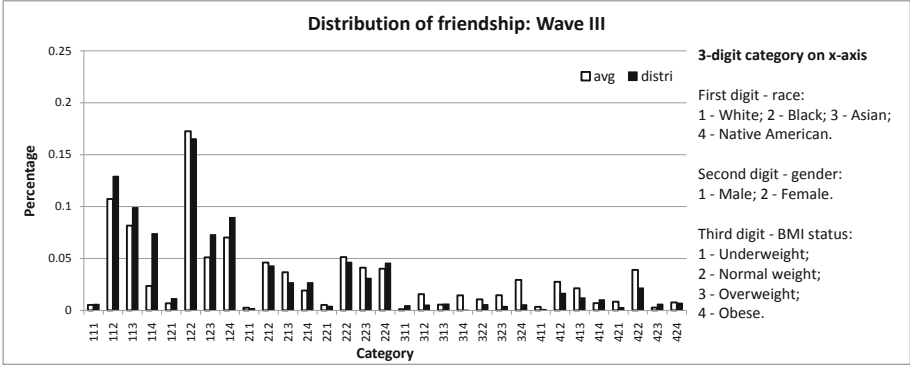
Fig. 3. Distribution versus average percentage of friends by each category split by race, gender and BMI status for Wave III

3.3   Causal Relationship

In this research, we test for the direction of causality to see whether social isolation causes obesity or obesity causes social marginalization, using the causality concept developed by Granger [8]. According to Granger, variable x causes variable y if past values of x contain information that helps explain the future values of y above and beyond the information contained in the past values of y alone. For instance, suppose we have

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + u_t.$$

Then x Granger causes y if $b_i \neq 0$, i.e., at least one coefficient of x is significantly different from zero. Another way to check whether x helps predict the future values of y is by testing if the variance of the residuals u_t drops when lagged values of x are included. For example when

$$y_t = a_1 + y_{t-1} + a_2 y_{t-2} + \ldots + v_t.$$

If x Granger causes y, variance of $(u_t) \leq$ variance of (v_t) since adding x in the regression increases the explanatory power and reduces the residuals.

   The concept of Granger causality is particularly relevant here because it accounts for the timing of the events i.e. x causes y if changes in x precede changes in y. We measure social isolation or marginalization by the number of friends a person has and, obesity by the BMI status. The following model (as well as its counter model where BMI and friends were swapped) was estimated:

$$Friends_3 = c + a_1 Friends_1 + a_2 Friends_2 + b_1 BMI_1 + b_2 BMI_2 + u_t, \quad (1)$$

$$Friends_3 = c + a_1 Friends_1 + a_2 Friends_2 + v_t, \quad (2)$$

where $Friends_x$ is the number of friends in Wave x, and BMI_x is the BMI status in Wave x.

The results are as follows:

$$Friends_3 = 1.45 - 0.25 Friends_1 + 0.24 Friends_2 - 0.22 BMI_1 - 0.06 BMI_2 + \hat{u}_t, \tag{3}$$

$$Friends_3 = 0.82 - 0.24 Friends_1 + 0.25 Friends_2 + \hat{v}_t \tag{4}$$

The t-values of the estimated coefficients show that all coefficients in both equations are statistically significant at 5% level or less. The variance of $(\hat{u}_t) <$ variance of $(\hat{v}_t)$. The coefficients of the lagged values of BMI show that the BMI status in Waves I and II negatively impacts the number of friends in Wave III in a statistically significant way. The observation that obesity precedes social marginalization lends credence to the notion that obesity causes social isolation rather than the other way around. We also tested the counter model where the BMI status and the friends were swapped in the above model to see if the lagged number of friends caused the future BMI status to change. The estimated coefficients were not statistically significant. The results are available from the authors upon request.

4   Summary and Conclusions

To address the relationship between weight and social marginalization, we first examine whether overweight and obese adolescents are discriminated in the friendship network. We stress that when discussing this issue, we need to account for the sample size of each category in the population, and consider if the seemingly lower popularity is simply representative of a smaller sample size. Using mutual nomination links as friends, over an 8-year period data from Wave I, II, and III, we conclude that overweight and obese adolescents and young adults have fewer friends, even after adjusting for their sample size. However, note that in Wave III, as the entire population is moving towards higher weight status, the overweight individuals' share of friends is roughly equal to the percentage of the group.

We further study the issue with divided race and gender groups and find distinctive patterns across different races. The most noticeable findings are the high popularity of Native American individuals of all BMI status, and low popularity of white obese individuals. This result highlights the cultural differences in the stigma associated with being overweight.

Most significantly, our results support the view that obesity causes social marginalization and not vice versa. This is an important finding for policy makers and planners who are looking for ways to reduce social marginalization of children. This research makes it clear that obesity not only has health consequences but also impacts the behavioral and social processes which are critical to the overall growth and development of children.

References

1. Bondonio, D.: Predictors of accuracy in perceiving informal social networks. Social Networks 20(4), 301–330 (1998)
2. Caprio, S., Daniels, S., Drewnowski, A., Kaufman, F., Palinkas, L., Rosenbloom, A., Schwimmer, J.: Influence of race, ethnicity, and culture on childhood obesity: Implications for prevention and treatment. Diabetes Care 31(11), 2211 (2008)
3. Christakis, N., Fowler, J.: The spread of obesity in a large social network over 32 years. N. Engl. J. Med. 357(4), 370–379 (2007)
4. Cohen-Cole, E., Fletcher, J.: Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. Journal of Health Economics 27(5), 1382–1387 (2008)
5. Dietz, W.: Health consequences of obesity in youth: childhood predictors of adult disease. Pediatrics 101(3), 518 (1998)
6. Dietz, W., Bellizzi, M.: Introduction: the use of body mass index to assess obesity in children. American Journal of Clinical Nutrition 70(1), 123S (1999)
7. Fowler, J., Christakis, N.: Estimating peer effects on health in social networks. Journal of Health Economics 27(5), 1386–1391 (2008)
8. Granger, C.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438 (1969)
9. Harris, K.M.: The national longitudinal study of adolescent health (add health), waves i & ii, 1994-1996; wave iii, 2001-2002; wave iv, 2007-2009 [machine-readable data file and documentation]. Carolina Population Center, University of North Carolina, Chapel Hill, NC (2009)
10. Ritenbaugh, C.: Obesity as a culture-bound syndrome. Culture, Medicine and Psychiatry 6(4), 347–361 (1982)
11. Strauss, R., Pollack, H.: Social marginalization of overweight children. Arch. Pediatr. Adolesc. Med. 157, 746–752 (2003)
12. Thorpe, K.: The future costs of obesity: National and state estimates of the impact of obesity on direct health care expenses. A Collaborative Report from United Health Foundation (2009)
13. Valente, T., Fujimoto, K., Chou, C., Spruijt-Metz, D.: Adolescent affiliations and adiposity: A social network analysis of friendships and obesity. Journal of Adolescent Health 45, 202–204 (2009)

Open Source Software Development: Communities' Impact on Public Good

Helena Garriga, Sebastian Spaeth, and Georg von Krogh

ETH Zurich, Department of Management, Technology, and Economics
Kreuzplatz 5, Zurich, 8032, Switzerland
{hgarriga,sspaeth,gvkrogh}@ethz.ch

Abstract. This study examines the innovation output of software development that produces public goods. We use resource dependence theory and collective action theory to explain the effects of interconnectedness on open source software (OSS) communities, and on contributions to public goods. We empirically test our proposals using an eight-year panel dataset on OSS projects based on the Eclipse Foundation, and conclude that interconnectedness negatively affects community mobilization and its contributions to public goods.

Keywords: Resource dependence, public good, open source software.

1 Introduction

Voluntary alliances in R&D are interesting because they involve no legal obligations to safeguard contributions while creating public good innovations, characterized by non-exclusivity and non-rivalry [1]. Examples of such alliances are OSS or the public human genome database, to which individuals and companies contribute voluntarily. Such alliances are typically subject to the "tragedy of the commons" [2] and the free-rider problem [3] that plagues any public good endeavor relying on collective action [4]. Some authors have proposed that contributions to public goods should be understood as a collective action problem, but so far, to our knowledge, no theoretical framework for the setting of such public goods has been developed and empirically investigated [5-7]. Many OSS projects are prominent examples of public good innovation: OSS is free for all to download, use, modify, and redistribute. Individual defection is immanent and resource under-provision may halt development—key problematic features of collective action. We explain community participation through collective action in public goods using elements of resource dependence theory (RDT) [8], which highlights the effects of interconnectedness in public goods.

OSS, by making software a public good, invites participation and contributions from volunteers. However, because OSS exhibits public good properties, it always faces the risk of under-provision and over-exploitation. A number of studies attempt to explain how some OSS benefits from volunteers using motivation models or software characteristics[9]. However, so far there is no clear picture of how community participation affects public good outcomes.

We examine this dilemma with the Eclipse Foundation, a group of OSS projects. We investigate how contributions to public good depend on community characteristics

and test our theory on Eclipse. The organization of this paper is as follows. We start with a review of existing literature on collective action and define core concepts of RDT. Subsequently, we present our extended model, hypotheses, and research design. Finally, we discuss our results, conclude the paper, summarize the insights and suggest some potential future research issues.

2   Theory

2.1   Collective Action in Public Goods

A branch of study on strategic alliances argues that their core objective is to create a public-good [5]. We understand "public good" to mean anything that results from collective action by interested parties, has no possibility of exclusion [7]. Past literature has focused on contributions made to creating and maintaining a public good [5-7]. Oliver and Marwell's work on critical mass [5] relates the input of resources and their interdependence to variable public good outputs. However, so far these studies have been limited to discussing public good production in terms of two simple functions, acceleration or deceleration, and have failed to test their propositions in a real setting. Monge et al. [7] use public good theory to understand alliance-based communication and information systems. Their study proposes that participants' individual characteristics and the resources contributed to the public good will impact future gains and further resource contributions, indicating that participant heterogeneity is favorable for collective action [5], and key for predicting the extent of output. Monge et al. [7] built a model of inter-organizational communication and information within public goods, drawing on earlier theory of interactive communication systems [10].

2.2   Resource Dependence Theory

RDT proposes that actors lacking essential resources to pursue certain goals or conduct their business seek to establish relationships with others in order to obtain the resources they need. However, this makes them dependent on others, which they may perceive as non-optimal [8]. Actors continuously attempt to alter dependency relationships by minimizing their own dependence on others or by gaining influence over common resources, and increasing others' dependence on them.

RDT models have been used to study organizational performance [11] in an environmental context. However, capturing all the externalities that affect the dependencies within an organization is a complicated procedure that resists operationalization. In this study, we define and construct some of the characteristics identified as environmental determinants by RDT and operationalize them in a public good based setting.

Interconnectedness [8] is defined as the connection between participants in a public good. Public goods with high interconnectedness face high uncertainty in producing a collective outcome, due to conflicting goals and high coordination costs [8]. Yet Simon [12] suggests that loosely connected organizational systems have higher chances of survival; inter-connectedness, even when producing conflict, is necessary for the success of a public good.

Community mobilization is the interaction between knowledge seekers and knowledge providers, and between developers and users of software, and it is critical for the good's success. Community mobilization, which depends on the environmental setting, has an impact on overall contributions and is crucial for the public good development strategy.

3   Hypotheses and Model

According to the theory of collective action presented above, a public good should motivate volunteers to contribute their own resources to it. Also, decisions to participate are interdependent, making people's willingness to contribute dependent on what others do [6]. In mature projects, where the time elapsed since creation is relatively high, more open source code is available, which will lead potential contributors to believe in the eventual success of the project [1, 13]. Working software will be used for further modification, development, and distribution [14].

Relative lack of project maturity, compared to other projects, leads to uncertainty [8], as collective goods have significant start-up costs that lead to an accelerating dynamic in contributions [6]. Maturity is indicative of abundant resource allocation in the public good; contributions leading to a public good increase over time. Therefore, we suggest that maturity has a positive impact on community mobilization:

Proposition P1. Public good's maturity positively impacts community mobilization.

In public good innovation, interconnectedness is important for the innovation itself and determines how the public good is structured [15]. Interconnectedness in OSS is defined as the connections between participants in a public good. Interconnectedness in public goods can be reduced if projects can be divided into modules that interface through limited and defined transactions. Public good innovation can have a monolithic or modular design. In the former, which creates high interdependence, one contribution impacts on all others. In the latter, individuals can specialize in contributing to a module. Coordination is through predefined interfaces between modules. These lower the cost of coordination; the fewer the partners and the smaller each module, the less effort will be required to understand the module architecture [15, 16]. Lower coordination costs mean higher marginal benefits, which in turn motivates higher contribution to the public good:

Proposition P2. Interconnectedness between modules negatively affects community mobilization.

Community mobilization is a valuable source of knowledge. It helps development through virtual forums, giving knowledge away for free. It is not governed by an employer or hierarchy [17], it challenges the established proprietary software, and participates in the production of new software [18], one of the reasons for initiating the movement. Individuals from the OSS community view their participation as an investment from which they might obtain future economic returns [19], such as revenues or self-marketing, and this motivates their engagement. A wide variety of contributions is expected in any OSS, from reports on software bugs to add-on suggestions. We suggest that community mobilization has a positive impact on knowledge made freely available to a public good:

Proposition P3. Community mobilization positively impacts contributions.

4   Research Design

We chose the Eclipse Foundation to test empirically the proposed theoretical model. Eclipse is a multilingual platform for developing and supporting highly integrated software engineering tools, and is the world's largest pool of OSS-sponsored projects and developers. It provides an ideal empirical context allowing us to investigate the effects of composition in public goods, contribution levels, and impact on community. Our empirical analysis draws on panel datasets containing information on 11 top projects provided by the Eclipse website, from the second quarter of 2001 to the third quarter of 2009.

4.1   Variables

We measure contributions by the community as the number of bugs periodically reported to any top project developed in Eclipse (see Graph 1). Bug reporting data are widely available and easy to extract, providing a reliable infrastructure for studying community contributions. They have been used as a participation measure in other OSS articles [20]. A total of 294,140 reported bugs were accounted for, and we define contributions as the number of bugs reported in each project quarterly.

   Research on community mobilization in OSS has focused on mailing-list systems [1, 21]. Mailing lists do not contribute lines of code, but support the interaction between knowledge seekers and knowledge providers, and between the developers and users of the software, playing a major role in its diffusion and improvement. We define community mobilization as the number of messages posted quarterly on mailing lists in each project (see Graph 2). We captured 147,725 messages posted on mailing lists.

Graph 1. Contributions to public goods   **Graph 2.** Community mobilization in public goods
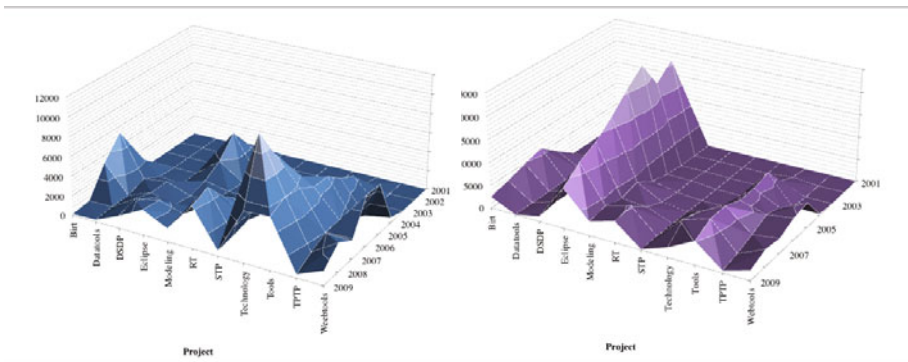
   In our model we relate process maturity to the time already invested in developing the public good. Next, we define interconnectedness as the level of modularity of the public good, and use project structure to size this variable. Our measure of project interconnectedness is the inverse of the active number of modules in a top project for each quarter. Finally, we use lines of code to control software size. We also control for the number of committers, identifying 1,149 incumbent committers in Eclipse at any point. Finally, we control for time-effects using a vector of 36 time dummies.

4.2 Data Analysis

The model considers the simultaneous relationships between a set of variables measured across time for a set of projects, a cross-panel dataset. Cross-sectional analysis allows extrapolation of results to a population, while time-series analysis is helpful in determining causality as well as seeing whether the same relationship holds across time for an individual. We analyze the data as a cross-panel dataset, with yearly quarters from to 2001Q2 to 2009Q3, 36 time periods, for 11 observed projects.

We checked for correlation of the independent variables, we estimated whether there are random, fixed (Lagrange-Multiplier test) or time effects in our data. Our results showed unbalanced panel data with significant fixed and time effects. Thus, a vector of 36 quarterly time dummies was incorporated in all models to control for unobserved time-varying characteristics common to all projects [22].

We also control for the problem of endogeneity using fixed effects [23]. Heteroskedasticity in the panel data is confirmed using the Lagrange multiplier test and serial correlation is confirmed using the Wooldridge approach [24]. We test for cross-sectional dependence using the Pesaran approach [25]. Accordingly, we use the robust standard errors [26], previously used in other panel data analysis [23], when the panel data has heteroskedasticity and serial correlation. Finally, we create four models (Tables 1 & 2) to evaluate our hypotheses.

5 Results and Conclusion

Recent interest in public good innovation, the contributions made to and by it, and its increasing occurrence makes it important to understand the incentives for individuals to allocate resources to these projects and participate actively in them. The effects of resource stability on community mobilization were measured using maturity, (Model II in Table 1). We hypothesized that maturity positively affects community mobilization (P1); however, our results do not support this statement conclusively, as maturity has an insignificant impact on community mobilization. We also proposed that interconnectedness has a negative impact on community mobilization (P2). Our results showed a negative and significant impact ($p > 0.05$), validating our hypothesis.

Table 1. Models explaining community mobilization in OSS public goods

Community mobilization	I	II
Interconnectedness		-2.015**
Maturity		-0.0387
Number of committers	0.030***	0.0275**
Size	-5.42e-08***	-3.78e-08*
R-sq	0.59	0.62

Statistical significance denoted as: * $p > 0.1$, ** $p > 0.05$, *** $p > 0.01$. Standard errors in parentheses. Included in models but not reported are 36 time dummies as control variables.

We studied the impact of community mobilization on contributions (Table 2). Model IV shows the positive and significant impact of community mobilization on contributions (P3). One key advantage of public good is usually said to be the involvement of volunteer individuals, and our analysis supports this statement. Since contributions are measured in absolute terms, could a community be expected to halt development? Future research should follow up this topic either through historical methods or simulations. Our study also shows how lower interconnectedness, through modules, translates into increased contributions. Although this was not one of our propositions, it is an intuitive extension of interdependence effects on community mobilization.

Table 2. Models explaining contributions through bug reports in OSS public goods

Contributions	III	IV
Interconnectedness		-1.451**
Maturity		-0.167
Community mobilization		0.250**
Number of committers	0.0348***	0.025***
Size	-9.62e-08***	-3.94e-08**
Adj R-sq	0.71	0.77

Statistical significance denoted as: * $p>0.1$, ** $p>0.05$, *** $p>0.01$. Standard errors in parentheses. Included in models but not reported are 36 time dummies as control variables.

As public goods rely on communities' involvement, it is crucial to have mechanisms that motivate this participation. Interconnectedness impacts community mobilization, as well as contributions to the public good. In Graph 3, we plot the estimated contributions to Eclipse using community mobilization and interconnectedness constructs. We show a trend of increased modules, which partly explains the increased participation. This has previously been proposed [15], but with limited empirical evidence. Our evidence, presented here, contributes to that discussion. However, other factors might add interdependence to the environment, such as concentration [8]. Communities might be reluctant to participate in projects where only one committer has the right to change the repository. Thus, community mobilization might be higher when many developers participate and independency is minimal.

Future research should focus on whether the negative correlation between interconnectedness and community mobilization is a general phenomenon for OSS. More studies will be necessary to see whether there is any significant difference in commercial software developments. We suggest that few parallels can be expected, only if they are highly dependent on voluntary contributions.

This article contributes to the integration of RDT with collective action theory, and the under-provision and free-rider problem [27]. We look closely at the parameters that hinder the potential of a public good, and, using RDT, explain how provisioning through collective action depends on environmental factors. We examine those factors, which lead to standards of deviance and conflict that alter provisioning [8].

Graph 3. Contributions through community mobilization and modules in OSS
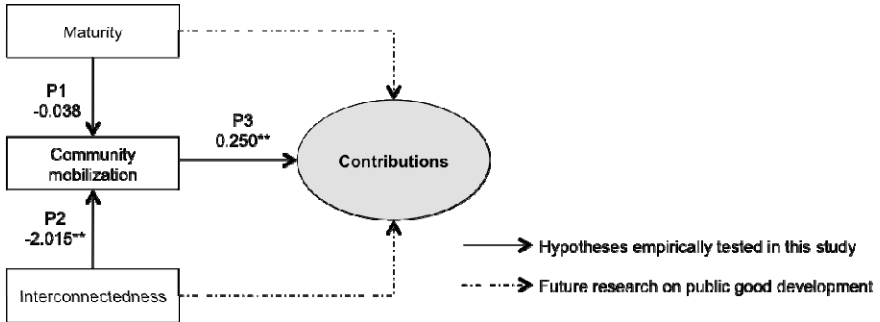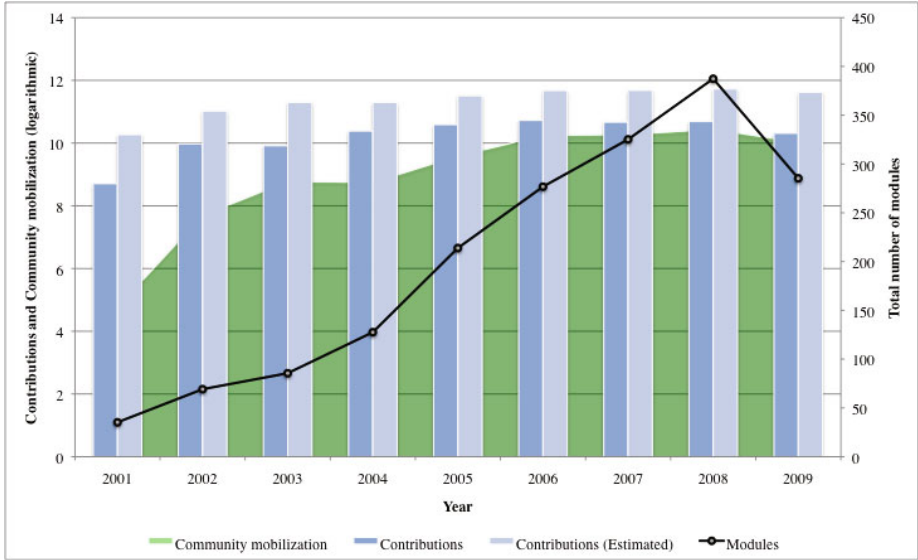
Fig. 1. Model of public good in OSS development using RDT and collective action

As more actors and resources become part of a public good, synergies can be found that boost contributions, but additional costs and dependencies also arise. This study highlights the importance of interconnectedness for community mobilization, and tentatively underlines its impact on overall contributions, which, however, we leave for future research. It is fundamental to the success of a public good not only to weigh the strengths and weaknesses of the innovation mechanisms available, but also to understand and predict how these mechanisms interact when used in tandem, and to recognize that more is not always better. Future research should consider these new clues, and extend our findings with more empirical evidence and joint theories.

References

1. von Hippel, E., von Krogh, G.: Open source software and the "private-collective" innovation model: Issues for organization science. Organization Science 14(2), 209–223 (2003)
2. Hardin, G.: The Tragedy of the Commons. Science 162(3859), 1243–1248 (1968)
3. Olson, M.: The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press, Cambridge (1965)
4. Osterloh, M., Rota, S.: Open source software development–Just another case of collective invention? Research Policy 36(2), 157–171 (2007)
5. Oliver, P.E., Marwell, G., Teixeira, R.: A theory of the critical mass. I. Interdependence, group heterogeneity, and the production of collective action. American Journal of Sociology 91(3), 522–556 (1985)
6. Marwell, G., Oliver, P.E., Prahl, R.: Social networks and collective action: A theory of the critical mass. III. American Journal of Sociology 94(3), 502–534 (1988)
7. Monge, P.R., et al.: Production of Collective Action in Alliance-Based Interorganizational Communication and Information Systems. Organization Science 9(3), 411–433 (1998)
8. Pfeffer, J., Salancik, G.R.: The external control of organizations: a resource dependence perspective, vol. xiii, p. 300. Harper & Row, New York (1978)
9. Shah, S.: Motivation, governance, and the viability of hybrid forms in open source software development. Management Science 52(7), 1000–1014 (2006)
10. Fulk, J., et al.: Connective and communal public goods in interactive communication systems. Communication Theory 6(1), 60–87 (1996)
11. Aldrich, H.: Resource Dependence and in Terorganiza Tional Relations: Local Employment Service Offices and Social Services Sector Organizations. Administration Society 7(4), 419–454 (1976)
12. Simon, H.A.: The sciences of the artificial. MIT Press, Cambridge (1996), 9780262691918
13. Lerner, J., Tirole, J.: Some simple economics of open source. The Journal of Industrial Economics 50(2), 197–234 (2002)
14. Haefliger, S., von Krogh, G.F., Spaeth, S.: Code Reuse in Open Source Software Development. Management Science 54(1), 180–193 (2008)
15. Baldwin, C.Y., Clark, K.B.: The Power of Modularity. In: Design Rules, vol. 1. The MIT Press, Cambridge (2000)
16. MacCormack, A., Rusnak, J., Baldwin, C.Y.: Exploring the structure of complex software designs: An empirical study of open source and proprietary code. Management Science 52(7), 1015–1030 (2006)
17. West, J., O'Mahony, S.: The Role of Participation Architecture in Growing Sponsored Open Source Communities, pp. 145–168. Carfax Pub. Co., Abingdon (2008)
18. Bechky, B.A.: Sharing meaning across occupational communities: The transformation of understanding on a production floor. Organization Science 14(3), 312–330 (2003)
19. Hars, A., Ou, S.: Working for free? Motivations for participating in open-source projects. International Journal of Electronic Commerce 6(3), 25–39 (2002)
20. Hertel, G., Niedner, S., Herrmann, S.: Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. Research Policy 32(7), 1159–1177 (2003)
21. von Krogh, G.F., Spaeth, S., Lakhani, K.: Community, joining, and specialization in open source software innovation: a case study. Research Policy 32(7), 1217–1241 (2003)

22. Bartels, B.: Beyond fixed versus random effects: a framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data. The Society for Political Methodology (2008)
23. Roebuck, M.C., Liberman, J.N.: Impact of Pharmacy Benefit Design on Prescription Drug Utilization: A Fixed Effects Analysis of Plan Sponsor Data. Health Services Research 44(3), 988–1009 (2009)
24. Wooldridge, J.M.: Econometric analysis of cross section and panel data. The MIT press, Cambridge (2002)
25. Pesaran, M.H.: General diagnostic tests for cross section dependence in panels (2004)
26. Hoechle, D.: Robust standard errors for panel regressions with cross-sectional dependence. Stata Journal 7(3), 281 (2007)
27. Ostrom, E.: Governing the commons: The evolution of institutions for collective action. Cambridge Univ. Pr., Cambridge (1990)

Location Privacy Protection on Social Networks

Justin Zhan and Xing Fang

Department of Computer Science
North Carolina A&T State University
{Zzhan,xfang}@ncat.edu

Abstract. Location information is considered as private in many scenarios. Protecting location information on mobile ad-hoc networks has attracted much research in past years. However, location information protection on social networks has not been paid much attention. In this paper, we present a novel location privacy protection approach on the basis of user messages in social networks. Our approach grants flexibility to users by offering them multiple protecting options. To the best of our knowledge, this is the first attempt to protect social network users' location information via text messages. We propose five algorithms for location privacy protection on social networks.

Keywords: Location information; privacy; social networks.

1 Introduction

Online social network has gained remarkable popularity since its debut. One of the most popular online social networks, Facebook, has reached more than 500 million active users [4]. Social network users are able to post their words, pictures, videos, etc. They can also write comments, share personal preferences, and make friends. Facebook even integrates messenger on their webpages supporting users instant online chatting with current online friends.

Online social networks typically deal with large amounts of user data. This may lead to privacy-related user data revelation [3, 14, 16]. Personal information, such as personal interests, contact information, photos, activities, associations and interactions, once revealed, may have different levels of impact, ranging from unexpected embarrassment or reputational damage [11] to identity theft [15]. Furthermore, effectively managing privacy for social network can be quite tricky. One reason is that different individuals have different levels of privacy-related expectations towards their information. For instance, some of them may be glad to publish their personal profiles tending to potentially develop additional friendships, while others may worry about the exposure of their identities so as to only reveal the profiles to their selected friends. In other situations, people are even willing to disclose their personal information to anonymous strangers rather than acquaintances [6]. Unfortunately, people may not often care about their privacy. Gross and Acquisti [6] found that only a few students change the default privacy preferences on Facebook by analyzing and evaluating the online behavior as well as the amount of private information disclosed from 4000 students in Carnegie Mellon University. Location information may be

sensitive because it enables the possibility of tracing someone. Motivated by this reason, we propose our privacy location information protection algorithms using encryption, k-anonymity and noise injection techniques.

The rest of this paper is organized as follows: In section 2, we briefly review some related work on location privacy protection. In section 3, we introduce our location privacy protection approach. We conduct experimental evaluation and present the result in section 4. We conclude the paper in section 6.

2   Related Work

2.1   Location Privacy Protection on Mobile Networks

Location privacy protection has gained much attention on mobile networks. With the rapid development of computing technology, computational capability harnessed from mobile phone hardware enables the running of some powerful software applications. Location-based service (LBS) is one of the applications where its service provider requires users to provide their location information in order to response their requests [2]. Schilit et al. [12] pointed out that location-based service is prone to privacy revelation. The service even incurs economic and reputation damages because of its potential privacy risks. Amoli et al. [2] classified privacy protections on LBS into two categories: policy-based approaches and data hiding techniques. In the policy-based approaches, a set of pre-defined policies are applied to users to inform that when and how the data can be stored, used, and revealed. K-anonymity, dummy-based approach, and obfuscation are three methods for the data hiding techniques. The k-anonymity works when a user's location can be hidden within a set of k members, who possess the same location information. In the dummy-based approach, a user is able to create some dummy location information to the service provider instead of sending out the real location information, while the obfuscation allows users to generalize their actual locations before sending them out. Since the aforementioned approaches have shortcomings, Amoli et al. proposed 2Ploc, a privacy preserving protocol of LBS based on one-time tickets [2]. Location protections also apply to mobile ad hoc routing protocols. Kamat et al. [8] introduced an idea of using fake routing sources. The fake sources are able to lead adversaries to fake locations. Kong and Hong [9] proposed ANDOR, an anonymous routing protocol, which suggests using route pseudonyms instead of node IDs during the routing process. ARM [13] is another anonymous routing protocol in which two nodes share a secret key and a pseudonym. Based on ANDOR and ARM, Taheri et al. [18] further proposed RDIS, an anonymous routing protocol with destination location privacy protection.

2.2   Location Privacy Protection on Social Networks

Location information on social networks is commonly treated as private data and is only available to a certain group of people. For instance, users of online social networks are able to configure their privacy settings so as to only reveal their locations to friends. Unfortunately, the majority of social network users only make changes of their default privacy settings after bad things occurred and the existing privacy settings are perplexed. Lipford et al. [10] presented a new privacy setting

interface based on Facebook that makes significantly improved understanding on the settings as well as better performance. The interface enables a set of HTML tabs, each providing a different browser's view of Facebook users' account information, in order to let the users properly manage their privacy settings. Ho et al. [7] analyzed the most popular social networks and discovered three privacy problems. First, users are not notified by social networks when their personal information is at privacy risks. Second, existing privacy protection tools in social networks are not flexible enough. Third, users cannot prevent information that may reveal the privacy of themselves from being uploaded by any other users. To solve the problems, Ho et al. [7] designed a privacy management framework, in which it enables data levels, privacy levels, and tracking levels, respectively. By designating data levels, users are able to sort their data into different privacy-related levels. Different amount of user data is available for different viewers, such as visitors, casual friends, normal friends, and best friends, according to their privacy levels. Finally, tracking levels can be applied to let users decide the magnitude of their account information that can be tracked. Content on social networks is usually tagged by its poster. This incurs privacy threats when location-related content is tagged via spatial-temporal tags. Freni et al. [5] summarized such threats as the location privacy and the absence privacy. The former one enables the possibility of referring other users' presence of specific locations at a given period of time. The latter one enables the possibility of referring other users' absence of specific locations at a given period of time. Freni et al. then introduced the generalization methods to prevent the privacy threats, where the methods are primarily space-based generalization and time-based generalization. Although all the aforementioned location privacy protection schemes have their advantages, to some extent, they are not applicable to protect location information revelation on user text messages. In this paper, we propose our location protection approach based on user text messages.

3   Our Approach

3.1   Preliminaries

Social network users frequently share text information with others. Abrol and Khan [1] presented a location information extraction approach on the basis of users' messages. In their approach, location information of a certain message is tagged and ranked based on its collision with their geo-location database entries. The one with the highest ranking score will be extracted. In our approach, we maintain a similar geo-location database, for the purpose of verifying location information's existence for a given user massage. We narrow down our geo-location information to city level in our database. Hence, for each entry of data, it includes *city*, *zip code*, *state*, and *country*. The three location information protection techniques provided in our approach are encryption, k-anonymity, and noise injection. Users are able to make choice of them based on their concerns. Encryption is recommended when a user wants to share her location information with designated individuals, who are the friends of the user. In this case, the entire user message will be encrypted by its receiver's public key. Both k-anonymity and noise injection techniques are suggested when a user wants to publicly post her message, in which it incorporates her location information.

In the rest of this subsection, we will formally elaborate some notions. Our geo-location database, *DB*, is a universe of location information of United States, specified to city level. Each data entry resembles as {*city*, *zip code*, *state*, *country*}. u stands for a user that the user's location information is tagged as $T(u)$. A user message is defined as m, where it is tagged as $T(m)$. By denoting $T(m) \subseteq T(u)$, we claim that m contains user location information. By denoting $|T(m) \in DB| \geq 0$, we claim that the tagged location information matches the database entries. For every user, she maintains a list of friends' public keys PK.

3.2 Encryption Approach

The encryption technique applies when a user wants to share her location information with her friends. Algorithm 1 is the pseudo code of the encryption algorithm.

Algorithm 1. The Encryption Approach

```
     Input: m
    Output: En(m)
 1: for every m do
 2:     T(m) = LocationTag(m)
 3: end for
 4: if (T(m) ∈ ∅ or T(m) ⊄ T(u)) then
 5:     return null
 6: else if (Encrypt = =1 and PK ∉ ∅) then
 7:     En(m) = {m}_PK
 8: else
 9:     return null
10: end if
```

In the algorithm, it takes user message m as its input and outputs the encrypted cyphertext, which is able to be decrypted by its receiver's private key. Location information of m is tagged by the $LocationTag$ function. The Encrypt is a Boolean variable. When its value equals one, it means the user enables the encryption technique. Otherwise, the encryption technique is disabled. This leaves flexibility to user when choosing location protection techniques.

3.3 K-anonymity Approach

K-anonymity technique is recommended when a user wants to publicly share her message, in which the message incorporates the user's location information.

Algorithm 2. The k-anonymity Approach

```
     Input: m
    Output: k_anon(m)
 1: for every m do
 2:     T(m) = LocationTag(m)
 3: end for
 4: if (T(m) ∈ ∅ or T(m) ⊄ T(u)) then
```

5:      **return** $null$
6: **else if** (ka = =1) **then**
7:      $k_{anon}(m) = k - anonymized(m)$
8: **end if**
9: **return** $k_{anon}(m)$

Algorithm 3. The k-anonymize algorithm

Input: $m, T(m)$
Output: $k_{anon}(m)$
1:**for** every $T(m) \in m$ **do**
2:      **if** $(|T(m) \in DB| \geq 2)$ **then**
3:          **return** $T(m)$
4:      **else** $T(m) \leftarrow *$
5:          **return** $T(m)$
6:      **end if**
7: **end for**
8: $k_{anon}(m) \leftarrow T(m)$
9: **return** $k_{anon}(m)$

Each data entry, in the DB, resembles as {*city*, *zip code*, *state*, *country*}. Since we only consider the geo-location information throughout United States, the data entry is able to be reduced to {*city*, *zip code*, *state*}, without losing any information. A totality tagging of a certain data entry resembles as $\{T_{city}, T_{state}, T_{zipcode}, T_{city,state,zipcode}\}$. It is widely known that a zip code is a unique identifier for a city. Therefore, the code in line 4 of the k-anonymity function is able to remove zip code tag for a given m. The collection of tags then becomes $\{T(city), T(state), T(city, state)\}$. The tag, $T(city, state)$, is a tagging of the Quasi-identifier of our DB. And this tag is able to be, at least, 2-anonymized via suppression, after the 4th step of the k-anonymity function. According to Sweeney [17], k-anonymity of a given message's location information is then able to be achieved as long as the Quasi-identifier of the location information is k-anonymized.

Given an example that the information, "Madison, SD, 57042", is tagged from a m. The k-anonymity algorithm first suppresses the zip code tag. The message transfers to "Madison, SD, *****". The Quasi-identifier tag then still uniquely matches our database entry, because there is only one city named Madison in South Dakota. Therefore, two ways of suppressions can be applied to anonymize the Quasi-identifier tag. One is to completely anonymize the tag, which it leads to "*******, **, *****". The alternative way is to anonymize the state tag in order to achieve city level anonymity.

3.4   Noise Injection Approach

The noise injection technique is aimed to protect user location information by injecting location noise, T_{noise}, in between $T(m)$. City level and state level anonymity can be achieved via noise injection. In the noise injection algorithm, we still leave user to decide which level's anonymity is most appropriate to her.

Algorithm 4. The Noise Injection Approach

Input: m

Output: $Noise_{injected}(m)$

1: **for** every m **do**
2:    $T(m) = LocationTag(m)$
3: **end for**
4: **if** $(T(m) \in \emptyset\ or\ T(m) \nsubseteq T(u))$ **then**
5:    **return** $null$
6: **else if** (ni $==1$) **then**
7:    $Noise_{injected}(m) = NoiseInject(m)$
8: **end if**
9: **return** $Noise_{injected}(m)$

Algorithm 5. The Noise-Inject algorithm

Input: $m, T(m)$

Output: $Noise_{injected}(m)$

1: **if** $(citylevel == 1)$ **then**
2:    **for** every $T(m) \in \{city\} \in DB$ **do**
3:        $T(m) \leftarrow T(m) \cup \{T_{noise}(city, zip\ code)\}$
4:    **end for**
5: **else if** $(statelevel == 1)$ **then**
6:    **for** every $T(m) \in \{city\} \cup \{zip\ code\} \in DB$ **do**
7:        $T(m) \leftarrow *$
8:    **end for**
9:    **for** every $T(m) \in \{state\} \in DB$ **do**
10:        $T(m) \leftarrow T(m) \cup \{T_{noise}(state)\}$
11:    **end for**
12:**else return** $null$
13:**end if**
14: $Noise_{injected}(m) \leftarrow T(m)$
15: **return** $Noise_{injected}(m)$

Recall that k-anonymity is achieved when $|T(m) \in DB| \geq 2$. Consider a $T(m) = \{T_{city}, T_{state}, T_{zipcode}\}$ and noise $T_{noise} = \{T_{n-city}, T_{n-zipcode}\}$, a $T(m)$ with injected noise is denoted as $\{T_{city}||T_{n-city}, T_{state}, T_{zipcode}||T_{n-zipcode}\}$. Therefore, a noise injected $T(m)$ matches at least two entries of the DB. Since noise city and zip code both belong to the same state as the original city and zip code, city level k-anonymity is accomplished that the original city is hidden with $1/k$ inferred possibility. For the state level, first, a $T(m)'s$ city and zip code tags are suppressed that only leaves state tag unsuppressed. Hence, state level k-anonymity is able to accomplish via injecting, at least, one noisy state. Similarly, the state level k-anonymity has $1/k$ inferred possibility.

4   Conclusion

Location privacy protection is an important research topic on mobile ad-hoc networks. However, location privacy protection on social networks has not been well studied. In this paper, we present our approach for the protection of location information based on social network user text messages. Three techniques including encryption, k-anonymity, and noise injection, are provided. We grant the right of selecting the techniques by users. Our experimental results show that our approach can significantly reduce location privacy revelation and can be efficiently deployed.

Even if our location protection approach is able to well performed, there are still some limitations haunting on it. First, the geo-location database has shortcomings. Currently, we only collected city level geo-locations in United States. However, social network users are scattered around the world, which requires us to expand the database entries in order to protect users outside of U.S. Additionally, we also need to expand our database to include any other geo-location information lower than city level. For instance, if a user message mentions locations like district, street, or buildings, our current database is then not able to deal with this situation. Second, our approach grants users the flexibilities for selecting protection techniques. On the other hand, it also impose burden such as the understanding of protection techniques to users. As our future work, we will integrate linguistic analysis towards user messages to allow system to automatically select the best protection technique for each user.

References

1. Abrol, S., Khan, L.: TweetHood: Agglomerative Clustering on Fuzzy k-Closest Friends with Variable Depth for Location Mining. In: Proceedings of the IEEE International Conference on Privacy, Security, Risk and Trust, Minneapolis, MN, USA, pp. 153–160 (August 2010)
2. Amoli, A., Kharrazi, M., Jalili, R.: 2Ploc: Preserving Privacy in Location-Based Services. In: Proceedings of the IEEE International Conference on Privacy, Security, Risk and Trust, Minneapolis, MN, USA, pp. 707–712 (August 2010)
3. Blakely, R.: Does Facebook's Privacy Policy Stack Up?,
   `http://business.timesonline.co.uk/tol/business/`
   `industry_sectors/technology/article2430927.ece`
4. Facebook Statistics,
   `http://www.facebook.com/press/info.php?statistics`
5. Freni, D., Vicente, C., Mascetti, S., Bettini, C., Jensen, C.: Preserving Location and Absence Privacy in Geo-Social Networks. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, Ontario, Canada (October 2010)
6. Gross, R., Acquisti, A.: Information Revelation and Privacy in Online Social Networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, Alexandria, Virginia, USA, pp. 71–80 (November 2005)
7. Ho, A., Maiga, A., Aimeur, E.: Privacy Protection Issues in Social Networking Sites. In: Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, Rabat, Morocco, pp. 271–278 (May 2009)

8. Kamat, P., Zhang, Y., Trappe, W., Ozturk, C.: Enhancing Source-Location Privacy in Sensor Network Routing. In: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems, Columbus, Ohio, USA (June 2005)
9. Kong, J., Hong, X.: ANODR: Anonymous on Demand Routing with Untraceable Routes for Mobile ad-hoc Networks. In: Proceedings of the 4th ACM International Symposium on Mobile ad hoc Networking and Computing, New York, NY, USA, June 2003, pp. 291–302 (2003)
10. Lipford, H., Besmer, A., Watson, J.: Understanding Privacy Settings in Facebook with an Audience View. In: Proceedings of the 1st Conference on Usability, Psychology, and Security, San Francisco, USA, pp. 1–8 (April 2008)
11. Rosenblum, D.: What Anyone Can Know: The Privacy Risks of Social Networking Sites. IEEE Security and Privacy 5(3), 40–49 (2007)
12. Schilit, B., Hong, J., Gruteser, M.: Wireless Location Privacy Protection. Computer 36(12), 135–137 (2003)
13. Seys, S., Preneel, B.: ARM: Anonymous Routing Protocol for Mobile ad hoc Networks. International Journal of Wireless and Mobile Computing 3(3) (October 2009)
14. Sogholan, C.: The Next Facebook Privacy Scandal, http://news.cnet.com/8301-13739_3-9854409-46.html
15. Strater, K., Lipford, H.: Strategies and Struggles with Privacy in an Online Social Networking Community. In: Proceedings of the 22nd British HCI Group Annual Conference, Liverpool, UK, pp. 111–119 (September 2008)
16. Stross, R.: When Everyone's a Friend, Is Anything Private?, http://www.nytimes.com/2009/03/09/business/worldbusiness/ 09iht-08digi.20688637.html
17. Sweeney, L.: k-anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10(5), 557–570 (2002)
18. Taheri, S., Hartung, S., Hogrefe, D.: Achieving Receiver Location Privacy in Mobile ad hoc Networks. In: Proceedings of the IEEE International Conference on Privacy, Security, Risk and Trust, Minneapolis, MN, USA, pp. 800–807 (August 2010)

Agent-Based Models of Complex Dynamical Systems of Market Exchange

Herbert Gintis

The Santa Fe Institute and Central European University
`hgintis@comcast.net`

Abstract. The standard Walrasian general equilibrium model is a static description of market clearing equilibria. Attempts over more than a century to provide a decentralized market dynamic that implements market equilibrium have failed. The reason is that a system of decentralized markets is a complex dynamical system in which the major form of learning is through experience (adaptive expectations) and imitating successful others (replicator dynamics).

I will present a model of decentralized market exchange where each individual produces one good and consumes many. I will show that an agent-based model of this economy converges to market equilibrium and is highly impervious to shocks. I will also present a model in which agents are firms and households, as in the standard Walrasian model. I will show that an agent-based model in this case converges strongly to market equilibrium, but with considerable excess volatility and large excursions from equilibrium. This is characteristic of error terms with "fat tails" as characterized in the complexity literature, and is due to the tendency of agents to imitate the successful, which leads to strongly correlated error distributions.

Computational and Statistical Models: A Comparison for Policy Modeling of Childhood Obesity

PANEL MEMBERS

Chair/Moderator:

Patricia L. Mabry, Ph.D.
Senior Advisor
Office of Behavioral and Social Sciences Research
Office of the Director
National Institutes of Health

Participants:

Ross Hammond, Ph.D.
Director, Center on Social Dynamics &
Policy
Senior Fellow, Economic Studies
Brookings Institution

Edward Hak-Sing Ip , Ph.D.
Professor and Associate Chair of
Research & Faculty Development
Division of Public Health Sciences
Wake Forest University Health
Sciences

Terry T-K Huang, PhD, MPH
Professor and Chair
Department of Health Promotion &
Social and Behavioral Health
College of Public Health
University of Nebraska Medical Center

Abstract. As systems science methodologies have begun to emerge as a set of innovative approaches to address complex problems in behavioral, social science, and public health research, some apparent conflicts with traditional statistical methodologies for public health have arisen. Computational modeling is an approach set in context that integrates diverse sources of data to test the plausibility of working hypotheses and to elicit novel ones. Statistical models are reductionist approaches geared towards proving the null hypothesis. While these two approaches may seem contrary to each other, we propose that they are in fact complementary and can be used jointly to advance solutions to complex problems. Outputs from statistical models can be fed into computational models, and outputs from computational models can lead to further empirical data collection and statistical models. Together, this presents an iterative process that refines the models and contributes to a greater understanding of the problem and its potential solutions. The purpose of this panel is to foster communication and understanding between statistical and computational modelers. Our goal is to shed light on the differences between the approaches and convey what kinds of research inquiries each one is best for addressing and how they can serve complementary (and synergistic) roles in the research process, to mutual benefit. For each

approach the panel will cover the relevant "assumptions" and how the differences in what is assumed can foster misunderstandings. The interpretations of the results from each approach will be compared and contrasted and the limitations for each approach will be delineated. We will use illustrative examples from CompMod, the Comparative Modeling Network for Childhood Obesity Policy. The panel will also incorporate interactive discussions with the audience on the issues raised here.

Detecting Changes in Opinion Value Distribution for Voter Model

Kazumi Saito[1], Masahiro Kimura[2], Kouzou Ohara[3], and Hiroshi Motoda[4]

[1] School of Administration and Informatics, University of Shizuoka
k-saito@u-shizuoka-ken.ac.jp
[2] Department of Electronics and Informatics, Ryukoku University
kimura@rins.ryukoku.ac.jp
[3] Department of Integrated Information Technology, Aoyama Gakuin University
ohara@it.aoyama.ac.jp
[4] Institute of Scientific and Industrial Research, Osaka University
motoda@ar.sanken.osaka-u.ac.jp

Abstract. We address the problem of detecting the change in opinion share over a social network caused by an unknown external situation change under the value-weighted voter model with multiple opinions in a retrospective setting. The unknown change is treated as a change in the value of an opinion which is a model parameter, and the problem is reduced to detecting this change and its magnitude from the observed opinion share diffusion data. We solved this problem by iteratively maximizing the likelihood of generating the observed opinion share, and in doing so we devised a very efficient search algorithm which avoids parameter value optimization during the search. We tested the performance using the structures of four real world networks and confirmed that the algorithm can efficiently identify the change and outperforms the naive method, in which an exhaustive search is deployed, both in terms of accuracy and computation time.

Keywords: Voter Model, Opinion Share, Change Point Detection.

1 Introduction

Recent technological innovation in the web such as blogosphere and knowledge/media-sharing sites is remarkable, which has made it possible to form various kinds of large social networks, through which behaviors, ideas and opinions can spread, and our behavioral patterns are strongly affected by the interaction with these networks. Thus, substantial attention has been directed to investigating the spread of influence in these networks [9,2,14].

Much of the work has treated information as one entity and nodes in the network are either active (influenced) or inactive (uninfluenced), i.e. there are only two states. However, application such as an on-line competitive service in which a user can choose one from multiple choices/decisions requires a model that handles multiple states. In addition, it is important to consider the value of each choice, e.g., quality, brand, authority, etc. because this impacts our choice. We formulated this problem using a value-weighted K opinion diffusion model and provided a way to accurately predict

the expected share of the opinions at a future target time from a limited amount of observed data [6]. This model is an extension of the basic voter model which is based on the assumption that a person changes its opinion by the opinions of its neighbors. There has been a variety of work on the voter model. Dynamical properties of the basic model have been extensively studied including how the degree distribution and the network size affect the mean time to reach consensus [10,12]. Several variants of the voter model are also investigated and non equilibrium phase transition is analyzed [1,15]. Yet another line of work extends the voter model by combining it with a network evolution model [3,2].

These studies are different from what we address in this paper. Almost all of the work so far on information diffusion assumed that the model is stationary. However, our behavior is affected not only by the behaviour of our neighbors but also by other external factors. We apply our voter model to detect a change in opinion share which is caused by an unknown external situation change. We model the change in the external factors as a change in the opinion value, and try to detect the change from the observed opinion share diffusion data. If this is possible, this would bring a substantial advantage. We can detect that something unusual happened during a particular period of time by simply analyzing the data. Note that our approach is retrospective, i.e. we are not predicting the future, but we are trying to understand the phenomena that happened in the past, which shares the same spirit of the work by Kleinberg [7] and Swan [13] in which they tried to organize a huge volume of the data stream and extract structures behind it.

Thus, our problem is reduced to detecting where in time and how long this change persisted and how big this change is. To make the analysis simple, we limit the form of the value change to a rect-linear one, that is, the value changes to a new higher level, persists for a certain period of time and is restored back to the original level and stays the same thereafter. We call this period when the value is high as "hot span" and the rest as "normal span". We use the same parameter optimization algorithm as in [6], i.e. the parameter update algorithm based on the Newton method which globally maximizes the likelihood of generating the observed data sequences. The problem here is more difficult because it has another loop to search for the hot span on top of the above loop. The naive learning algorithm has to iteratively update the patten boundaries (outer loop) and the value must also be optimized for each combination of the pattern boundaries (inner loop), which is extraordinary inefficient. We devised a very efficient search algorithm which avoids the inner loop optimization during the search. We tested the performance using the structures of four real world networks (blog, Wikipedia, Enron and coauthorship), and confirmed that the algorithm can efficiently identify the hot span correctly as well as the opinion value. We further compared our algorithm with the naive method that finds the best combination of change boundaries by an exhaustive search through a set of randomly selected boundary candidates, and showed that the proposed algorithm far outperforms the native method both in terms of accuracy and computation time.

2   Opinion Formation Models

The mathematical model we use for the diffusion of opinions is the value-weighted voter model with K (≥ 2) opinions [6]. A social network is represented by an undirected

(bidirectional) graph with self-loops, $G = (V, E)$, where V and E ($\subset V \times V$) are the sets of all the nodes and links in the network, respectively. For a node $v \in V$, let $\Gamma(v)$ denote the set of neighbors of v in G, that is, $\Gamma(v) = \{u \in V; (u, v) \in E\}$. Note that $v \in \Gamma(v)$.

In the model, each node of G is endowed with $(K + 1)$ states; opinions $1, \cdots, K$, and *neutral* (i.e., no-opinion state). It is assumed that a node never switches its state from any opinion k to neutral. The model has a parameter w_k (> 0) for each opinion k, which is called the *value-parameter* and must be estimated from observed opinion diffusion data. Let $f_t : V \to \{0, 1, 2, \cdots, K\}$ denote the opinion distribution at time t, where $f_t(v)$ stands for the opinion of node v at time t, and opinion 0 denotes the neutral state. We also denote by $n_k(t, v)$ the number of v's neighbors that hold opinion k at time t for $k = 1, 2, \cdots, K$, i.e., $n_k(t, v) = |\{u \in \Gamma(v); f_t(u) = k\}|$. Given a target time T, and an initial state in which each opinion is assigned to only one distinct node and all other nodes are in the neutral state, the evolution process of the model unfolds in the following way. In general, each node v considers changing its opinion based on the current opinions of its neighbors at its $(j-1)$th update-time $t_{j-1}(v)$, and actually changes its opinion at the jth update-time $t_j(v)$, where $t_{j-1}(v) < t_j(v) \le T$, $j = 1, 2, 3, \cdots$, and $t_0(v) = 0$. It is noted that since node v is included in its neighbors by definition, its own opinion is also reflected. The jth update-time $t_j(v)$ is decided at time $t_{j-1}(v)$ according to the exponential distribution of parameter λ (we simply use $\lambda = 1$ for any $v \in V$)[1]. Then, node v changes its opinion at time $t_j(v)$ as follows: If node v has at least one neighbor with some opinion at time $t_{j-1}(v)$, $f_{t_j(v)}(v) = k$ with probability $w_k n_k(t_{j-1}(v), v)$ / $\sum_{k'=1}^{K} w_{k'} n_{k'}(t_{j-1}(v), v)$ for $k = 1, \cdots, K$, otherwise, $f_{t_j(v)}(v) = 0$ with probability 1. Note here that $f_t(v) = f_{t_{j-1}(v)}(v)$ for $t_{j-1}(v) \le t < t_j(v)$. If the next update-time $t_j(v)$ passes T, that is, $t_j(v) > T$, then the opinion evolution of v is over. The evolution process terminates when the opinion evolution of every node in G is over.

Given the observed opinion diffusion data $\mathcal{D}(T_s, T_e) = \{(v, t, f_t(v))\}$ in time-interval $[T_s, T_e]$ (a single example), we consider estimating the values of value-parameters $w_1, \cdots, w_K$, where $0 \le T_s < T_e \le T$. From the evolution process of the model, we can obtain the following log likelihood function

$$\mathcal{L}(\boldsymbol{w}; \mathcal{D}(T_s, T_e)) = \log \prod_{(v,t,k) \in C(T_s, T_e)} \frac{n_k(t, v) w_k}{\sum_{k'=1}^{K} n_{k'}(t, v) w_{k'}}, \tag{1}$$

where $\boldsymbol{w} = (w_1, \cdots, w_K)$ stands for the K-dimensional vector of value-parameters, and $C(T_s, T_e) = \{(v, t, f_t(v)) \in \mathcal{D}(T_s, T_e); |\{u \in \Gamma(v); f_t(u) \ne 0\}| \ge 2\}$. Thus, our estimation problem is formulated as a maximization problem of the objective function $\mathcal{L}(\boldsymbol{w}; \mathcal{D}(T_s, T_e))$ with respect to $\boldsymbol{w}$. We find the optimal values of $\boldsymbol{w}$ by employing a standard Newton method (see [6] for more details).

3   Change Detection Problem

We investigate the problem of detecting the change in behavior of opinion diffusion in a social network G based on the value-weighted voter model with K opinions, which is referred to as the *change detection problem*. In this problem, we assume that some

---

[1] Note that this is equivalent to picking a node randomly and updating its opinion in turn $|V|$ times.

change has happened in the way the opinions diffuse, and we observe the opinion diffusion data in which the change is embedded, and consider detecting where in time and how long this change persisted and how big this change is.

Here, we mathematically formulate the change detection problem. For the opinion diffusion data $\mathcal{D}(0, T)$ in time-interval $[0, T]$, let $[T_1, T_2]$ denote the hot (change) span of the diffusion of opinions. This implies that the intervals $[0, T_1)$ and $(T_2, T]$ are the normal spans. Let $\boldsymbol{w}_n$ and $\boldsymbol{w}_h$ denote the value-parameter vectors for the normal span and the hot span, respectively. Note that $\boldsymbol{w}_n/\|\boldsymbol{w}_n\| \neq \boldsymbol{w}_h/\|\boldsymbol{w}_h\|$ since the opinion dynamics under the value-weighted voter model is invariant to positive scaling of the value-parameter vector $\boldsymbol{w}$, where $\|\boldsymbol{w}_n\|$ and $\|\boldsymbol{w}_h\|$ stand for the norm of vectors $\boldsymbol{w}_n$ and $\boldsymbol{w}_h$. Then, the change detection problem is formulated as follows: Given the opinion diffusion data $\mathcal{D}(0, T)$ in time-interval $[0, T]$, detect the anomalous span $[T_1, T_2]$, and estimate the value-parameter vector $\boldsymbol{w}_h$ of the hot span and the value-parameter vector $\boldsymbol{w}_n$ of the normal span.

Since the value-weighted voter model is a stochastic process model, every sample of opinion diffusion can behave differently. This means that it is quite difficult to accurately detect the true hot span from only a single sample of opinion diffusion. Methods that use only the observed bursty activities, including those proposed by Swan and Allan [13] and Kleinberg [7] would not work. We believe that an explicit use of underlying opinion diffusion model is essential to solve this problem. It is crucially important to detect the hot span precisely in order to identify the external factors which caused the behavioral changes.

4   Detection Methods

4.1   Naive Method

Let $\mathcal{T} = \{t_1, \cdots, t_N\}$ be a set of opinion change time points of all the nodes appearing in the diffusion results $\mathcal{D}(0, T)$. We can consider the following value-parameter vector switching when there is a hot span $S = [T_1, T_2]$:

$$\boldsymbol{w} = \begin{cases} \boldsymbol{w}_n & \text{if } t \in \mathcal{T} \setminus S, \\ \boldsymbol{w}_h & \text{if } t \in \mathcal{T} \cap S. \end{cases}$$

Then, an extended objective function $\mathcal{L}(\boldsymbol{w}_n, \boldsymbol{w}_h; \mathcal{D}(0, T), S)$ can be defined by adequately modifying Equation (1) under this switching scheme. Clearly, the extended objective function is expected to be maximized by setting S to be the true span $S^* = [T_1^*, T_2^*]$, for which $\mathcal{D}(0, T)$ is generated by the value-weighted voter model, provided that $\mathcal{D}(0, T)$ is sufficiently large. Therefore, our hot span detection problem is formalized as the following maximization problem.

$$\hat{S} = \arg\max_S \mathcal{L}(\hat{\boldsymbol{w}}_n, \hat{\boldsymbol{w}}_h; \mathcal{D}(0, T), S), \tag{2}$$

where $\hat{\boldsymbol{w}}_n$ and $\hat{\boldsymbol{w}}_h$ denote the maximum likelihood estimators for a given S.

In order to obtain $\hat{S}$ according to Equation (2), we need to prepare a reasonable set of candidate spans, denoted by $\mathcal{S}$. One way of doing so is to construct $\mathcal{S}$ by considering all pairs of observed activation time points. Then, we can construct a set of candidate

spans by $S = \{S = [t_1, t_2] : t_1 < t_2, t_1 \in \mathcal{T}, t_2 \in \mathcal{T}\}$. Equation (2) can be solved by a naive method which has two iterative loops. In the inner loop we first obtain the maximum likelihood estimators, $\hat{w}_n$ and $\hat{w}_h$, for each candidate S by maximizing the objective function $\mathcal{L}(w_n, w_h; \mathcal{D}(0, T), S)$ using the Newton method. In the outer loop we select the optimal $\hat{S}$ which gives the largest $\mathcal{L}(\hat{w}_n, \hat{w}_h; \mathcal{D}(0, T), S)$ value. However, this method can be extremely inefficient when the number of candidate spans is large. Thus, in order to make it work with a reasonable computational cost, we consider restricting the number of candidate time points to a small value, denoted by J, i.e., we construct $\mathcal{T}_J = \{t_1, \cdots, t_J\}$ by selecting J points from $\mathcal{T}$; then we construct a restricted set of candidate spans by $\mathcal{S}_J = \{S = [t_1, t_2] : t_1 < t_2, t_1 \in \mathcal{T}_J, t_2 \in \mathcal{T}_J\}$. Note that $|\mathcal{S}_J| = J(J-1)/2$, which is large when J is large.

4.2   Proposed Method

It is easily conceivable that the naive method can detect the hot span with a reasonably good accuracy when we set J large at the expense of the computational cost, but the accuracy becomes poorer when we set J smaller to reduce the computational load. We propose a novel detection method below which alleviates this problem and can efficiently and stably detect a hot span from diffusion results $\mathcal{D}(0, T)$.

We first obtain the maximum likelihood estimators, $\hat{w}$ based on the original objective function of Equation (1), and focus on the first-order derivative of the objective function $\mathcal{L}(w; \mathcal{D}(0, T))$ with respect to the value-parameter vector w at each individual opinion change time. More specifically, let w_t be the value-parameter vector at time $t \in \mathcal{T}$. Then we obtain the following formula for the maximum likelihood estimators due to the uniform parameter setting and the globally optimal condition.

$$\frac{\partial \mathcal{L}(\hat{w}; \mathcal{D}(0, T))}{\partial w} = \sum_{t \in \mathcal{T}} \frac{\partial \mathcal{L}(\hat{w}; \mathcal{D}(0, T))}{\partial w_t} = 0. \tag{3}$$

Now, we can consider the following partial sum for a given hot span $S = [T_1, T_2]$.

$$g(S) = \sum_{t \in \mathcal{T} \cap S} \frac{\partial \mathcal{L}(\hat{w}; \mathcal{D}(0, T))}{\partial w_t}. \tag{4}$$

Clearly, $\|g(S)\|$ is likely to have a sufficiently large positive value if $S \approx S^*$ due to our problem setting. Namely, the hot span is detected as follows:

$$\hat{S} = \arg\max_{S \in \mathcal{S}} \|g(S)\|. \tag{5}$$

Here note that we can incrementally calculate $g(S)$. More specifically, let $\mathcal{T} = \{t_1, \cdots, t_N\}$ be a set of candidate time points, where $t_i < t_j$ if $i < j$; then, we can obtain the following formula.

$$g([t_i, t_{j+1}]) = g([t_i, t_j]) + \frac{\partial \mathcal{L}(\hat{w}; \mathcal{D}(0, T))}{\partial w_{t_{j+1}}}. \tag{6}$$

The computational cost of the proposed method for examining each candidate span is much smaller than the naive method described above. When $|\mathcal{T}| = N$ is very large, we construct a restricted set of candidate spans $\mathcal{S}_J$ as explained above. We summarize our proposed method below.

1. Maximize $\mathcal{L}(\boldsymbol{w}; \mathcal{D}(0, T))$ by using the Newton method.
2. Construct the candidate time set $\mathcal{T}$ and the candidate span set $\mathcal{S}$.
3. Detect a hot span $\hat{S}$ by Equation (5) and output $\hat{S}$.
4. Maximize $\mathcal{L}(\boldsymbol{w}_n, \boldsymbol{w}_h; \mathcal{D}(0, T), \hat{S})$ by using the Newton method, and output $(\hat{\boldsymbol{w}}_n, \hat{\boldsymbol{w}}_h)$.

Here note that the proposed method requires likelihood maximization by using the Newton method only twice.

5   Experimental Evaluation

We adopted four datasets of large real networks. They are all bidirectionally connected networks. The first one is a trackback network of Japanese blogs used in [5], which has $12,047$ nodes and $79,920$ directed links (the blog network). The second one is a network of people that was derived from the "list of people" within Japanese Wikipedia, used in [4], and has $9,481$ nodes and $245,044$ directed links (the Wikipedia network). The third one is a network derived from the Enron Email Dataset [8] by extracting the senders and the recipients and linking those that had bidirectional communications. It has $4,254$ nodes and $44,314$ directed links (the Enron network). The fourth one is a coauthorship network used in [11], which has $12,357$ nodes and $38,896$ directed links (the coauthorship network).

For each of these networks, we generated opinion diffusion results for three different values of K (the number of opinions), i.e., $K = 2$, 4, and 8, by choosing the top K nodes with respect to node degree ranking as the initial K nodes and simulating the model mentioned in section 2 from 0 to $T = 25$. We assumed that the value of all the opinions were initially 1.0, i.e. the value-parameters for all the opinions are 1.0 for the normal span, and further assumed that the value of the first opinion changed to double for a period of [10, 15], i.e. the value-parameter of the first opinion is 2.0 and the value-parameters of all the other opinions are 1.0 for the hot span. We then estimated the hot span and the value-parameters for both the spans (normal and hot) by the two methods (the proposed and the naive), and compared their accuracy and the computation time. We adopted $1,000$ as the value of J (the number of candidate time points) for the proposed method, and 5, 10, and 20 for the naive method.

Figures 1 and 2 show the experimental results[2] where each value is the average over 10 trials for 10 distinct diffusion results. We evaluated the accuracy of the estimated hot span $[\hat{T}_1, \hat{T}_2]$ by the absolute error $|\hat{T}_1 - T_1| + |\hat{T}_2 - T_2|$, and the accuracy of the estimated opinion values $\hat{\boldsymbol{w}}$ by the mean absolute error $\Sigma_{i=1}^{K}(|\hat{w}_{in} - w_{in}| + |\hat{w}_{ih} - w_{ih}|)/K$, where w_{in} and w_{ih} are values of opinion i for the normal and the hot spans, respectively.

From these results, we can find that the proposed method is much more accurate than the naive method for both the networks. The average error for the naive method decreases as J becomes larger. But, even in the best case for the naive method ($J = 20$), its average error in the estimation of the hot span is maximum about 30 times larger than that of proposed method (in the case of the Enron network under $K = 2$), and it is maximum about 6 times larger in the estimation of value-parameters (in the case of the coauthorship network under $K = 2$). It is noted that the naive method needs much

---

[2] We only show the results for the two networks (Enron and coauthorship) due to the space limitation. In fact, we obtained similar results also for the other two networks (blog and Wikipedia).

(a) Error of $[\hat{T}_1, \hat{T}_2]$ (b) Error of $\hat{w}$ (c) Computation time
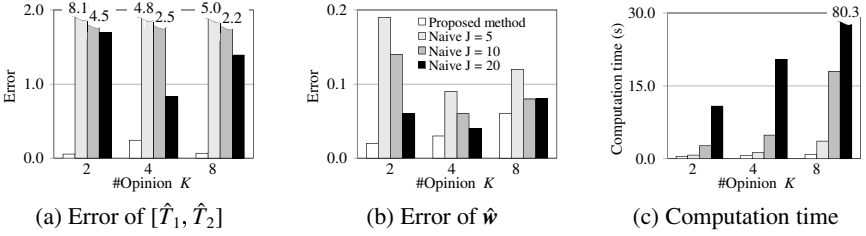
Fig. 1. Comparison on the Enron network

(a) Error of $[\hat{T}_1, \hat{T}_2]$ (b) Error of $\hat{w}$ (c) Computation time
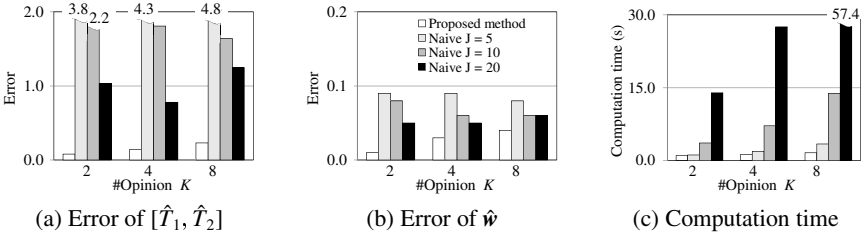
Fig. 2. Comparison on the coauthorship network

longer computation time to achieve these best accuracies than the proposed method although the number of candidate time points for the naive method is 50 times smaller. Indeed, it is about 20 times longer for the former case, about 13 times longer for the latter case, and maximum about 95 times longer for the whole results (in the case of the Enron network under $K = 8$). From these results, it can be concluded that the proposed method is able to detect and estimate the hot span and value-parameters much more accurately and efficiently compared with the naive method.

6 Conclusions

In this paper, we addressed the problem of detecting the unusual change in opinion share from the observed data in a retrospective setting, assuming that the opinion share evolves by the value-weighted voter model with multiple opinions. We defined the hot span as the period during which the value of an opinion is changed to a higher value than the other periods which are defined as the normal spans. A naive method to detect such a hot span would iteratively update the pattern boundaries that form a hot span (outer loop) and iteratively update the opinion value for each hot span candidate (inner loop) such that the likelihood function is maximized. This is very inefficient and totally unacceptable. We developed a novel method that avoids the inner loop optimization during search. It only needs to estimate the value twice by the iterative updating algorithm (Newton method), which can reduce the computation times by 7 to 95 times, and is very efficient. We applied the proposed method to opinion share samples generated from four real world large networks and compared the performance with the naive method that considers only the randomly selected boundary candidates. The results clearly indicate that the proposed method far outperforms the naive method both

in terms of accuracy and efficiency. Although we assumed a simplified problem setting in this paper, the proposed method can be easily extended to solve more intricate problems. As the future work, we plan to extend this framework to spatio-temporal hot span detection problems.

Acknowledgments

References

1. Castellano, C., Munoz, M.A., Pastor-Satorras, R.: Nonlinear q-voter model. Physical Review E 80, 041129 (2009)
2. Crandall, D., Cosley, D., Huttenlocner, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and sociai infuence in online communities. In: Proceedings of KDD 2008, pp. 160–168 (2008)
3. Holme, P., Newman, M.E.J.: Nonequilibrium phase transition in the coevolution of networks and opinions. Physical Review E 74, 056108 (2006)
4. Kimura, M., Saito, K., Motoda, H.: Minimizing the spread of contamination by blocking links in a network. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI 2008), pp. 1175–1180 (2008)
5. Kimura, M., Saito, K., Motoda, H.: Blocking links to minimize contamination spread in a social network. ACM Transactions on Knowledge Discovery from Data 3(9), 9:1–9:23 (2009)
6. Kimura, M., Saito, K., Ohara, K., Motoda, H.: Learning to predict opinion share in social networks. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010), pp. 1364–1370 (2010)
7. Kleinberg, J.: Bursty and hierarchical structure in streams. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2002), pp. 91–101 (2002)
8. Klimt, B., Yang, Y.: The enron corpus: A new dataset for email classification research. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 217–226. Springer, Heidelberg (2004)
9. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. In: Proceedings of the 7th ACM Conference on Electronic Commerce (EC 2006), pp. 228–237 (2006)
10. Liggett, T.M.: Stochastic interacting systems: contact, voter, and exclusion processes. Spriger, New York (1999)
11. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818 (2005)
12. Sood, V., Redner, S.: Voter model on heterogeneous graphs. Physical Review Letters 94, 178–701 (2005)
13. Swan, R., Allan, J.: Automatic generation of overview timelines. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000), pp. 49–56 (2000)
14. Wu, F., Huberman, B.A.: How public opinion forms. In: Papadimitriou, C., Zhang, S. (eds.) WINE 2008. LNCS, vol. 5385, pp. 334–341. Springer, Heidelberg (2008)
15. Yang, H., Wu, Z., Zhou, C., Zhou, T., Wang, B.: Effects of social diversity on the emergence of global consensus in opinion dynamics. Physical Review E 80, 046108 (2009)

Dynamic Simulation of Community Crime and Crime-Reporting Behavior

Michael A. Yonas, Jeffrey D. Borrebach, Jessica G. Burke,
Shawn T. Brown, Katherine D. Philp,
Donald S. Burke, and John J. Grefenstette

Graduate School of Public Health,
University of Pittsburgh,
Pittsburgh, PA 15261, USA
{may24,jdb108,jgburke,stb60,kdp5,donburke,gref}@pitt.edu
http://www.phdl.pitt.edu

Abstract. An agent-based model was developed to explore the effectiveness of possible interventions to reduce neighborhood crime and violence. Both offenders and non-offenders (or *citizens*) were modeled as agents living in neighborhoods, with a set of rules controlling changes in behavior based on individual experience. Offenders may become more or less inclined to actively commit criminal offenses, depending on the behavior of the neighborhood residents and other nearby offenders, and on their arrest experience. In turn, citizens may become more or less inclined to report crimes, based on the observed prevalence of criminal activity within their neighborhood. This paper describes the basic design and dynamics of the model, and how such models might be used to investigate practical crime intervention programs.

Keywords: computer simulation, agent-based model, neighborhood violence, crime prevention.

1 Introduction

While reductions in neighborhood violence have been observed, violent crime and victimization continue to be a critical issue of public safety and public health concern. The impact of violence disproportionately impacts young people living within low-income impoverished neighborhoods. For youth, violence is the second leading cause of death for youth ages 15-24 [1] and the leading cause of death for African American youth in this same age range [2]. Frequently, interventions to address community violence focus on individuals, whereas increased efforts to understand and illustrate the role of factors at all ecological levels are necessary for comprehensively addressing risk factors for neighborhood violence at the individual, family and neighborhood levels [3, 4]. Previous research has identified a variety of contextual factors such as collective efficacy, neighborhood disorganization, decreased social support, and limited resources associated with influencing neighborhood violent crime [5–8].

Many neighborhoods affected by violence have invested heavily in developing community-level intervention responses to neighborhood violence [1, 9]. Comprehensive community mobilization interventions and such as block watch programs have been shown to be effective to address crime and violence [10, 11]. Studies assessing factors influencing citizens and victims report violence and crime have noted that a substantial number of violent crimes, approximately 50%, go unreported. Reasons for failing to report include fear of retaliation, cultural constructions related to gender, and violence in the home, work or neighborhood [12]. While some community-level approaches have shown evidence of effectiveness, these approaches are often expensive, difficult to sustain and evaluate due to lack of causal data, and challenging to replicate [11].

Modeling and simulation have the ability to evaluate potential community intervention strategies at low cost and facilitate interdisciplinary collaboration. We utilize modeling and simulation to examine and evaluate potential community crime intervention strategies. In order to inform our exploratory agent-based model, we integrate key behavioral and community factors associated with neighborhood mobilization and watch programs. This paper describes the basic design and dynamics of the model, and describes how such models might be used to investigate practical neighborhood crime intervention programs.

Previous agent-based models (ABMs) have been developed to explore mitigation and dynamics of criminal activity. Epstein [13] constructed an early ABM of civil violence and rebellion with agents having levels of grievance against central authority determined by perceived-hardship and government legitimacy. The model produced a punctuated equilibrium, or periods of peace alternating with periods or rebellion. Groff [14] developed an ABM of street robbery crimes and found that inclusion of explicit geographic distributions reproduced patterns more similar to empirical patterns than other models. Furtado et al [15] utilized an ABM crime simulation including self-organizing behavior for criminals to model the swarm intelligence paradigm (i.e. most crime events in urban centers are concentrated in few places) to show that criminals' social networks are essential for them to aggregate into organized groups. SimDrugPolicing is an ABM developed by Dray et al [16] to analyze the effects of patrol, hotspot policing and problem-oriented policing law-enforcement strategies on a street-based illegal drug market. While these previous efforts provide valuable insights into the behavior of criminals and police, the model described here contributes a novel focus on the crime-reporting behavior of citizens in a community experiencing crime.

2   Methods

An ABM was designed to study the effects of change in crime-reporting behavior as a result of exposure to incidents of crime. Both offenders and non-offenders (or *citizens*) were modeled as agents living in neighborhoods, with a set of rules controlling changes in behavior based on individual experience. Offenders may become more or less inclined to actively commit offenses, depending on the behavior of the neighborhood residents and other offenders, and on their experience
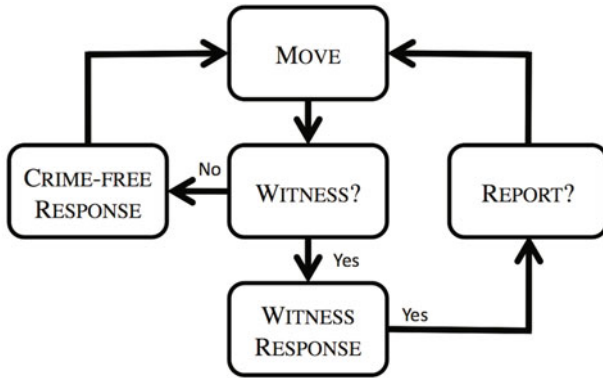
Fig. 1. Flow chart of the activity of a citizen, for each simulation time step

with the criminal justice system. Citizens may become more or less inclined to report offenses, based on the observed prevalence of offenses within their neighborhood. The model was implemented in NetLogo 4.1.

The community is represented by a two-dimensional toroidal grid; any box on the grid may be occupied by an *agent*. Agents move around within individually-defined neighborhoods and may observe the behavior of other agents in their immediate surroundings. In particular, agents may witness offenses being committed and observe whether or not the offenses are punished. The behavior of an agent is controlled by a few simple rules, described in the following sections. Agents are divided into two disjoint groups: *citizens*, who do not commit offenses, and *offenders*, who may commit crimes.

2.1   Citizen Model

Citizens in the model are agents who do not commit offenses, but who may affect the prevalence of crime through their actions. The decision flow chart for a citizen is shown in Fig. 1. At each time step, the citizen moves to a random location within an individually fixed local neighborhood. Citizens may witness an offense that occurs within a given distance of the citizen's current location. Upon witnessing a crime, the citizen experiences a *witness-response* that may alter future reporting behavior, and decides whether to report the crime to the police. If the citizen witnesses no crime during this time step, the citizen experiences a *crime-free-response*, which may also alter the citizens future reporting behavior.

We represent the reporting behavior of a citizen as a two-step process: witnessing a nearby crime, and then reporting it. These steps are modeled by two probabilities for each agent i, the agents *witness-probability* W_i and the agents *report-probability* R_i. If an offense occurs within a given distance of the citizen, then the citizen witnesses the offense with probability W_i. If the citizen witnesses an offense, the agent reports the crime with probability R_i. A citizen's
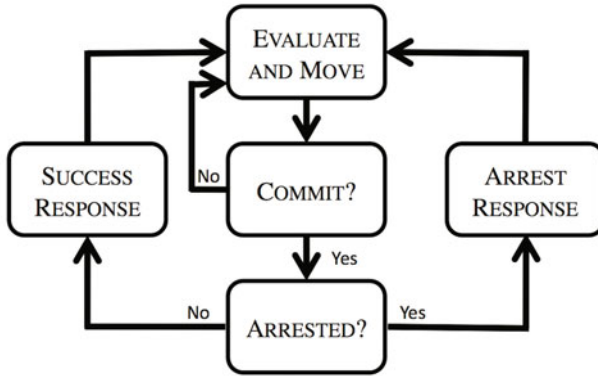
Fig. 2. Flow chart of the activity of an offender for each simulation time step

report-probability R_i can be affected by crime in the citizen's neighborhood. If a crime is witnessed, the citizen experiences a *witness-response*:

$$R_i = R_i * (1 - \alpha) + \alpha, \qquad \qquad \text{if} \quad \alpha > 0 \qquad \qquad (1)$$
$$R_i = R_i * (1 + \alpha), \qquad \qquad \text{if} \quad \alpha \leq 0$$

where α is the *witness response rate*. The effect of formula (1) is to raise R_i by α percent if α is positive or the decrease R_i by $|\alpha|$ percent if α is negative. The magnitude of α represents the tendency of a citizen to respond to an occurrence of crime by increasing the citizen's likelihood of reporting crimes. A negative value of α reflects a situation in which a citizen becomes less likely to report crimes, due to fear of reprisal or loss of hope. Alternatively, if no crime is witnessed, the citizen experiences a *crime-free-response*:

$$R_i = R_i * (1 - \beta) + \beta, \qquad \qquad \text{if} \quad \beta > 0 \qquad \qquad (2)$$
$$R_i = R_i * (1 + \beta), \qquad \qquad \text{if} \quad \beta \leq 0$$

where β is the *crime-free response rate*. The magnitude of β represents the tendency of a citizen to respond to an absence of crime by increasing the citizen's likelihood of reporting crimes. A negative value of β reflects a situation in which a citizen becomes less likely to report crimes, perhaps due to complacency.

2.2   Offender Model

Offenders in the model are agents who may commit criminal acts. The behavior of offenders during each simulation time step is shown in Fig. 2. Each offender moves around within the community seeking promising opportunities for crimes. If the current location appears attractive enough, the offender commits an offense. Upon committing an offense, the offender may be reported to the police by a witness. When an offender is reported, there is a fixed chance that the offender will be arrested.

Each offender i has a probability of offending, C_i. If offender i commits a crime, the agent's probability of offending is altered by whether or not the agent is arrested. For this model we assume that the probability of offending decreases when arrested; that is, the *arrest-response* is:

$$C_i = C_i * (1 - \gamma) \tag{3}$$

where $\gamma \geq 0$ is the *arrest response rate*. Alternatively, if an offender commits a crime but is not arrested, we assume that his probability of offending increases due to his successful criminal activity; that is, the *success response* is:

$$C_i = C_i * (1 - \delta) + \delta \tag{4}$$

where $\delta \geq 0$ is the *offender success response rate*.

The behavior of offenders is affected by their perception of the relative risks and opportunities in a given location. The model reflects this by associating with each location an *opportunity index* $Op(j)$, defined as an estimate of the probability of committing an crime at location j and not being arrested. At each time step, each offender moves to a adjacent neighborhood location j^* with the maximum value of $Op(j^*)$ among all neighboring locations. Once the offender moves to location j^*, the decision to commit a crime is based on the comparison between the perceived opportunity at that location and the offender's probability of offending C_i: Offender i commits a crime at current location j if $Op(j) \leq C_i$. As a result, offenders tend to migrate toward those areas of the community associated with high rates of unreported criminal activity.

3 Results

The model was developed as a platform for exploring the effects of altering citizen behavior through community intervention programs. To this end, we performed an initial survey of the effects of altering citizen reporting-behavior in the model by systematically altering the witness response rate α and the crime-free response rate β. Witness response α might be increased through interventions such as decreasing tolerance for neighborhood crime through the removal of graffiti, and making it easier to report crimes by providing neighborhood phone-trees (allowing citizens to call a neighbor who reliably contacts the police). The crime-free response β might be increased by programs targeted to raise expectations of a crime-free neighborhood and decrease citizen complacency, such as encouraging citizens to spend more time outdoors communicating with neighbors.

A series of simulations was run, varying α from -0.2 to 0.2 and varying β from -0.02 to 0.02. (Initial studies showed that the results of the model do not vary significantly outside these parameter ranges). The values of fixed parameters are shown in Table 1.

For each scenario, 10 replications of the model with identical initial conditions were performed. Figure 3 shows mean results for (a) total crime, (b) proportion

Table 1. Fixed Parameters for Crime ABM Simulation

Parameter	Value
Number of Citizens	900
Number of Offenders	100
Number of Locations	10,000
Witness Probability (W_i)	U(0,0.5)
Initial Report Probability (R_i)	U(0,1)
Initial Probability of Offending (C_i)	U(0,1)
Probability of Arrest when Reported	0.5
Arrest Response Rate (γ)	0.1
Offender Success Response Rate (δ)	0.5
Jail Term	U(1,30) days
Length of Simulation	365 days

U(a,b): uniform distribution between a and b

of the community that has been crime-free for 30 days, and (c) number of active citizens (defined as citizens whose probability of reporting a crime is greater than 50%).

The results show some expected patterns. For example, total crime is minimized in the model when large positive values of α and β are observed, which represent citizens becoming more active reporters regardless of changes in crime prevalence in their neighborhood. Perhaps a more surprising result occurs when $\alpha = 0.2$ and $\beta = -0.02$ (lowest corner in the plots). These values correspond to scenarios in which citizens become more likely to report when crime occurs, but
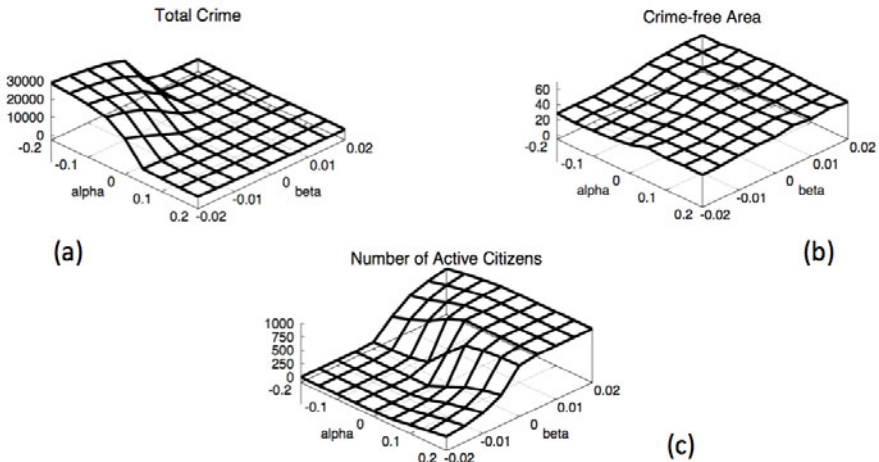
Fig. 3. Summary results over variations in values for α and β. (a) Total crime. (b) Final proportion of crime-free area, over variations in values for α and β. (c) Final number if active citizens.

less likely to report when the neighborhood becomes crime-free. This scenario yields an expected low total crime (Fig 3a), relatively high crime-free area (Fig 3b) and interestingly requires surprisingly few citizens who are likely to report crime (Fig 3c). Assuming that the overall cost of an intervention program may be correlated with the number of citizens whose reporting behavior needs to change, this scenario provides support for community watch crime prevention programs.

4 Discussion

The current model provides useful insights into which interventions at the neighborhood level might be most effective in reducing the presence of crime in a community. Future studies will elaborate the model to reflect more detailed behavior for both citizens and offenders (e.g., retaliatory behavior). For example, in developing this formative model, the arrest probability was fixed at 50%. In future models reflecting real-world settings, α, β, γ, and δ are likely to be functions of the probability of citizen action and arrest. Future models will also include explicit elements of the law enforcement system (e.g., increasing police presence in neighborhoods experiencing high crime rates).

Of course, it is recognized that computational models cannot reflect the complexity of actual communities and of individual responses to crimes within a community. Nonetheless, such models provide an additional set of tools for generating inclusive dialogue to inform the development of intervention policies which equitably integrate interdisciplinary expertise of neighborhood residents, law enforcement, policy makers and researchers. As more realistic computational models are developed, it may be possible to virtually evaluate a broad range of possible interventions with unique multi-level characteristics, and potentially identify those that may be likely to be successful in preventing community crime.

References

1. Center for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS), `www.cdc.gov/injury/wisqars/` (accessed 2010 June 14)
2. Center for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS), Youth Violence Data Sheet, `www.cdc.gov/violenceprevention/pdf/YV-DataSheet-a.pdf` (accessed 2010 June 14)
3. Kellerman, A.L., Fuqua-Whitley, D.S., Rivara, F.P., Mercy, J.: Preventing Youth Violence: What Works? Annual Review of Public Health 19, 271–292 (1998)
4. National Youth Violence Prevention Resource Center (NYVPRC), `www.safeyouth.gov/Resources/Prevention/Pages/PreventionStrategies.aspx`

5. Anderson, E.: Code of the Street: Decency, Violence and the Moral Life of the Inner City. Norton, New York (1999)
6. Earls, F.J.: Violence and Todays Youth: Critical Health Issues for Children and Youth. Future of Children 4(3), 4–23 (1994)
7. Sampson, R.J., Raudenbush, S.W., Earls, F.: Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. Science 277, 918–924 (1997)
8. Taylor, R.B., Gottfredson, S.D., Brower, S.: Understanding Block Crime and Fear. Journal of Research in Crime and Delinquency 21, 303–331 (1984)
9. Wilson, J.M., Chermak, S., McGarrell, E.F.: Community-Based Violence Prevention: An Assessment of Pittsburgh's One Vision One Life Program. RAND Corp., Santa Monica (2010)
10. Bibb, M.: Gang Related Services of Mobilization for Youth. In: Klein, M.W. (ed.) Juvenile Gangs in Context: Theory, Research, and Action. Prentice-Hall, Englewood Cliffs (1967)
11. Bennett, T.H., Holloway, K.R., Farrington, D.P.: Effectiveness of Neighbourhood Watch in Reducing Crime. National Council on Crime Prevention, Stockholm (2008)
12. Reiss, A.J., Roth, J.A.: Measuring Violent Crime and Their Consequences. In: Understanding and Preventing Violence. National Research Council, pp. 404–429. National Academy Press, Washington DC (1993)
13. Epstein, J.: Modeling Civil Violence: An Agent-based Computational Approach. Proceedings of the National Academy of Sciences of the United States of America 99 (suppl. 3), 7243–7250 (2002)
14. Groff, E.: Characterizing the Spatio-temporal Aspects of Routine Activities and the Geographical Distribution of Street Robbery. In: Liu, L., Eck, J. (eds.) Artificial Crime Analysis Systems, pp. 226–251. IGI Global, Hershey (2008)
15. Furtado, V., Melo, A., Coelho, A.L.V., Menezes, R., Belchio, M.: Simulating Crime against Properties using Swarm Intelligence and Social Networks. In: Liu, L., Eck, J. (eds.) Artificial Crime Analysis Systems, pp. 300–318. IGI Global, Hershey (2008)
16. Dray, A., Mazerolle, L., Perez, P., Ritter, A.: Drug Law Enforcement in an Agent-Based Model: Simulating the Disruption. In: Liu, L., Eck, J. (eds.) Artificial Crime Analysis Systems, pp. 352–371. IGI Global, Hershey (2008)

Model Docking Using Knowledge-Level Analysis

Ethan Trewhitt, Elizabeth Whitaker, Erica Briscoe, and Lora Weiss

Georgia Tech Research Institute, 250 14th Street NW, Atlanta, GA 30332-0832
{ethan.trewhitt,elizabeth.whitaker,erica.briscoe,
lora.weiss}@gtri.gatech.edu

Abstract. This paper presents an initial approach for exploring the docking of social models at the knowledge level. We have prototyped a simple blackboard environment allowing for model docking experimentation. There are research challenges in identifying which models are appropriate to dock and the concepts that they should exchange to build a richer multi-scale view of the world. Our early approach includes docking of societal system dynamics models with individual and organizational behaviors represented in agent-based models. Case-based models allow exploration of historical knowledge by other models. Our research presents initial efforts to attain opportunistic, asynchronous interactions among multi-scale models through investigation and experimentation of knowledge-level model docking. A docked system can supply a multi-scale modeling capability to support a user's what-if analysis through combinations of case-based modeling, system dynamics approaches and agent-based models working together. An example is provided for the domain of terrorist recruiting.

Keywords: model docking, knowledge level modeling, federated models, socio-cultural modeling.

1 Introduction

Modeling complex, social behaviors requires representations provided by different modeling approaches and across multiple scales [1]. An analyst may need information about individual, organizational, and societal behaviors in a system model that explores the interactions of variables at each of these scales of behavior. These federated models need to interoperate or become *docked*, i.e., communicate, collaborate, and share asynchronous information, to represent the system being studied. This paper focuses on developing a knowledge-level or semantic approach to federated model interoperability. There are many approaches for enabling the interaction of models at the data level, such as the High Level Architecture (HLA) for distributed simulations or MATREX[1] for battlespace modeling. This paper presents a research approach to enabling the interaction of multi-scale models at the knowledge-level. The specific advances that are made in this paper include presenting the types of knowledge (as opposed to specific data structures) that need to be exposed across models of different scales to enable seamless interoperability.

---

[1] Modeling Architecture for Technology Research & Experimentation.

There are many research challenges in identifying which models are appropriate to dock and the relevant concepts that they should exchange to build a richer multi-scale view of the world being modeled. Our early experiments [2] include docking of societal system dynamics models [3] with individual and organizational behaviors represented in agent-based models [4]. More recently, we have expanded our docking approach to include the docking of case-based models [5], which allow exploration of historical knowledge by other models.

Macro-scale models of a society, e.g., system dynamics-based models, rarely track the movements or opinions of individuals, and instead work with trends of the overall population. System dynamics models are useful for monitoring the effects of policies or actions on broad demographic groups. However, the decisions of people within the population are nuanced and are often dependent on those individuals' personal experiences, and individual traits are best simulated by a micro-scale model, such as an agent-based model. Historical experiences are best captured by a case-based reasoning system and can be used to inform the building of an agent that behaves rationally. Each of these model types has advantages and disadvantages. Rather than try to capture an entire system's behavior with a single model type, docking allows each model to represent the scale and type of information for which it is best suited. The research presented in this paper consists of our initial efforts to attain this goal through the investigation and experimentation of knowledge-level model docking.

2   Model Docking Framework

2.1   Existing Model Docking Frameworks

Existing model docking frameworks, such as the US Defense Advanced Research Projects Agency's COMPOEX[2] [6], offer effective tools for implementing the communication-layer and timing details necessary to create federated models. The Army's OneSAF (One Semi-Automated Forces) modeling environment combines tactical operations up to the battalion and brigade levels, with variable levels of resolution and supports composability within a single simulation environment. The Stabilization and Reconstruction Operations Model (SROM) supports analyzing interdependencies of PMESII[3] variables, but it is still predominantly a macro-scale model. Although there has been substantial work to enable multi-scale analyses and communication through a shared vocabulary or ontology, there is still much work to be done in understanding how to enable interactions of multiple models developed by different researchers, where the interactions occur at the knowledge-level and across different scales. The research presented in this paper aims to improve understanding these knowledge-level interactions.

2.2   The Knowledge Level

The *Knowledge Level* is a concept developed by Newell [7] to describe a level of representation above the symbolic level. At the knowledge-level, the research focuses

---

[2] Conflict Modeling, Planning, and Outcomes Experimentation.
[3] Political, Military, Economic, Social, Infrastructure, and Information.

on the study of the interaction of model content and semantics rather than the data structures, algorithms, or implementation used to enable the interaction. Our research focuses on exposing information between models at the knowledge-level. Thus, a mechanism for enabling knowledge-level experimentation for model docking is necessary. We use a blackboard approach to support the loose integration of model components through a shared knowledge area and enable the opportunistic use of shared knowledge by federated components. This allows for an easily reconfigurable experimentation environment enabling the research focus to remain on understanding the content to be exposed at the knowledge level.

2.3   Communicating across a Blackboard

The Blackboard [8, 9] environment uses a shared working memory area, called the *blackboard*, for collaborative problem solving amongst a set of autonomous knowledge sources that write information to and read information from the blackboard. Blackboard frameworks were developed to address complex and ill-defined problems for which existing techniques were inadequate. The knowledge sources are typically executable models or model components. They can post partial solutions, e.g., information that they have learned or generated, that may be of use to other knowledge sources. This information may include goals (or subgoals) and partial or complete problem solutions. The knowledge sources post requests for services or information from other knowledge sources and may choose to respond to requests. In this way, the blackboard approach supports asynchronous, opportunistic problem-solving by a "team" of collaborating knowledge sources.

Exploring and prototyping model docking using a blackboard allows models of different types and scales to post summaries or interpretations of scenarios or situations that may be useful to other models. The asynchronicity of the blackboard enables models to take information when they need it and based on their time-scales.

The blackboard approach is one that inherently enforces very little structure upon the models that make use of the shared memory area. This makes it an ideal method of communicating and sharing knowledge in an experimental system where the concepts being exchanged and the problem solving collaborations are being reconfigured in order to explore model docking issues.

This paper shows how a blackboard docking approach can be used for linking agent-based, system dynamics, and case-based models. This approach helps address the information gap caused by separate models and enables systematic reasoning of individual, social, and cultural behaviors. It also enables interactions between dynamic models resulting in increased capabilities for analysis and interpretation of behaviors, and it facilitates what-if analyses and analytical exploration associated with understanding multi-scale behavior.

2.4   Model Considerations

When developing a federated model, several questions must be kept in mind so as to ensure the results are relevant to the analysis. These questions helped support the development of the federated approach presented in this paper:

- What is the use case for the overall federated model?
- What are the individual knowledge-level elements that require modeling?
- What type of model is best for representing each system component? (Some models may represent multiple components.)
- What knowledge does each model require from other models?
- What knowledge can each model contribute to other models?
- What connecting concepts exist between each pair of models?

3   Docking Concept and Example Execution

Fig. 1 presents a graphic of three model types docked through blackboard.
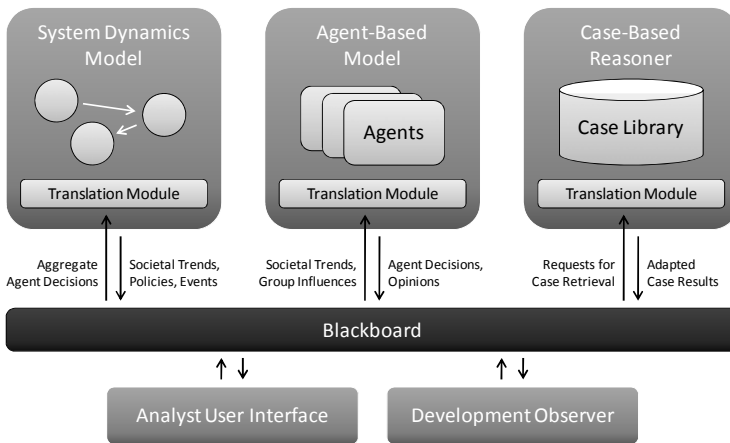
Fig. 1. Example of three model types docked with a blackboard

3.1   Identifying the Connecting Concepts

In building a complex multi-scale docked model, the individual models should be chosen so that concepts and vocabularies provide different views of the same system. The connecting concepts must be identified. These are the concepts represented in each model that could be supplied to other models docked to enable a richer representation of the system. An example is presented to illustrate this approach.

3.2   Example Use Case

To illustrate the approach, we present an example use case built around a terrorist recruiting scenario that includes elements at three different scales of analysis: micro, meso, and macro. The federated system is meant to answer questions about terrorist recruitment patterns.

Four independent executable models are included, each of which performs a specialized task:

  – Agent-Based Models. Two agent-based models are included, one with agents
    simulating individuals and one with agents representing organizations. The in-
    dividual agents make a recruitment decision based on their own personal traits
    and experiences. Group agents (representing terrorist groups, religious organiza-
    tions, and community groups) accept societal influences and exert influence
    over their members and other groups.
  – Case-Based Model. A case-based model is included that provides a set of past
    cases of individual recruiting. Each case consists of a single individual targeted
    for recruitment, including the properties of the individual, the group doing the
    recruiting, and the context. The solution consists of the individual's decision
    whether to join the group.
  – System Dynamics Model. A system dynamics model is included that provides a
    representation of societal interactions. This model represents the context in
    which the individuals and organizations make their decisions, including vari-
    ables such as economic and political stability, government policy constraints,
    and leadership popularity.

Individually, these models are capable of modeling only a piece of the overall system,
but in a federation, they provide a much broader environment for analysis.

Figure 2 presents an informal, example schedule of communications between mod-
els, as well as the expected knowledge to be passed from one model to another. Each
model's output represents the most relevant information that it has produced, rather
than being customized for use by a specific model. Because model communications
are loosely defined, modifications to the simulation require less modification of the
models themselves. In each model interaction, the sending model's output is written
to the blackboard, while the receiving model reads the data from the blackboard. It is
the obligation of the sending model to create knowledge that is as broadly useful as
possible, while it is the obligation of the receiving model to use that knowledge for its
own needs.

To understand how the system cycles and executes, we walk through the process in
Fig. 2. Here, an analyst sets up execution parameters for baseline values that are rele-
vant across all scales of knowledge. The end goal is for the analyst to use the system
to perform what-if experiments by altering input parameters and monitoring changes
in the outcome of the overall simulation. Once the inputs for a particular simulation
have been entered, the system begins execution.

The execution occurs as follows. Input parameters are passed to the appropriate
sub models for initiation. The system dynamics model begins by simulating macro-
scale (society-level) variables to support agent creation. Agents are used in two ways.
They are created to represent a hypothetical population (meso-scale behavior) using
demographics determined by the system dynamics model, and they are created to
represent individual (micro-scale behavior) that affect individual decisions. The case-
based reasoning model retrieves and adapts cases in which past stories are similar to
the current individuals and context. For cases that do not satisfy a certain similarity
metric from those available in the case library, the agent-based models prepare to
simulate the responses of those individuals using an internal influence model. The
agent-based knowledge for the historical cases and simulated individuals are aggre-
gated and used as numerical values by the system dynamics model. The system
dynamics model outputs societal decisions to the agents.
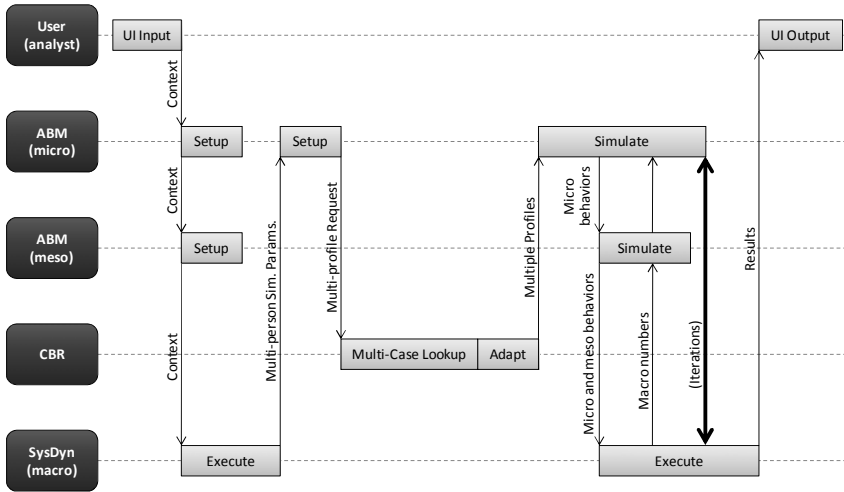
Fig. 2. Sample timing diagram of example model interactions

Table 1. Example connecting concepts

Model	Inputs	Outputs
Agent model, individual	— Societal trends — Conservatism of government — Opinions of society: gov't, morality, etc. — Events occurring in environment — Economic conditions — Targeted recruitment efforts — Influence of leaders, friends	— Opinions about terrorist group — Decision whether to join terrorist group — Interaction history
Case-based model	— Request for case retrieval — An individual's situation — Details of recruitment efforts — Question: "Is this person likely to be recruited?"	— Solution: yes or no — Explanation of solution? — Probabilities
System dynamics model	— Aggregate individual behaviors — Decisions to be recruited — Opinions of government, coalition actions — Aggregate details about environment, events	— Trends about societal decisions, opinions — Trends about recruitment efforts — Cultural variables

The system cycles amongst all the models, simulating individual attitudes, group behaviors, and societal responses, until the simulation reaches the terminating condition specified by the analyst. This configuration allows the docked models to autonomously control their execution and timing, referencing the overall modeling timeline and to allow meaningful information exchange across the models with differing time scales. This is known as opportunistic model execution.

An important element of this approach is visibility. Since a goal is to enable experimentation with model docking approaches, it is important to be able to view

the explicit knowledge exchanges among models to understand and evaluate the connecting concepts. Since each model's output is written to the blackboard for other models' use, a module, called the *development observer*, is incorporated to enable monitoring of interactions between models and to display the interactions to the model developer. This information can help the developer improve the design of such interactions. Table 1 shows some of the example connecting concepts.

3.3   Information to Expose

One of the benefits of docking independent models is that factors that influence behavior can exist at their 'natural' scale, i.e., the scale where they ordinarily occur. There is enough consistency to identify common, critical behavioral influences at their natural levels of representation. Typical macro-scale information includes demographics, culture, economics, and political structure. Also significant at this level is the representation of events affecting a large number of people or occurring over a long period a time. Representation at the meso-scale may include groups and organizations which may hold organizational beliefs and demonstrate behaviors. The structure of groups is also an important characteristic which may be used to determine how micro-scale agents communicate and are influenced by meso-scale variables. At the micro-scale, agents may represent individuals and typically demonstrate behaviors that are the result of the individuals' beliefs and attributes. Behaviors are especially relevant to other scales of analysis, i.e., many individual decisions may be aggregated to create a higher-scale attribute. Micro-scale aggregation of beliefs can result in a different representation of beliefs than when perceived from the meso-scale representation of group beliefs. A loosely structured classification of the types of information most commonly shared between scales should facilitate an analytical approach docking models that represent different behavioral aspects and scales of the same general problem. A recent paper [10] provides a more comprehensive description of the information types to expose at each scale.

4   Conclusions

This paper presents an initial approach for exploring the docking of social models at the knowledge level. We have prototyped a simple blackboard environment that allows for experimentation with docking models and iterative improvement of the approach. There are many research challenges in identifying which models are appropriate to dock and the relevant concepts that they should exchange to in order to build a richer multi-scale view of the world being modeled. Our early approach includes docking of societal system dynamics models with individual and organizational behaviors represented in agent-based models and case-based models, which allow exploration of historical knowledge by other models. The research presented in this paper represents initial efforts to attain opportunistic and asynchronous interactions among multi-scale models through the investigation and experimentation of knowledge-level model docking.

References

1. Zacharias, G., MacMillan, J., Van Hemel, S. (eds.): Behavioral Modeling and Simulation: from Individuals to Societies. National Research Council, Washington, D.C. (2008)
2. Weiss, L., Whitaker, E., Briscoe, E., Trewhitt, E.: Cultural and Behavioral Model Docking Research. In: Human Social Cultural and Behavioral Modeling Focus 2010 Conference, Chantilly, VA (2009),
   `http://www.gtri.gatech.edu/files/media/`
   `Weiss_Abstract_HSCB_Conference_2010_public.pdf`
3. Forrester, J.: Counterintuitive Behavior of Social Systems. Technology Review 73(3), 52–68 (1971)
4. Macal, C., North, M.: Tutorial on Agent-Based Modeling and Simulation. In: Proceedings of the 2005 Winter Simulation Conference, Argonne (2005)
5. Kolodner, J.: Case-Based Reasoning. Morgan Kaufmann, San Mateo (1993)
6. Defense Advanced Research Projects Agency COMPOEX,
   `http://www.darpa.mil/ipto/programs/compoex/compoex.asp`
7. Newell, A.: The Knowledge Level. AI Magazine 2(2), 1–33 (1981)
8. Engelmore, R.S., Morgan, A. (eds.): Blackboard Systems. Addison-Wesley, Reading (1988)
9. Whitaker, E., Bonnell, R.: A Blackboard Model for Adaptive and Self-Improving Intelligent Tutoring Systems. Journal of Artificial Intelligence in Education (1992)
10. Briscoe, E., Trewhitt, E., Whitaker, E., Weiss, L.: Multi-Scale Interactions in Behavior Modeling, Submitted to: Social Computing, Behavioral-Cultural Modeling and Prediction 2011, March 29-31, Sundance, UT (2011)

Using Web-Based Knowledge Extraction
Techniques to Support Cultural Modeling

Paul R. Smart[1], Winston R. Sieck[2], and Nigel R. Shadbolt[1]

[1] School of Electronics and Computer Science, University of Southampton,
Southampton, SO17 1BJ, UK
{ps02v,nrs}@ecs.soton.ac.uk
[2] Applied Research Associates, Inc., 1750 Commerce Center Blvd, N., Fairborn, OH,
45324, USA
wsieck@ara.com

Abstract. The World Wide Web is a potentially valuable source of information about the cognitive characteristics of cultural groups. However, attempts to use the Web in the context of cultural modeling activities are hampered by the large-scale nature of the Web and the current dominance of natural language formats. In this paper, we outline an approach to support the exploitation of the Web for cultural modeling activities. The approach begins with the development of qualitative cultural models (which describe the beliefs, concepts and values of cultural groups), and these models are subsequently used to develop an ontology-based information extraction capability. Our approach represents an attempt to combine conventional approaches to information extraction with epidemiological perspectives of culture and network-based approaches to cultural analysis. The approach can be used, we suggest, to support the development of models providing a better understanding of the cognitive characteristics of particular cultural groups.

Keywords: cultural modeling, ontology-based information extraction, culture, cognition, knowledge extraction, world wide web.

1 Introduction

The World Wide Web (WWW) is a valuable source of culture-relevant information, and it is therefore an important resource for those interested in developing cultural models. The exploitation of the WWW in the context of cultural modeling is, however, hampered both by the large-scale nature of the Web (which makes relevant information difficult to locate) and the current dominance of natural language formats (which complicates the use of automated approaches to information processing). In this paper, we describe an approach to support the use of the Web in cultural modeling activities. The approach is based on the development of ontology-based information extraction capabilities, and it combines the use of Semantic Web technologies and natural language processing (NLP) techniques with an epidemiological perspective of culture [1] and the use

of network-based approaches to cultural analysis [2]. Technological support for the approach is currently being developed in the context of the IEXTREME[1] project, which is funded by the U.S. Office of Naval Research.

The structure of the paper is as follows. In Section 2, we outline what is meant by the term 'culture', and we describe an approach to cultural modeling that is based on the development of models representing the ideas associated with particular cultural groups. In Section 3, we describe our approach to the development of Web-based knowledge extraction capabilities to support cultural model development. This approach combines conventional approaches to information extraction with semantically-enriched representations of cultural models, and it seeks to provide a culture-oriented ontology-based information extraction (OBIE) capability for the WWW.

2   Cultural Models and Cultural Network Analysis

Before addressing the use of the Web to study culture, we first need to define what is meant by the term 'culture'. As is to be expected in any highly interdisciplinary field, there are a variety of conceptions of culture. Our conception is distinctly cognitive in nature, and it is based on an epidemiological perspective [1]. A fundamental assumption of this perspective is that shared developmental experiences lead to important similarities in the mental representations (e.g. values and causal knowledge) that are distributed among members of a population. Culturally widespread ideas ground the distribution of behavioral norms, discussions, interpretations, and affective reactions, and researchers working within the epidemiological perspective thus seek to describe and explain the prevalence and spread of ideas within specific populations.

Working from this perspective, we previously developed a technique called cultural network analysis (CNA), which is a method for describing the 'ideas' that are shared by members of cultural groups [2]. CNA discriminates between three kinds of ideas, namely, concepts, values, and causal beliefs, which together constitute the contents of what are called 'cultural models'. These cultural models typically rely on the use of belief network diagrams to show how the set of relevant ideas relate to one another (see Fig. 1 for an example).

In general, we can distinguish two types of cultural models: qualitative and quantitative cultural models (see [2]). Qualitative cultural models present the ideas associated with a particular group, whereas quantitative models add information about the prevalence of those ideas in the target population. In addition to seeing the approach described in this paper as a means to validate and refine qualitative cultural models, it is also possible to see the approach as enabling a cultural analyst to estimate the relative frequency of ideas in a target population and thus develop quantitative extensions of qualitative cultural models (see Section 3.6).

---

3   Web-Based Knowledge Extraction for Cultural Model Development

In this section, we describe an approach to cultural model development that combines CNA with state-of-the-art approaches to knowledge representation and Web-based information extraction. The aim is to better enable cultural model developers to exploit the WWW as a source of culture-relevant information. The approach we describe is based on a decade of research into OBIE systems (see [3] for a review), and it combines conventional approaches to information extraction with an ontology that provides background knowledge about the kinds of entities and relationships that are deemed important in a cultural modeling context. The first step in the process is to develop an initial qualitative cultural model using a limited set of knowledge sources (see [2] for more details on this step). The second step involves the development of a cultural ontology using the qualitative cultural model as a reference point. This ontology is represented using the Ontology Web Language (OWL), which has emerged as a de facto standard for formal knowledge representation on the WWW. The third step is to manually annotate sample texts using the cultural ontology in order to provide a training corpus for rule learning. Rule learning, in the current context, is mediated by the $(LP)^2$ algorithm, which is a supervised algorithm that has been used to develop a variety of adaptive information extraction and semantic annotation capabilities [4,5]. Following the development of information extraction rules, the rules are then applied to Web resources in the fourth step in order to identify instances of the entities defined in the initial qualitative cultural model. Step five consists in the identification and extraction of causal relations. The extraction of causal relationships is a difficult challenge because techniques for information extraction have tended to focus on the extraction of particular entities in a text, rather than the relationships between those entities. We attempt to extract causal relationships using an approach that combines the use of background knowledge in the form of a domain ontology with the general purpose lexical database, WordNet [6]. Finally, in step six, the extracted cultural knowledge is integrated, stored, and used to estimate the relative frequencies of the various ideas presented in the initial qualitative cultural model. We briefly describe each of these steps in subsequent sections.

3.1   Step 1: Develop Qualitative Cultural Model

The technique used to develop qualitative cultural models has been described in previous work [2], and we will not reiterate the details of the technique here. Fig. 1 illustrates a simplified qualitative cultural model that represents an extremist Sunni Muslim's beliefs about current socio-political relationships between Islam and the West. The set of ideas represented in Fig. 1 were extracted from articles that describe jihadist narratives, and they are presented here for illustrative purposes. The cultural model illustrates concepts shared by the group, as well as their common knowledge of the causal relationships between those concepts. This shared knowledge influences expectations about how socio-political
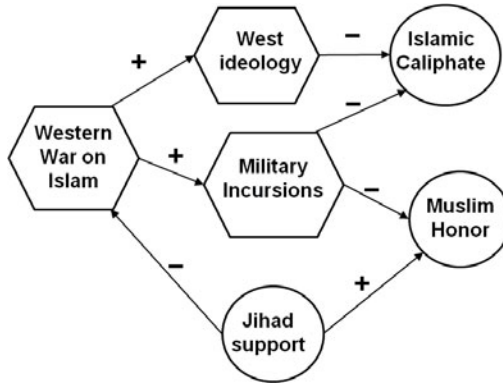
Fig. 1. Sunni extremist cultural model of jihad (simplified)

relationships will unfold, and it provides a basis for the selection of particular actions and decision outcomes; for example, the decision to support jihad.

Fig. 1 shows the three different kinds of ideas that are the targets of a culture-oriented knowledge extraction system. These ideas include simple concepts such as "Western ideology" and "Muslim Honor", each represented as closed shapes in Fig. 1. They also include causal beliefs; for example, the idea that Western ideology (e.g. secularism, nationalism) is inhibiting the formation of a unified Islamic caliphate and the idea that the West promotes this ideology because it is engaged in a covert war against Islam. These causal beliefs are represented as arrows in the figure, with the +/- symbols indicating the polarity of the causal relationship. Finally, Fig. 1 portrays values using specific shapes, with circles indicating "positive" outcomes and hexagons indicating "negative" outcomes. Developing an Islamic caliphate is thus a good thing according to the cultural model. Maintaining (and enhancing) Muslim honor is likewise valued. According to the model, jihad is viewed positively and should be supported by the model's adherents due to the perceived anticipated consequences for Muslims. Most directly, support for jihad decreases the chances that the West will continue its war against Islam, and it enhances collective Muslim honor.

3.2   Step 2: Develop Cultural Ontology

Once an initial qualitative cultural model has been developed, the next step in the process is to develop an ontology that represents the contents of the model. The main reason this step is undertaken is that it enables the cultural model to be used to support information extraction. Over recent years, a rich literature has emerged concerning the use of ontologies in information extraction, and a number of important tools have been developed to support OBIE [3]. By converting the cultural model into an ontology using standard knowledge representation languages, such as OWL, we are able to capitalize on the availability of these pre-existing tools and techniques, and we are also able to compare the success of our approach with other OBIE approaches.

The ontologies developed to represent the contents of cultural models are based around the notion of ideas as being divided into concepts, causal beliefs and values. These three types of ideas constitute the top-level constructs of the ontology, and subtypes of these constructs are created to represent the various elements of the cultural model. For example, if we consider the notion of "Jihad support", as depicted in Fig. 1, then we can see that this is a type of concept, and it is regarded as a positive thing, at least from the perspective of the target group. Within the ontology developed to support this cultural model we have the concept of 'support-for-jihad', which is represented as a type of 'jihad-related-action', which is in turn represented as a type of 'action', which is in turn represented as a type of 'concept'. Given the focus of the cultural model in representing causal beliefs, the notions of actions, events and the causally-significant linkages between these types of concept are often the most important elements of the cultural ontology.

3.3   Step 3: Develop OBIE Capability

In this step of the process, the aim is to create rules that automatically detect instances of the concepts, beliefs and values contained in the cultural model. There are clearly a number of ways in which this might be accomplished, especially once one considers the rich array of information extraction techniques and technologies that are currently available [7]; however, we prefer an approach that delivers symbolic extraction rules (i.e. rules that are defined over the linguistic features and lexical elements of the source texts) because the knowledge contained in the rules can be easily edited by subject matter experts. In addition, it is easier to provide explanation-based facilities for rule-based symbolic systems than it is for systems based on statistical techniques.

The approach to rule creation that we have adopted in the context of the IEXTREME project is based on the use of the $(LP)^2$ learning algorithm, which has been used to create a number of semantic annotation systems [4,5]. The basic approach is to manually annotate a limited number of source texts using the cultural ontology that was created in the previous step. These annotated texts are then used as the training corpus for rule induction (see [8] for more details). During rule induction, the $(LP)^2$ algorithm generalizes from an initial rule that is created from a user-defined example by using generic shallow knowledge about natural language. This knowledge is provided by a variety of NLP resources, such as a morphological analyzer, a part-of-speech (POS) tagger and a gazetteer. The rules that result from the learning process thus incorporate a variety of lexical and linguistic features. Previous research has suggested that rules can be defined over a large number of features. For example, Bontcheva et al [9] used a variety of NLP tools to generate 94 features over which information extraction rules could be defined. Of course, not all of these features are likely to be of equal importance in creating information extraction systems, and further empirical studies are required to assess their relative value in the domain of cultural modeling.

3.4   Step 4: Extract Concepts

Once extraction rules have been created, they can be applied to potential knowledge sources in order to detect occurrences of the various ideas expressed in the cultural model. Because most of the user-defined annotations will be based on the nodal elements of the cultural model networks, such as those seen in Fig. 1, this step is particularly useful for detecting mentions of specific concepts in source texts. In the case of Sunni extremist cultural models this could, for example, include mentions of jihad-related concepts, for example "Jihad is a means to expel the Western occupiers", as well as references to aspects of Western ideology. In general, information extraction systems based on the machine learning technique described above (i.e. the $(LP)^2$ algorithm) have proved highly effective in identifying instances of the terms defined in an ontology, so we can expect reasonable extraction performance for this step of the process.

3.5   Step 5: Extract Causal Relations

There have been a number of attempts to extract relational information in a Web-based context (see [7]). The use of ontologies in such systems plays an important role because they provide background knowledge about the possible semantic relationships that are likely to exist between the various entities identified in previous processing steps. Thus, if a system first subjects a textual resource to entity-based semantic annotation, then it is able to use the ontology to form expectations about the kind of relationships that might be apparent in particular text fragments. When this background knowledge is combined with lexical and linguistic information, a relation extraction system is often able to identify relationships that would be impossible to detect using a text-only analysis.

The approach to relation extraction that we have adopted in the case of the IEXTREME project is based on a technique that was previously developed to support information extraction in the domain of artists and artistic works [10]. The approach builds on the outcome of the previous step, which is concerned with the detection of concepts in the source texts. Importantly, once these concept annotations are in place, the relation extraction subsystem is provided with a much richer analytic substrate than would otherwise have been the case. In fact, it is only once such annotations are in place that the real value of the ontology (for the detection and extraction of relationships) can be appreciated. In particular, the ontology provides background knowledge that drives the formation of expectations about the kinds of relationships that could appear between concepts, and once such expectations have been established, they can be supported or undermined by subsequent lexical analysis of the sentence in which the concepts occur.

Obviously, the nature of the natural language processing that is performed on the sentence is key to this relation-based annotation capability. It is not sufficient for a system to simply form an expectation about the kind of relationships that might occur between identified entities in the text; the system also needs to ascertain whether the linguistic context of the sentence supports the

assertion of a particular relationship. The decision regarding which relationship (if any) to assert in a particular sentential context is based on a strategy similar to that used in previous research [10]. Essentially, each relationship in the ontology is associated with a 'synset' (a set of synonyms) in the general-purpose lexical database WordNet [6]. When the relation extraction system executes, it attempts to match the words in a sentence against the WordNet-based linguistic grounding provided for each expected relationship. In addition to representing information about synonyms, the WordNet database also represents hypernymy (superordinate) and hyponymy (subordinate) relationships. These can be used to support the matching process by avoiding problems due to transliteration.

3.6   Step 6: Exploit Knowledge Extraction Capability

The knowledge extraction capability outlined in the previous steps provides support for the refinement, extension and validation of the knowledge contained in cultural models. The ability to detect instances of the ideas expressed in cultural models across a range of Web resources (including blogs, organizational websites, discussion threads and so on) provides a means by which new knowledge sources can be discovered and made available for a variety of further model development and refinement activities. The use of OBIE technology therefore provides a means by which the latent potential of the Web to serve as a source of culture-relevant knowledge and information can be exploited in the context of qualitative cultural modeling initiatives.

Aside from the development of better qualitative cultural models, the use of knowledge extraction techniques can also support the development of quantitative cultural models. As discussed above, quantitative cultural models extend qualitative cultural models by including information about the relative frequencies of particular ideas within the population to which the cultural model applies [2]. By harnessing the power of OBIE methods, the current approach provides a means by which ideas (most notably concepts and causal beliefs) can be detected across many hundreds, if not thousands, of Web resources. This provides an estimate of the prevalence of particular ideas in the target population of interest, and it provides a means by which the Web can be used to support the development of quantitative cultural models.

4   Conclusion

This paper has described an approach to harnessing the latent potential of the Web to support cultural modeling efforts. The approach is based on the development of culture-oriented knowledge extraction capabilities and the use of techniques that support a cognitive characterization of specific cultural groups. Systems developed to support the approach may be seen as an important element of iterative cultural modeling efforts, especially ones in which an initial qualitative cultural model drives the acquisition of information from a large number of heterogeneous Web-based resources.

References

1. Sperber, D.: Explaining Culture: A Naturalistic Approach. Blackwell, Malden (1996)
2. Sieck, W.R., Rasmussen, L., Smart, P.R.: Cultural Network Analysis: A Cognitive Approach to Cultural Modeling. In: Verma, D. (ed.) Network Science for Military Coalition Operations: Information Extraction and Interaction. IGI Global, Hershey (2010)
3. Wimalasuriya, D.C., Dou, D.: Ontology-based information extraction: An introduction and a survey of current approaches. Journal of Information Science 36(3), 306–323 (2010)
4. Vargas-Vera, M., Motta, E., Domingue, J., Lanzoni, M., Stutt, A., Ciravegna, F.: MnM: ontology driven semi-automatic or automatic support for semantic markup. In: 13th International Conference on Knowledge Engineering and Knowledge Management, Siguenza, Spain (2002)
5. Ciravegna, F., Wilks, Y.: Designing adaptive information extraction for the Semantic Web in Amilcare. In: Handschuh, S., Staab, S. (eds.) Annotation for the Semantic Web. IOS Press, Amsterdam (2003)
6. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography 3(4), 235–244 (2004)
7. Sarawagi, S.: Information extraction. Foundations and Trends in Databases 1(3), 261–377 (2008)
8. Ciravegna, F.: Adaptive information extraction from text by rule induction and generalisation. In: 17th International Joint Conference on Artificial Intelligence, Seattle, Washington, USA (2001)
9. Bontcheva, K., Davis, B., Funk, A., Li, Y., Wang, T.: Human Language Technologies. In: Davies, J., Grobelnik, M., Mladenic, D. (eds.) Semantic Knowledge Management: Integrating Ontology Management, Knowledge Discovery, and Human Language Technologies. Springer, Berlin (2009)
10. Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P., Shadbolt, N.R.: Automatic Ontology-Based Knowledge Extraction from Web Documents. IEEE Intelligent Systems 18(1), 14–21 (2003)

How Corruption Blunts Counternarcotic Policies in Afghanistan: A Multiagent Investigation

Armando Geller, Seyed M. Mussavi Rizi, and Maciej M. Łatek*

Department of Computational Social Science
George Mason University
4400 University Drive
Fairfax VA 22030, USA
{ageller1,smussavi,mlatek}@gmu.edu

Abstract. We report the results of multiagent modeling experiments on interactions between the drug industry and corruption in Afghanistan. The model formalizes assumptions on the motivations of players in the Afghan drug industry, quantifies the tradeoffs among various choices players face and enables inspection of the time, space and level of supply chain in which one can expect positive and negative impacts of counternarcotic policies. If reducing opium exports is one measure of effectiveness for NATO operations in Afghanistan, grasping the links between corruption and the drug industry should provide a better picture of the second-order interactions between corruption and investment in improving the governance quality, in deploying security forces tasked with eradication and interdiction and in programs to enhance rural livelihoods.

Keywords: Afghanistan, Corruption, Multiagent Simulation, Empirical Models, Bounded Rationality.

1 Introduction

In Afghanistan, an illicit economy, a weak government, factionalization and the erosion of norms [1] recur as symptoms of long-lasting conflict along reinforcing divides of modernity versus tradition, urban versus rural, and center versus the periphery. Rampant corruption in a de facto narco-state [2] and a booming drug industry in a de facto highjacked government [3] show two faces of the same Janus guarding the gates to the complex Afghan political economy. In 2009, Afghanistan ranked 179 out of 180 countries in government corruption [4]; 59% of Afghans consider government corruption a greater concern than lack of security and unemployment [5] while bribery is estimated to impose a billion-dollar burden on the Afghan gross domestic product (GDP) annually [6]. Afghanistan has long dominated global opium markets [7] with no less than 30% of the country GDP coming from drug trafficking [5]. Drug money funds both warlords who often double as government officials and the insurgents who fight NATO

forces and the government. Both sides tax narcotics trade, protect shipments, run heroin labs, and organize farm output in areas they control [8].

After 4 years of refusing to take on the drug industry in Afghanistan [9], the U.S. adopted eradication without alternative livelihoods as the dominant counternarcotic policy in 2006. The policy failed to curb poppy cultivation, but it is credited as "the single biggest reason many Afghans turned against the foreigners" [10]. In 2007, the U.S. sought to complement eradication with rural development programs like the ineffective wheat distribution program in 2008: the program was based on unusually high price ratio of wheat to poppy; failed to address farmers' incentives for poppy cultivation, and undermined the local seed market. Mindful of the antagonizing effects of eradication on Afghan farmers and convinced of the relative efficiency of interdiction [11,12], in 2009 the Obama administration deemphasized eradication and focused on interdiction and rural development instead [13].

Experts agree that illicit economic activities, especially cultivating poppies and trading narcotics feed government corruption in Afghanistan [14,15]. Yet, identifying the weaknesses of current counternarcotic efforts in Afghanistan has proven easier than explaining the causal links between corruption and the drug industry; thus outlining effective policies to combat them both. Goodhand [16] finds that counternarcotic policies fuel conflict, because they render the complex interdependencies between rulers and peripheral elites unstable and Felba-Brown [17] blames the same complex interdependencies between economic incentives and local power structures for the eventual failure of the current counternarcotic and anti-corruption policies.

We report the results of multiagent experiments on the interaction between the drug industry and corruption in Afghanistan. We have argued in [18] that multiagent modeling is the only approach that can explain historical variations in opium export volumes and prices given the incentives and constraints of Afghan farmers, traders, traffickers and power brokers. Our model formalizes assumptions on the motivations of players in the Afghan drug industry, quantifies the tradeoffs among various choices players face and enables the inspection of the time, space and the level of supply chain in which one can expect positive and negative impacts of counternarcotic policies. On a more applied note, if reducing opium exports is one measure of effectiveness for NATO operations in Afghanistan, grasping the links between corruption and the drug industry should provide a better picture of the second-order interactions between corruption and investment in improved governance, in security forces tasked with eradication and interdiction and in programs to enhance rural livelihoods.

2   The Model: Actors, Behaviors and Validity

Figure 1 outlines the static structure of our model that consists of three basic agent classes of farmers, traders and power brokers. While power brokers can be used to represent ISAF and insurgents, they mainly represent district and province governors who engage in interdiction, eradication or racketeering. Each agent class needs to operate across multiple contexts and has multiple types of decisions to make:
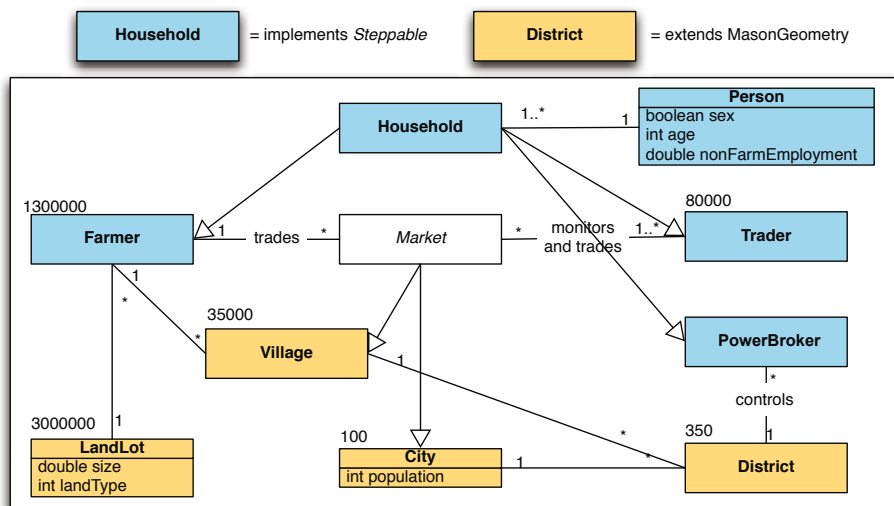
Fig. 1. Simplified UML class diagram of the static structure of the model. Main agent types—farmers, traders and power brokers—extend the household class. Therefore, they are located in a specific place and embedded in kinship and social networks. Their welfare and security conditions can also be monitored. Numbers denote approximate counts of each agent class in the model.

- Traders (*a*) negotiate price, place and size of trades; (*b*) route shipments, react to interdiction and banditry, corrupt power brokers; and (*c*) adjust the size of target stock.
- Farmers (*a*) allocate land, decide when to harvest; and (*b*) rent and offer labor, react to eradication, corrupt power brokers.
- Power brokers (*a*) determine protection rates; (*b*) respect or renege on mutual agreements; and (*c*) allocate forces to interdiction and eradication.

We faced two challenges in our effort: (*a*) translating qualitative information into agent purpose, cognition and reasoning and (*b*) endowing agents with necessary attributes and placing them in the environment. This challenge was amplified by the fact that we had to find a right balance between using historical behaviors coded as if-then-else rules and representing agent adaptation and anticipation driven by first principles, because the purpose of the model is to support policy decisions. For instance, take a farmer's decision on what crops to cultivate. To make this decision, a farmer needs to account for government and ISAF eradication policies, crop and labor market conditions, family demographics and food security conditions, and availability of seeds and fertilizer. Farmers are endowed with plots of land characterized by land types with specific yields for any combination of potential crops and climate conditions. Some land types are irrigated, so climatic conditions do not influence their yield much. Some crops, for example poppy, are quite resistant to weather conditions regardless of whether they are
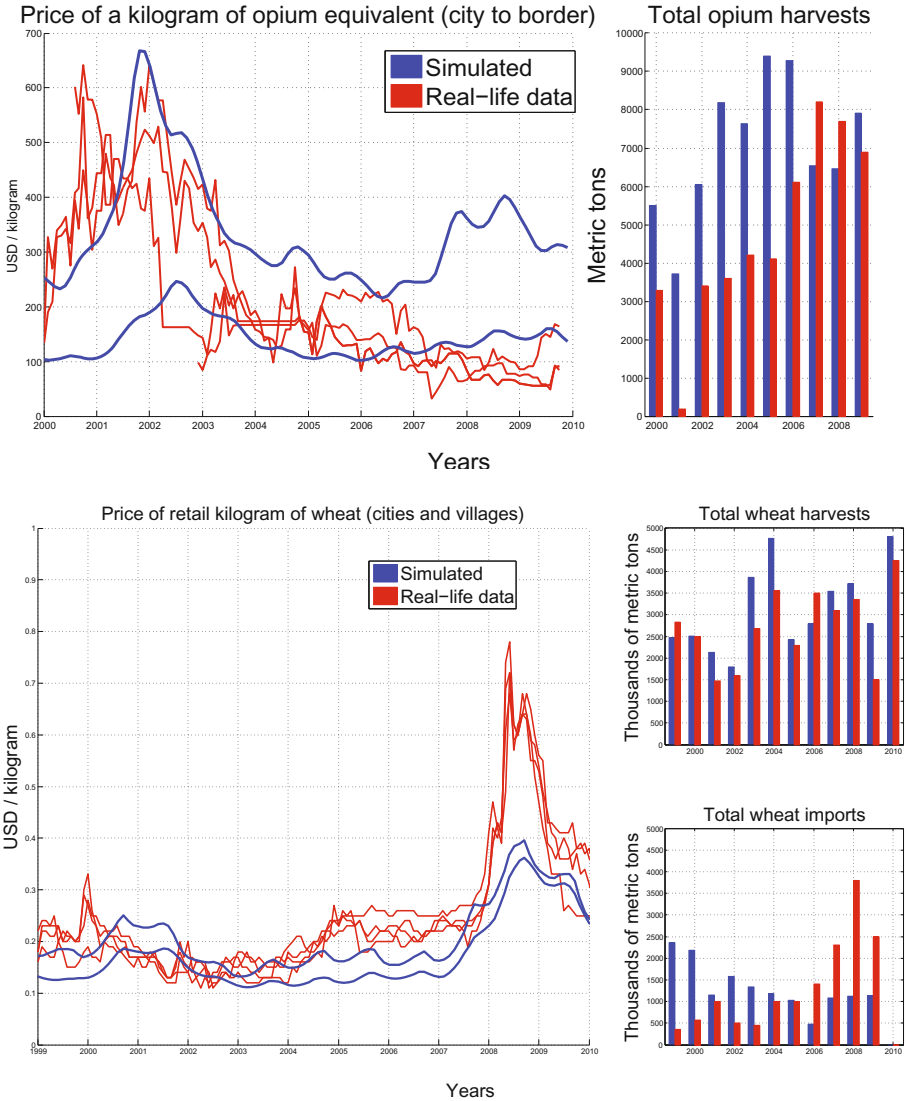
Fig. 2. Output validation against historical data for 1999–2009 compiled from [14, 19], corresponding to city- or province-level observations from Nangarhar, Kandahar, Helmand, Badakhshan, Mazari Sharif and Herat for opium, and Kabul, Kandahar, Jalalabad, Herat, Mazari Sharif and Faizabad for wheat on the left panel. Simulation price time series on the left panel are min–max ranges for the same markets. The right panel shows the country-level poppy harvest and wheat imports and harvest for the same period.

cultivated on irrigated land or not. In [20] we proposed and validated an approach to data fusion that helps build all necessary connections between agents and the environment using available data.
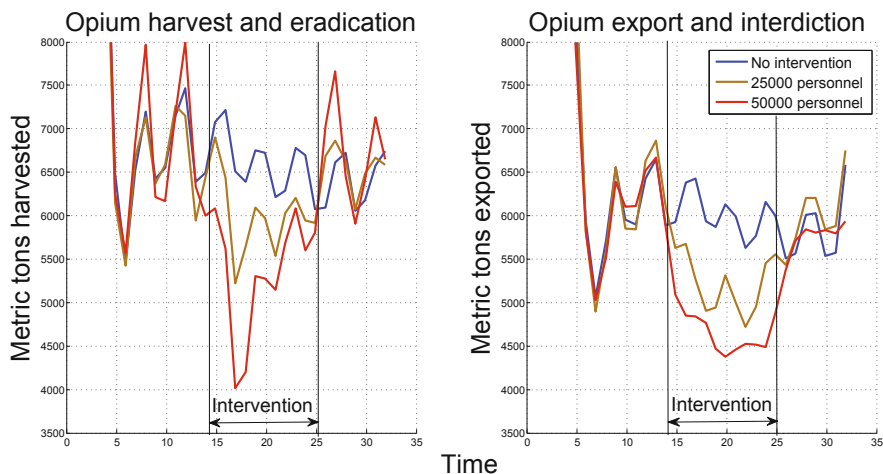
Fig. 3. Results of interventions with corruption. On the left panel, we use eradication to suppress opium harvests. On the right panel, we target traders with interdiction to suppress opium exports. Intervention is introduced in year 15 and lifted after year 25. Opium losses before reaching border and export are about 10%. Each condition is an average of 10 independent model runs.

The complexity of interdependencies among agents' decisions is significant. For example, if a farmer with a large family expects imminent drought when poppy prices are high enough, he may decide to forgo grain cultivation and plant only drought-resistant, labor-intensive poppy, hoping to buy wheat from the market with opium income later. Similarly, some farmer households may decide to fallow some of their land and offer accessory labor for farm wages to other households. At the same time, they need to remember that if they choose to cultivate poppy, they will be either exposed to eradication or need to pay protection bribe. [21,22,23] and [24] provide detailed accounts of farmers' crop choice behavior along with their objectives, reasoning mechanisms and information they take into account. Incorporating such lower-level decision factors into our model guarantees that agents can adapt to changing environment plausibly.

After instantiating the model at full scale, we ran it in open loop for 10 years to eliminate transients, then introduced historical weather[1] and global wheat prices approximated by the wholesale price of wheat in Pakistan. We also represented the global heroin market and major events in security conditions.[2] The resulting dashboard with simulated and historical dynamics compared for the 1999–2009 period is presented on Figure 2.

---

[1] We specified monthly, local weather conditions down to the village level, using vegetation indexes in [25] to correlate yields at harvest.

[2] We kept the value of all opium and heroin exports from Afghanistan pinned to $2 billion. We also provided a crude estimate of the number of people dealing with eradication and interdiction at the national level.

(a) Eradication

(b) Interdiction
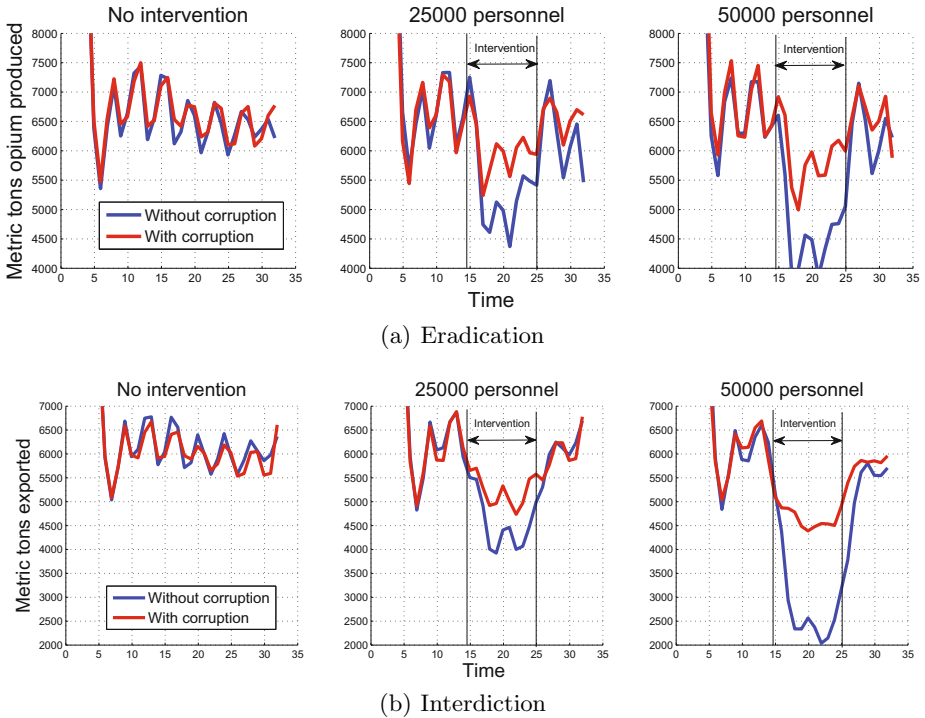
Fig. 4. Sensitivity of interventions to corruption. The same experiment design as on Figure 3 has been used. Each condition is an average of 10 independent runs.

3   Policy Experiments

We performed experiments against stationary environmental conditions, with different magnitudes of forces dedicated to either eradication or interdiction after running the model for 15 years. The results of interventions in the presence of corruption are summarized on Figure 3. On Figures 4(a) and 4(b) we repeat the process, this time contrasting the default situation with a universe without corruption. The results show that corruption is merely one mechanism that causes opium production and exports to rebound during and after intervention. Under the eradication regime, prices shift and the spatial distribution of poppy cultivation changes; under the interdiction regime traffickers adapt shipment routes. These adaptive and anticipatory processes provide complementary mechanisms explaining the resilience of the system. In a perfect world, interdiction would be much more cost-effective than eradication, by some 30%. In reality, due to the aforementioned factors, the counternarcotic effects of interdiction diminish faster than those of eradication.

4 Conclusions

We have outlined our approach to building an empirically-informed multiagent model of the drug industry in Afghanistan and used it to study the question of resilience of the opium economy to counternarcotic policies implemented by a government riddled with corruption. We have presented the results of model validation against historical conditions and used experiments to quantify the reduction in the effectiveness of eradication and interdiction, due to the presence of corruption of officials tasked with implementing these policies. We have demonstrated that corruption blunts any eradication and interdiction efforts.[3]

Our model suffers from a number of limitations that open windows for further model development, experiments and field research. On the production side, we have forgone the "salaam" credit system and how eradication compounds farmers' debt. Our farmers and traders live in a world devoid of existing or emergent social norms: They simply seek survival. On the scenario analysis side, we have made the sweeping assumption that all policy enforcers are subject to corruption. While the generality of this assumption helps us construct "grim scenarios" with conservative estimates, in reality, multiple agencies, some of whom under direct operational supervision of U.S. counternarcotic personnel, carry out operations.

References

1. Rubin, B.: Saving Afghanistan. Foreign Affairs 86(1), 57–78 (2007)
2. Burnett, V.: Crackdown On Afghanistan's Cash Crop Looms: In War on Drugs. Authorities Seek to Uproot Poppies. Boston Globe, September 18 (2004)
3. Baldauf, S., Bowers, F.: Afghanistan Riddled with Drug Ties; The Involvement of Local as Well as High-Level Government Officials in the Opium Trade Is Frustrating Efforts to Eradicate Poppy Fields. In: Christian Science Monitor, May 13 (2005)
4. Transparency International: Annual Report (2009)
5. United Nations Office of Drugs and Crime: Corruption in Afghanistan. Bribery as Reported by the Victim (2010)
6. Integrity Watch Afghanistan: Afghan Perceptions and Experiences of Corruption. A National Survey (2010)
7. Macdonald, D.: Drugs in Afghanistan: Opium, Outlaws and Scorpion Tales. Pluto Press (2007)
8. Peters, G.: Taliban Drug Trade: Echoes of Colombia. In: Christian Science Monitor, November 21 (2006)
9. Risen, J.: Poppy Fields Are Now a Front Line in Afghanistan War. In: New York Times, May 16 (2007)
10. Hari, J.: Legalize It: Why Destroy Poppies and Afghan Farmers When the World Needs Legal Opiates? In: Los Angeles Times, November 6 (2006)

---

[3] This paper focuses on drugs suppression policies; however, the model includes alternative development, alternative crops, infrastructure development and specific anti-corruption policies whose discussions were omitted, because they were not directly relevant to our main findings.

11. Rubin, B.R., Sherman, J.: Counter-Narcotics to Stabilize Afghanistan: The False Promise of Crop Eradication. Technical report (2008)
12. Mansfield, D., Pain, A.: Counter-Narcotics in Afghanistan: The Failure of Success? Technical report, Afghanistan Research and Evaluation Unit (2009)
13. Stack, L.: US Changes Course on Afghan Opium, Says Holbrooke. Christian Science Monitor, June 28 (2009)
14. Buddenberg, D., Byrd, W.: Afghanistan's Drug Industry: Structure, Functioning, Dynamics and Implications for Counter-Narcotics Policy. Technical report, United Nations Office on Drugs and Crime (2009)
15. Katzman, K.: Afghanistan: Politics, Elections, and Government Performance. Technical Report RS21922, Congressional Research Service Report (2010)
16. Goodhand, J.: Corrupting or Consolidating the Peace? The Drugs Economy and Post-conflict Peacebuilding in Afghanistan. International Peacekeeping 15(3), 405–423 (2008)
17. Felbab-Brown, V.: Testimony before the Senate Caucus on International Narcotics Control, October 21 (2009)
18. Latek, M.M., Rizi, S.M.M., Geller, A.: Persistence in the Political Economy of Conflict: The Case of the Afghan Drug Industry. In: Proceedings of AAAI Complex Adaptive Systems Fall Symposium (2010)
19. Food and Agriculture Organization of the United Nations: Global Information and Early Warning System (2010)
20. Rizi, S.M.M., Latek, M.M., Geller, A.: Merging Remote Sensing Data and Population Surveys in Large, Empirical Multiagent Models: The Case of the Afghan Drug Industry. In: Proceedings of the Third World Congress of Social Simulation (2010), http://css.gmu.edu/projects/irregularwarfare/remotesensing.pdf
21. Maletta, H., Favre, R.: Agriculture and Food Production in Post-war Afghanistan: A Report on the Winter Agricultural Survey 2002-2003. Technical report, Food and Agriculture Organization (2003)
22. Mansfield, D.: What is Driving Opium Poppy Cultivation? Decision Making Amongst Opium Poppy Cultivators in Afghanistan in the 2003/4 Growing Season. In: UNODC/ONDCP Second Technical Conference on Drug Control Research, pp. 19–21 (2004)
23. Kuhn, G.: Comparative Net Income from Afghan Crops. Technical report, Roots of Peace Institute (2009)
24. Chabot, P., Dorosh, P.: Wheat Markets, Food Aid and Food Security in Afghanistan. Food Policy 32(3), 334–353 (2007)
25. National Aeronautics and Space Administration: The Moderate Resolution Imaging Spectroradiometer (MODIS) Data (2010)

Discovering Collaborative Cyber Attack Patterns Using Social Network Analysis⋆

Haitao Du and Shanchieh Jay Yang

Department of Computer Engineering
Rochester Institute of Technology, Rochester, New York 14623

Abstract. This paper investigates collaborative cyber attacks based on social network analysis. An Attack Social Graph (ASG) is defined to represent cyber attacks on the Internet. Features are extracted from ASGs to analyze collaborative patterns. We use principle component analysis to reduce the feature space, and hierarchical clustering to group attack sources that exhibit similar behavior. Experiments with real world data illustrate that our framework can effectively reduce from large dataset to clusters of attack sources exhibiting critical collaborative patterns.

Keywords: Network security, Collaborative attacks, Degree centrality, Hierarchical clustering.

1 Introduction

The increasing and diverse vulnerabilities of operating systems and applications, as well as freely distributed cyber attack tools have led to significant volume of malicious activities on the Internet. Hackers can collaborate across the world or control zombie machines to perform Distributed Denial-of-service (DDoS) or other coordinated attacks. One of the common challenges in cyber attack analysis is the effectiveness in analyzing large volume of diverse attacks across the Interent. Ongoing research and practice has focused on investigating attacks via statistical analysis [12] [2], and defining traffic level similarities to cluster Internet hosts [11]. This set of work essentially answers the questions: "what are the most attacked services?" and "what are the most notorious attack sources and their profiles?". For Botnet attacks, which is a type of collaborative attacks, Gu et al. have proposed to use X-means clustering to analyze TCP/UDP traffic level features [5].

This paper analyzes general collaborative cyber attacks and aims at addressing the questions of "what is the relationship between attack sources?" and "what are the collaborative attack patterns?" by using only the source and target address information in a large dataset. We consider the UCSD Network

---

Telescope dataset [1], which monitors a Class-A network and provides malicious traffic that presumably represents $1/256$ of global attacks [8]. We extract the most fundamental information, source and destination address, from a 2-day data collected in Nov. 2008. An Attack Social Graph (ASG) is defined based on the extracted information. Borrowing the idea of degree centrality, six features are derived based on the ASG to describe the relationship among attack sources. Agglomerative hierarchical clustering are utilized to highlight critical attackers and group similar sources. The result shows our methodology can be used to provide abstract information on a group of attack sources and identify possible leader and conspirators from a large dataset.

2   Social Network Modeling of Cyber Attacks

Definition 1 (Attack Social Graph). *An Attack Social Graph $ASG_T(V, E)$ is a directed graph representing the malicious traffic within a time interval $[0, T]$, where a vertex $v \in V$ is a host, and an edge $e_{(u,v)} \in E$ exists if attacks are observed between u and v. Edge direction is from the attacking host to its target.*

Network Telescope does not respond to any requests, nor do they send any traffic within the Class-A network; Also, it does not collect traffic flowing outside Network Telescope. Therefore, ASG_T for the Network Telescope data is a bipartite graph. The vertex set V can be divided into two disjoint sets, V_s and V_t, which represent the set of attack sources and the set of attacked targets, respectively. Moreover, the edge directions can only be from $u \in V_s$ to $v \in V_t$. Let $d_{\mathrm{in}}(v)$ and $d_{\mathrm{out}}(v)$ denote the in-degree and out-degree of vertex v, respectively. The degree of vertex in V_s and V_t satisfies: $\forall u \in V_s$, $d_{\mathrm{in}}(u) = 0$ and $\forall v \in V_t$, $d_{\mathrm{out}}(v) = 0$.

ASG_T is comprised of many subgraphs that are disconnected from each other. Each subgraph denotes "local" attacks occurring between $V_s' \subset V_s$ and $V_t' \subset V_t$. The subgraphs vary significantly in size. Figure 1 shows two subgraph examples in ASG_5, where Fig. 1(a) contains $87,781$ vertices[1], and Fig. 1(b) contains 35 vertices.

Definition 2 (Attack Conspirators). *Let $T_u \doteq \{v \mid e_{(u,v)} \in E, \forall v \in V_t\}$ be the targets attacked by $u \in V_s$, the attack conspirators of $u \in V_s$, denoted as C_u, is a set of vertices: $C_u \doteq \{v \mid T_u \cap T_v \neq \emptyset, \ \forall v \in V_s\}$.*

Consider Fig. 1(b), host B is attacked by 20 sources, which is significantly higher than the average number of attacks, if targets are randomly chosen. Therefore, the sources attacking B might be collaborating or controlled as zombie machines. Similar to B, host C is attacked by a number of sources. Note that host A is a common culprit to attack both B and C which makes A suspicious to controlling all other attacking sources. The hosts who attack the same targets(s) as A does are referred to as the conspirators of A. With a more careful examination, A is the only host in Fig. 1(b), that has a large number of conspirators $|C_A|$, and for most $v \in C_A$, $d_{\mathrm{out}}(v) = 1$. This forms the basis of our algorithms that extract suspicious leaders of collaborative attacks.

---

[1] For better visualization, the vertices that has $d_{\mathrm{in}}(v) = 1$ are removed in Fig. 1(a).

(a) A Large Subgraph of ASG_5

(b) A Small Subgraph of ASG_5
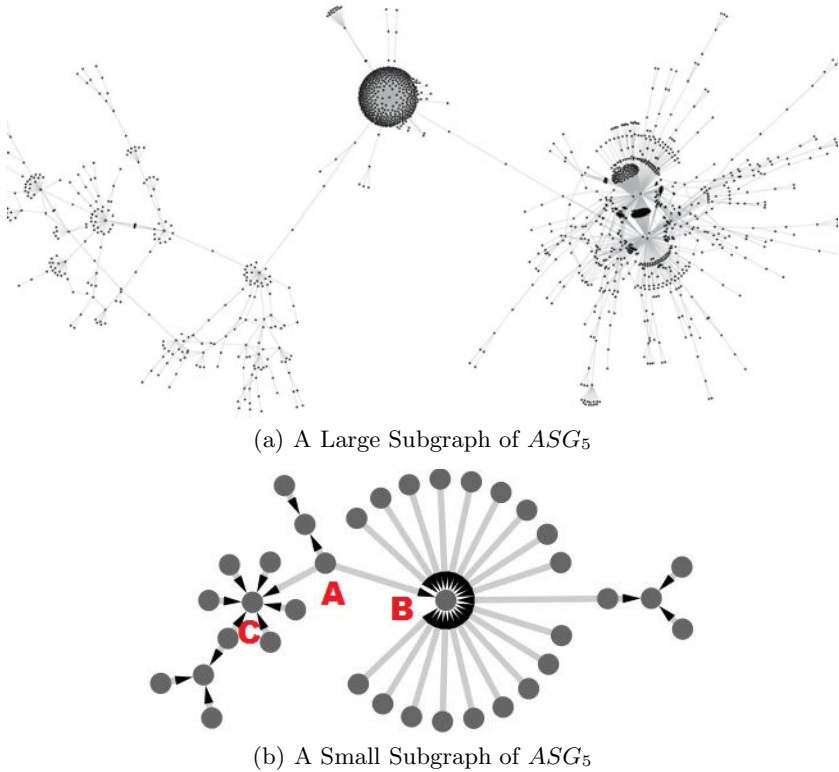
Fig. 1. Examples of Connected Subgraphs of ASG_5

3   Social Network Analysis of Collaborative Attacks

3.1   Subgraphs with Collaborative Attacks

ASG_T can be too large to be visually analyzed, Table 1 shows the size of ASG_T for different T, based on the UCSD Network Telescope dataset. Figure 2(a) plots $|V_s|, |V_t|$ and $|E|$ in log scale. The fact $|V_s| \ll |V_t|$ indicates a relatively small number of hosts attacking a Class-A network. While ASG_T is large, much of it do not contain collaborative attacks. The bipartite subgraphs can be categorized as: *one-attacks-one*, *one-attacks-many*, *many-attack-one*, and *many-attack-many* cases. Examples of these four types are given in Fig. 3 from left to right. *One-attacks-one* and *one-attacks-many* do not constitute collaborative attacks since $C_u = \emptyset$, for all attack sources u in these subgraphs.

Table 1 also shows the size of reduced ASG_T after removing the *one-attacks-one* and *one-attacks-many* cases. Figure 2(b) shows the proportion of remaining vertices and edges, i.e., $|V'_s|/|V_s|$, $|V'_t|/|V_t|$, $|E'|/|E|$ for different T values. The reduction seems less effective for a large T: almost 90% of attack sources has at least one conspirator at $T = 60$. This is due to the higher chance of selecting

Table 1. Attack Social Graph Size for Different Time Interval T (minutes)

	Original ASG_T			Reduced ASG_T		
T	$|V_s|$	$|V_t|$	$|E|$	$|V_s'|$	$|V_t'|$	$|E'|$
1	1,685	5,818	6,080	367	109	371
3	7,852	43,862	46,498	3,457	1,118	3,754
5	15,158	115,457	122,738	8,707	2,886	10,167
10	37,342	437,780	470,442	27,524	12,937	45,599
30	118,485	2,558,898	3,098,040	103,144	356,098	895,240
60	203,932	5,577,978	7,844,522	186,632	1,529,333	3,795,877

(a) Original Graph Size   (b) Proportion of Remaining Graph Size
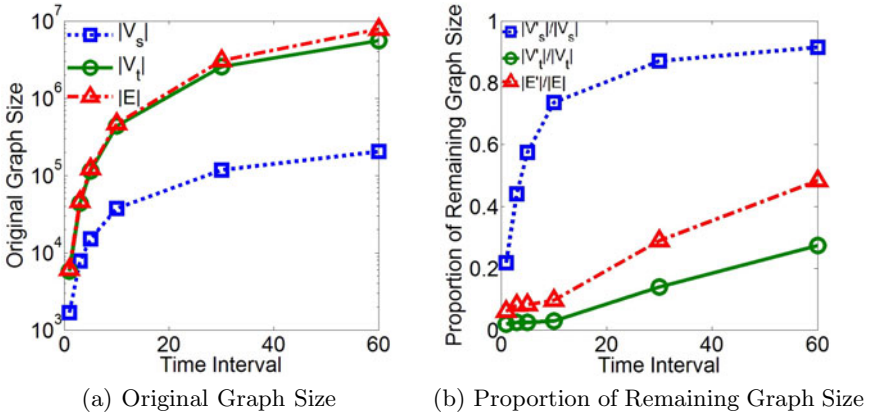
Fig. 2. Original Graph Size and Proportion of Remaining Graph Size
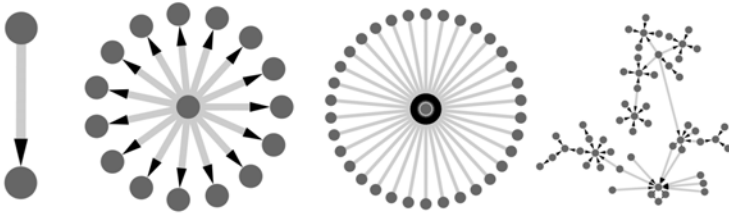
Fig. 3. Examples of ASG_T Subgraph Patterns

overlapping targets over a longer time interval. To reduce the possibility of ana-
lyzing collaborative patterns that contain targets due to random selections, we
suggest to analyze ASG_T for relatively small T, such as 5 to 10 minutes. This
paper considers $T = 5$ as an example to illustrate our methodology.

3.2   Feature Reduction

The reduced ASG_T contains collaborative attacks of different types. Borrowing
the idea from social network analysis, collaborative attacks are treated similarly
to co-authorship, in which, the authors have a lot of publications with different

co-authors are the "leaders" in the research community. We adopt Freeman's notion of degree centrality measurement [4] to derive six features for each $u \in V_s$ to differentiate sources of collaborate attacks in the reduced ASG_T. Consider the attack sources $u \in V_s$ in a reduced ASG_T, the six features are the number of targets $|T_u|$, the sum of the targets' in-degree $\sum_{x \in T_u} d_{\text{in}}(x)$, the standard deviation of targets' in-degree $\sigma_{x \in T_U} d_{\text{in}}(x)$, the number of conspirators $|C_u|$, the sum of conspirators' out-degree $\sum_{x \in C_u} d_{\text{out}}(x)$ and the standard deviation of conspirators' out-degree $\sigma_{x \in C_u} d_{\text{out}}(x)$. These six features are further analyzed using Principle Component Analysis (PCA) [7]. The usage of PCA produces uncorrelated principle components that gives the best separation among the attack sources.

As a result, two principle components are found, and the attack source set V_s' is further reduced into fewer "feature points" on a 2D plane. Each feature point represents one or more attack sources sharing the same degree centrality features. For example, the ASG_5 shown in Table 1 is reduced to 837 feature points, as shown in Fig. 5(a).

4   Clustering Analysis of Collaborative Patterns

The feature points on the 2D plane can be in close proximity and it is logical to perform clustering algorithms to discover similar collaborative patterns. For the reduced ASG5 dataset, most feature points are in cluttered regions on the 2D plane and few are outliers. When zooming into the cluttered region, similar distribution with cluttered regions and few outliers again occurs. This phenomenon can continue for several iterations. In order to better discover the community structure, Agglomerative Hierarchical Clustering [6] is used for analysis.

Figure 4 shows the resulting log-scale dendrogram with respect to the reduced ASG_5. The slow and steady staircase climbing exhibited in the left side of the log-scale dendrogram suggests that it is challenging to find the community structure from this data. Therefore, as discussed earlier, we investigate the dendrogram and the corresponding feature points by regions.

Figure 5(a) shows all the feature points of the reduced ASG_5 in the 2D plane and Fig. 5(b) plots the corresponding dendrogram showing only the top-30 branches. Plotting the top-30 branches reveals good community structure, as will be illustrated below. The four branches shown on the left of Fig. 5(b) actually
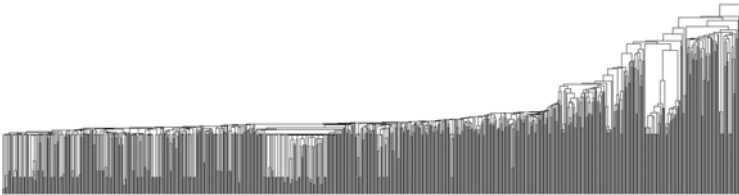
Fig. 4. The Log Scale (y-axis) Dendrogram of Feature Points on ASG_5

(a) Feature Points of Reduced ASG_5

(b) Dendrogram of Points for Fig. 5(a)

(c) Feature Points within Cluster A

(d) Dendrogram of Points for Fig. 5(c)

(e) Feature Points within Cluster D

(f) Dendrogram of Points for Fig. 5(e)
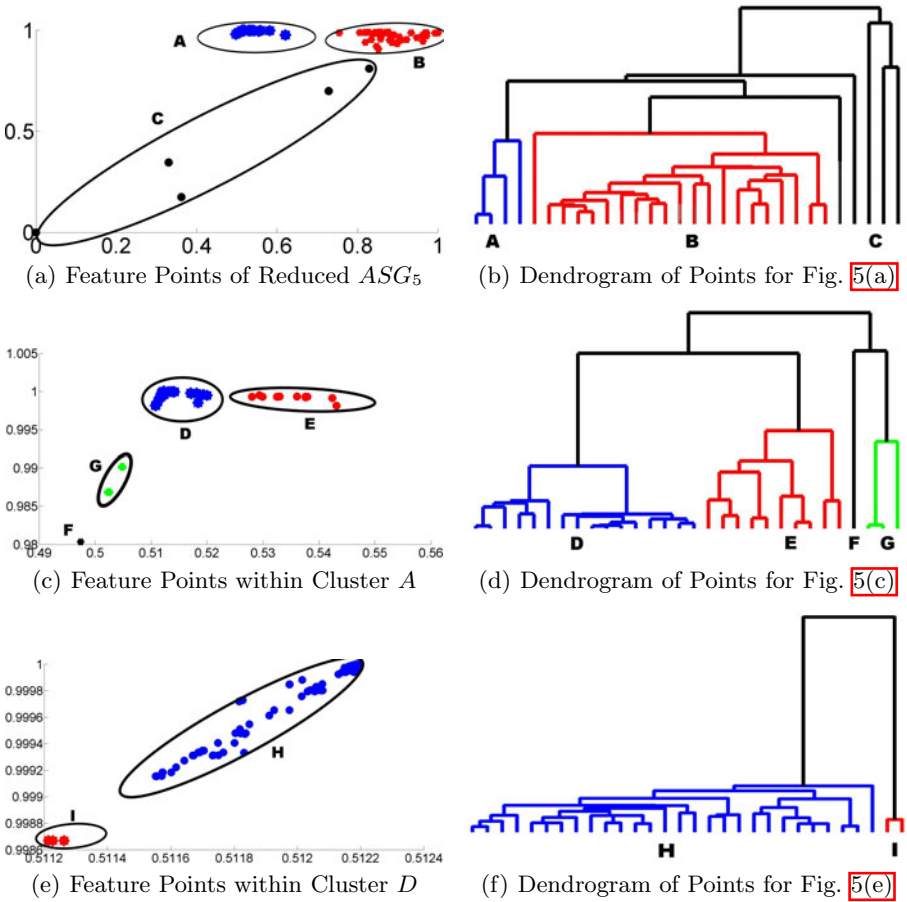
Fig. 5. Feature Points and The Corresponding Dendrogram

contains 777 out of the 837 points, i.e., the slow and steady staircase in Fig. 4. These 777 points are clustered as Cluster A and a zoom-in view of Cluster A is shown in Fig. 5(c). Figure 5(d) shows the corresponding dendrogram of feature points within Cluster A. Again only the top branches are chosen to reveal the community structure. The most dense region within Cluster A is Cluster D, and the feature points within Cluster D are further analyzed. Additional iterations can be run to reveal the sub-community structure. Figures 5(e) and 5(f) show the feature points in 10 th iteration and the corresponding dendrogram.

By analyzing the clusters and sub-clusters from Fig. 5, we find several sets of interesting collaborative attack patterns. First, the five feature points within Cluster C are actually five distinct attack sources that attack a large number of targets within the Class-A network monitored by the Network Telescope. These five sources are outliers on the 2D plane (see Cluster C in Fig. 5(a)), and a zoom-in plot of Fig. 1(a) is shown in Fig. 6(a) to identify these sources in the ASG_5.

(a) Outliers in Reduced ASG_5

(b) Sample Hosts in Cluster B

(c) Sample Patterns in Cluster H
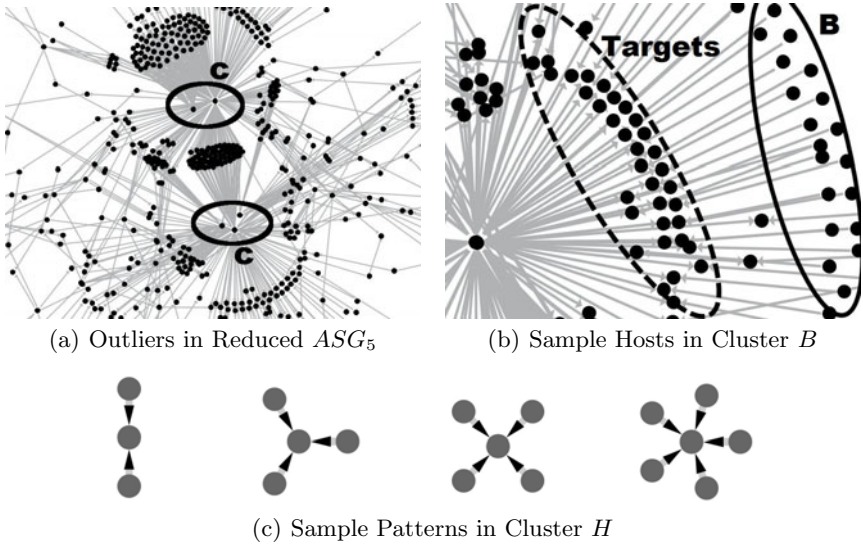
Fig. 6. An Example of Discovered Outliers and Clusters

Second, all sources in Cluster B have $\sum_{x \in C_u} d_{\text{out}}(x) > 10^4$. This is due to that these sources all have a "special" conspirator 125.211.214.160, which is an attack sources in Cluster C and its out-degree is $18,920$. Being a conspirator of such host makes their features significantly different from others and thus form a cluster. Figure 6(b) identifies some of the Cluster B sources in a zoom-in view of ASG_5. Third, Cluster A contains a large number of attack sources that are distinct from Cluster A and B, but do not share much common features. By analyzing the sub-regions within Cluster A, we can reveal some of the collaborative attack patterns. Figure 6(c) shows several examples of "several-attack-one" patterns identified in Cluster H – recall Fig. 5(e) and 5(f).

5   Conclusion and Future Work

Going beyond classical statistical approaches, cyber attack analysis can benefit from social network analysis. This work defined Attack Social Graph to provide an effective representation of relationship among attack sources. Borrowing the idea of degree centrality measurement and agglomerative hierarchical clustering, we discover various types of collaborative attack patterns. Our framework effectively reduces from large dataset and reveals possible leaders of collaborative cyber attacks.

Future extension and ongoing work include the use of weighted ASG, where the weights can be defined based on the number of packets exchanged. In addition, a moving time window allows for analysis of how collaborative patterns evolve. Finally, other features, such as geographical and application level information can be used to further differentiate collaborative attacks.

References

1. Aben, E., et al.: The CAIDA UCSD Network Telescope Two Days in November 2008, Dataset (2008), http://www.caida.org/data/passive/telescope-2days-2008_dataset.xml

2. Allman, M., et al.: A brief history of scanning. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, p. 82 (2007)

3. Batagelj, V., Mrvar, A.: Pajek-program for large network analysis. Connections 21(2), 47–57 (1998)

4. Freeman, L.: A set of measures of centrality based on betweenness. Sociometry 40(1), 35–41 (1977)

5. Gu, G., et al.: Botminer: clustering analysis of network traffic for protocol and-structure independent botnet detection. In: Proceedings of the 17th Conference on Security Symposium, pp. 139–154 (2008)

6. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Computing Surveys 31(3), 264–323 (1999)

7. Jolliffe, I.: Principal component analysis. Springer Series in Statistics (2002)

8. Moore, D., Shannon, C., Voelker, G., Savage, S.: Network telescopes: Technical report. CAIDA (2004)

9. NWB Team: Network workbench tool. Indiana University, Northeastern University, and University of Michigan (2006), http://nwb.slis.indiana.edu

10. Shannon, P., et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13(11), 24–98 (2003)

11. Wei, S., et al.: Profiling and clustering internet hosts. In: Proceedings of the International Conference on Data Mining (2006)

12. Yegneswaran, V., et al.: Internet intrusions: Global characteristics and prevalence. In: Proceedings of the International Conference on Measurement and Modeling of Computer Systems, p. 147 (2003)

Agents That Negotiate Proficiently with People

Sarit Kraus

Dept. of Computer Science Bar-Ilan University, Ramat Gan 52900 Israel and
Institute for Advanced Computer Studies University of Maryland, MD 20742 USA
`sarit@umiacs.umd.edu`

Abstract. Negotiation is a process by which interested parties confer with the aim of reaching agreements. The dissemination of technologies such as the Internet has created opportunities for computer agents to negotiate with people, despite being distributed geographically and in time. The inclusion of people presents novel problems for the design of autonomous agent negotiation strategies. People do not adhere to the optimal, monolithic strategies that can be derived analytically, as is the case in settings comprising computer agents alone. Their negotiation behavior is affected by a multitude of social and psychological factors, such as social attributes that influence negotiation deals (e.g., social welfare, inequity aversion) and traits of individual negotiators (e.g., altruism, trustworthiness, helpfulness). Furthermore, culture plays an important role in their decision making and people of varying cultures differ in the way they make offers and fulfill their commitments in negotiation.

In this talk I will present the following two agents that negotiate well with people by modeling several social factors: The PURB agent that can adapt successfully to people from different cultures in complete information settings, and the SIGAL agent that learns to negotiate successfully with people in games where people can choose to reveal private information. These agents were evaluated in extensive experiments including people from three countries. I will also demonstrate how agents' opponent modeling of people can be improved by using models from the social sciences.

References

1. Gal, Y., Grosz, B., Kraus, S., Pfeffer, A., Shieber, S.: A framework for Investigating Decision-Making in Open Mixed Networks. Artificial Intelligence Journal 174(18), 1460–1480 (2010)
2. Gal, Y., Kraus, S., Gelfand, M., Khashan, H., Salmon, E.: Negotiating with People across Cultures using an Adaptive Agent. ACM Transaction on Intelligent Systems and Technology (2011)
3. Peled, N., Kraus, S., Gal, K.: A Study of Computational and Human Strategies in Revelation Games. In: Proc. of AAMAS 2011 (2011)

Cognitive Modeling for Agent-Based Simulation of Child Maltreatment

Xiaolin Hu[1] and Richard Puddy[2]

[1] Department of Computer Science, Georgia State University, Atlanta, GA 30303
[2] Atlanta, GA 30341

Abstract. This paper extends previous work to develop cognitive modeling for agent-based simulation of child maltreatment (CM). The developed model is inspired from parental efficacy, parenting stress, and the theory of planned behavior. It provides an explanatory, process-oriented model of CM and incorporates causality relationship and feedback loops from different factors in the social ecology in order for simulating the dynamics of CM. We describe the model and present simulation results to demonstrate the features of this model.

Keywords: Child maltreatment, cognitive modeling, agent-based simulation, parental efficacy, parenting stress, theory of planned behavior.

1 Introduction

Child maltreatment (CM) negatively impacts child development and results in long term health problems as well. Understanding and preventing CM is a challenging task due to the dynamic and complex nature of this phenomenon. The complexity results from a system containing multi-level (individual, relationship, community, societal) factors across the social ecology (Belsky, 1980), diversity of actors (such as families, schools, government agencies, health care providers) that potentially affect maltreatment, and multiplicity of mechanisms and pathways that are not well studied or understood. Complex systems science and agent-based modeling offer tremendous promise in this area because they have proven to be a powerful framework for exploring systems with similar characteristics (Hammond, 2009).

Using a complexity science-informed approach, in previous work (Hu and Puddy, 2010) we developed an agent-based model (ABM) for studying CM and CM prevention. The ABM uses a resource-based conceptual model to simulate the occurrence of CM. It models a community of agents, each of which corresponds to a family unit and includes a parent-child relationship. CM occurs when there exists unmet child need, i.e., the parent care is insufficient to meet the child need. Agents are connected to each other through social network, which acts as a potential resource for agents to obtain support in providing child care. The developed ABM focuses on the community level of the social ecology. It was able to simulate the impacts of several community level factors such as the density of social connection, community openness, and community resource on the rate of CM (Hu and Puddy, 2010).

While the previous work built a foundation for applying agent-based modeling and simulation to studying CM, extensions are needed in order to adequately model the dynamics of CM. In particular, in the previous work each agent itself is pretty simple in the sense that it only models the resource aspect to account for the occurrence of CM. Although resource is an important factor in CM, the occurrence of CM is a result of a cognitive process that impacts parents' decision to engage in aggression toward children (Milner, 1993, 2000). Thus in order to adequately model the dynamics and the complexity of CM it is necessary to build a cognitive process into the agent model. Motivated by this need, this paper presents cognitive modeling of CM. The developed cognitive model is inspired from several sources of cognitive science and CM psychology, including the Theory of Planned Behavior (TPB) (Ajzen, 1985; Ajzen, 1991), self-efficacy theory (Bandura, 1977; Bandura, 1994), and models of parenting stress (Belsky, 1984; Hillson and Kuiper, 1994). TPB postulates three conceptually independent determinants of behavioral intention: attitude toward the behavior, subjective norm, and the perceived behavioral control. Among them the perceived behavior control is originated from the concept of self efficacy. For the phenomenon of CM, a specialized type of self-efficacy, parental efficacy, can be defined as the extent to which parents believe they can influence the context in which their child is growing (Shumow and Lomax 2002). This concept of parental efficacy is used in our model to influence the perceived behavior control in TPB. A brief introduction of these theories/models is omitted here and can be found in an extended version of this paper (Hu and Puddy, 2011). Our work of cognitive modeling aims to provide an explanatory, process-oriented model of CM, incorporating causality relationship and feedback loops from different factors in the social ecology of CM. It is hoped that the developed model will support more accurate agent-based simulation of CM and provide more accountable insights for preventing CM.

It is important to note that there are different types of child maltreatment, including physical abuse, emotional abuse, sexual child abuse, and neglect. While these different types of maltreatment may follow different cognitive pathways, in this work we are less concerned about the specific type of maltreatment. Instead, as a simulation model aiming towards a computational analysis of the phenomenon of CM, this work pays more attention to a general structure of the cognitive process of CM, regardless its specific format. Because of this, in our paper we use abstract terms "*responsive to child need*" and "*not responsive to child need*" to describe parents' behavior, where "*not responsive to child need*" means maltreatment. With that said, however, we note that the development of our model is mainly informed from the physical abuse and neglect types of maltreatment, and thus may not be meaningful for the other maltreatment types. Also note that this paper focuses on the modeling of a single agent. Multi-agents interactions and community level factors can be added later (see, e.g., Hu and Puddy, 2010) and are not explicitly considered in this paper.

2   Cognitive Modeling of Child Maltreatment

In developing agent's cognitive model of child maltreatment, we employ TPB as a basic mechanism to compute the behavioral *intention* of the agent (regarding being responsive or not responsive to child need). A significant portion of the cognitive model deals with how to compute the *perceived behavior control* used in TPB. In our

work, the perceived behavior control is the result of *parental efficacy* modulated by the *stress level*. Fig. 1 shows the major elements and their relationships of the cognitive model, followed by the description of how the model works. Due to space limitation, the description focuses on the computational aspect by showing how the major elements are computed and updated. Corresponding mathematical equations are omitted in this paper, and can be found in the Appendix of an extended version of this paper (Hu and Puddy, 2011). We note that the purpose of this paper is to explore the possible dynamics of CM without an attempt to have fidelity to real world observations (real world fidelity and validation will be attempted in later phases). Because of this, some parameter values (e.g., coefficients, weights) used in the equations are empirically determined based on what we think are reasonable within the context of CM. Due to the same reason, we simplify the quantification of all the elements of our model (except for *parenting stress* and *contextual stress*) by representing them on an arbitrary scale of 0-100. The units of the scale are purposely not specified in order to simply represent relative values of the variables under study. Also note that in the following description, elements such as *child need*, *family resource*, and *community* (with *social network*) have been defined in previous work (Hu and Puddy, 2010) and thus are used in this paper with minimal explanations.
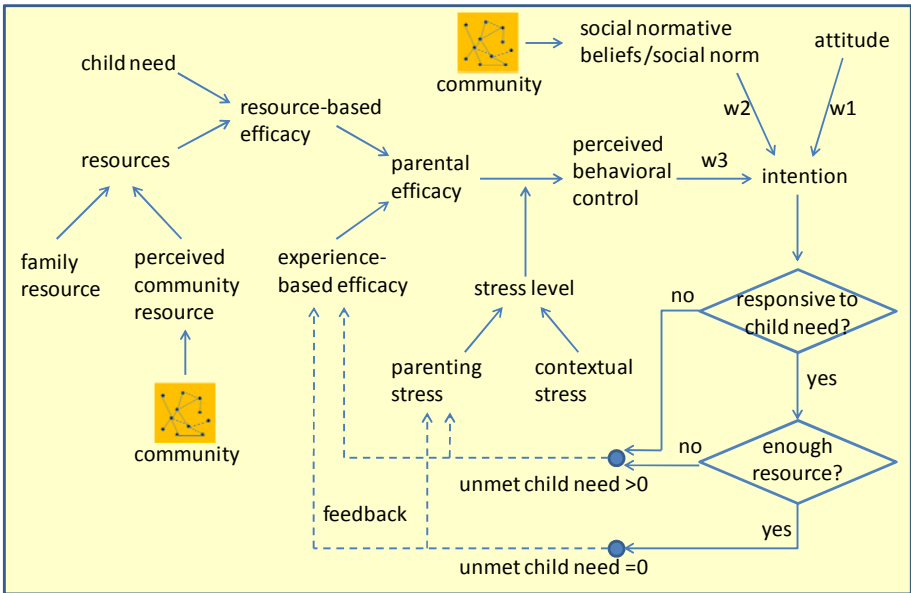
Fig. 1. Cognitive modeling of child maltreatment (of a single agent)

Parental Efficacy

An agent's *parental efficacy* defines the agent's belief about its capability of taking care of the child (i.e., meeting the child's need). In our work, an agent's parental efficacy is formed from two sources: the *resource-based efficacy* and the *experience-based efficacy*. The resource-based efficacy captures the portion of efficacy induced from the agent's *resource* compared to the *child need*. It thus depends on the

difference between the resource and child need – the larger the difference, the higher the resource-based efficacy. In our model, this relationship is captured by a linear equation between the resource-based efficacy and the difference between resource and child need. Note that an agent's resource can include both *family resource* and *perceived community resource*. The later depends on community level factors such as social network and community resource (Hu and Puddy, 2010).

The experience-based efficacy captures the portion of efficacy induced from the agent's past experience of taking care of the child. The experience-based efficacy changes over time and is dynamically updated based on the agent's experience of meeting the child need. In general, if the agent successfully meets the child need, its experience-based efficacy increases; otherwise, its experience-based efficacy decreases. In our model, the experience-based efficacy increases or decreases in a predefined rate based on whether the child need is met or not. A coefficient is used to model the ceiling effect when increasing experience-based efficacy: as the experience-based efficacy becomes larger, the increase rate of the experience-based efficacy becomes slower.

The parental efficacy is the weighted sum of the resource-based efficacy and experience-based efficacy. We assign a larger weight for the experience-based efficacy because we consider parent-child interactions as routine events. Under routine circumstances, an individual may well utilize previous performance level (obtained from experience) as the primary determinant of self-efficacy (Gist and Mitchell, 1992).

Stress Level

An agent's *stress level* comes from two different sources: *contextual stress* and *parenting stress*. Belsky (1984) has suggested that there are three contextual sources of stress: the marital relationship, social networks, and employment that can affect parenting behavior. In our work, these stresses are modeled by a single entity called contextual stress and is assumed to be unchanged during the simulation. Besides the contextual sources of stress, parenting stress (i.e., stress specific to the parenting role) also directly and indirectly affected parenting behavior (Rodgers, 1998). Parenting stress results from the disparity between the apparent demands of parenting and the existing resources to fulfill those demands (Abidin, 1995). In our model, the parenting stress is due to being a parent for satisfying the child need. It is dynamic in nature as described below: 1) when the amount of parent care provided by the agent is much less than the family resource (this means, e.g., the agent can "easily" satisfy the child need), the parenting stress decays over time– it decreases from the previous value; 2) otherwise when the amount of parent care is close to the full capacity of family resource (this means, e.g., the agent needs to exert its full effort in order to satisfy the child need), the parenting stress accumulates over time – it increases from the previous value. In our model, we use a threshold ratio 80%: when more than 80% of the family resource is used, parenting stress increases; otherwise the parenting stress decreases. Also note that in the current model, when an agent chooses "*not responsive to child need*", it uses zero of its family resource (as in neglect). Thus in this case, the parenting stress decreases because the agent is not engaged in the parenting behavior. An agent's stress level is the sum of its contextual stress and parenting stress.

Behavioral Intention

Based on TPB, an agent's behavioral *intention* is calculated from *attitude*, *social norm*, and *perceived behavior control* (denoted as *PBC*), multiplying by their corresponding *weights*. The attitude indicates the agent's general tendency or orientation on the parenting behavior. The social norm reflects the social context of the agent, and can be computed from the agent's social network (e.g., the average of all neighbors' attitudes). The PBC is the result of parental efficacy modulated by the stress level. In our model, the modulation is modeled by a stress-based coefficient that modulates the parental efficacy into the PBC. In general, the higher the stress level, the smaller the stress-based coefficient. In computing the behavioral intention, three weights are used to define the relative percentages of contribution of the three elements: attitude, social norm, and PBC in computing the intention. In our implementation, these weights are determined based on the level of PBC: when an agent has high level of PBC, the attitude, social norm, and PBC have about the same weight; as the level of PBC drops, the weight for the PBC becomes larger. This approach, to some extent, incorporates the idea of "hot" cognition (Abelson, 1963), which is a motivated reasoning phenomenon in which a person's responses (often emotional) to stimuli are heightened.

Based on its attitude, social norm, and PBC, an agent's behavioral intention is the weighted sum of them. This intention determines the agent's probability of choosing the "*responsive to child need*" behavior. Thus the higher the intention level, the more likely the agent will be engaged in the parenting behavior for meeting the child need. Finally, if the "*responsive to child need*" behavior is chosen, the agent will check if it actually has the resource (family resource plus community resource) to meet the child need. Only when this is true, the child need is successfully satisfied. Otherwise (either if the "*not responsive to child need*" is chosen or there is insufficient resource), there exists unmet child need, i.e., child maltreatment.

Simulation of the agent model runs in a step-wise fashion. In every step, the parental efficacy is computed from the current experience-based efficacy and resource-based efficacy (based on the difference between resource and child need). The stress level is also computed from the contextual stress and the current parenting stress. Then the perceived behavior control is calculated, and the behavioral intention is calculated to determine the probability of choosing "*responsive to child need*" behavior. Finally, due to the feedback loops, the experience-based efficacy is updated based on if the child need is met or not, and the parenting stress is updated based on the parent care to family resource ratio.

3   Simulation Results

We carried out a series of simulation experiments to illustrate the fundamental features of our cognitive model and see what insights we might gain by them. Our experiments aim to provide qualitative understanding of the model dynamics, *i.e.*, to explore the potential directional impact on CM rates when efficacy, stress, and family resource are varied. To do this, we fix the other model parameters as follows: attitude =90, social norm =90, child need =50, perceived community resource =0, initial value of parenting stress =0. The value we chose for each of these elements was designed to reflect a rough cut at the relative contribution of that element in relation to the others.

Results in this section are based on simulations using a single agent. Due to space limitation, the result of family resource is not shown in this paper.

Our first experiment shows how the initial value of experienced-based efficacy impacts the CM result. In this experiment, we set a high contextual stress 20 (66.7 in a 100 scale), and set the family resource to be 55, which is enough to meet the child need (=50) but still results in parenting stress because it is not significantly larger than the child need. Fig. 2(a) shows how the *parenting stress*, *total stress level* (parenting stress plus contextual stress), *experience-based efficacy* (denoted as *EBE*), *perceived behavior control* (denoted as *PBC*), and the *intention* level change over time for 2000 time steps when the EBE starts with an initial value of 30. Fig. 2(b) shows the same information when the *EBE* starts with an initial value of 40. The two figures show two clearly different behavior patterns due to the different initial value of *EBE*. In Fig. 2(a), since the initial EBE of 30 is relatively small, the agent's intention of choosing "*responsive to child need*" is not high enough, leading to significant number of CM. The large number of CM in turn decreases the EBE further, resulting in an adverse self-reinforcing loop that makes EBE rapidly decreases to zero. On the contrary, in Fig. 2(b) the initial EBE of 40 was able to make the agent have high enough intention so the experience of meeting the child need becomes dominant. This in turn increases the agent's EBE further, resulting in a favorable self-reinforcing loop that gradually raises EBE, PBC, and the intention to a high level. Note that as the agent becomes more and more engaged in the parenting behavior, the parenting stress gradually increases too. The influence of parenting stress will be studied in the next experiment.

(a) Initial experience-based efficacy=30   (b) initial experience-based efficacy=40
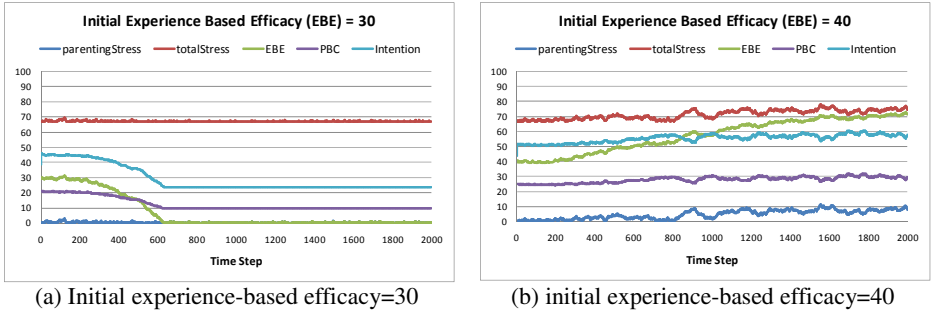
Fig. 2. Dynamic behavior for different initial experience-based efficacy values

To further see the impact of the initial EBE, we vary the initial EBE from 10 to 100 (increase by 10 each time) and carry out the simulations. Fig. 3 shows the dynamic change of EBE over time for different initial values. As can be seen, when the initial value is 30 or smaller, the EBEs all decrease to 0 due to the adverse self-reinforcing loop explained above. When the initial value is 40 or more, the EBEs either gradually increase or gradually decrease and eventually all converge to the same value (about 72). This experiment reveals two important properties of the system impacted by EBE. First, the initial EBE can play a "critical" role in defining the overall experience of parenting – a relative small change in the initial EBE (e.g., from 30 to 40) may result in significantly different system behavior. Second, regardless the initial values,
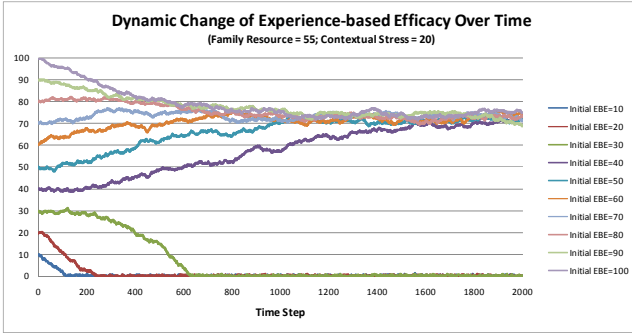
Fig. 3. The impact of different initial experience-based efficacy values
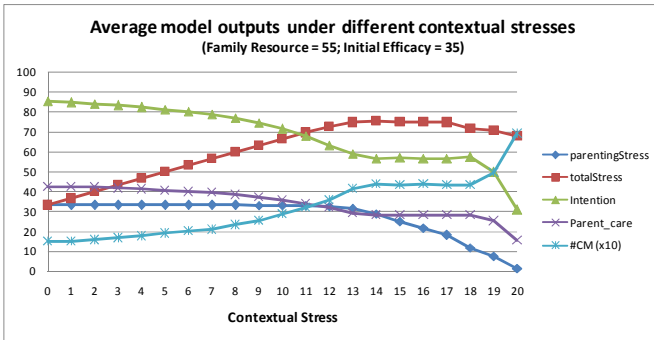
Fig. 4. The impact of contextual stress

eventually the system always converge to one of the two relatively stable states, that is, to two attractors. This type of bifurcation (convergence of the system on one of two relatively stable states) is common in complex systems. Observing this phenomenon here points to the need for further exploration of this system in future work.

Our second experiment tests the impact of the contextual stress on the CM result. In this experiment, we set the family resource to be 55, the initial experience-based efficacy to be 35. We vary the contextual stress from 0 to 20 (increase by 1 each time) and run the simulations for 2000 time steps. For each simulation we ran the model 10 times and averaged the results which we present in Fig. 4. Since it takes time for the system to reach a stable state as shown above, we only measure the results between time 1000 and time 2000, and then compute the average of the 1000 time steps. Fig. 4 shows the averaged *parenting stress*, *total stress*, *intention*, *parent care*, and *number of CM*. Note the parent care is the total amount of parent care averaged over the 1000 time steps, and the number of CM is the total number of times when there are unmet child need averaged over the 1000 time steps. From the figure we can see that the contextual stress has important impacts on CM. In most situations, when the contextual stress reduces, the number of CM decreases too. However, counter intuitively, this is not always true, as in Fig. 4 when the contextual stress reduces from 18 to 14, the number of CM maintains at the same level. This is because the reduction of

contextual stress causes the increase of parenting stress (as shown in Fig. 4) due to the increase of intention. Together they make the overall stress level stay at a constant level. It should be noted that this result does not mean there is no benefit to reduce the contextual stress from 18 to 14 – as the contextual stress is reduced further to below 14, the number of CM decreases too. This experiment shows that the interplay between contextual stress and parenting stress can cause complex, and sometimes counter-intuitive, system behavior.

4 Conclusions

This paper extends previous work to develop cognitive modeling for agent-based simulation of CM. The developed model is inspired from the theory of planned behavior, self-efficacy theory, and parenting stress models. Preliminary results using a single agent are presented. Future work includes more analysis and tests of the cognitive model, adding community level factors to carry out multi-agent simulations, and aligning the model parameters with real world data and model validation.

References

1. Abelson, R.P.: Computer simulation of "hot cognition". In: Tomkins, S.S., Messick, S. (eds.) Computer Simulation of Personality, pp. 277–302. Wiley, New York (1963)
2. Abidin, R.R.: Parenting Stress Index: Professional Manual, 3rd edn. Psychological Assessment Resources, Inc., Odessa (1995)
3. Ajzen, I.: From intentions to actions: A theory of planned behavior. In: Kuhl, J., Beckman, J. (eds.) Action-control: From Cognition to Behavior, pp. 11–39. Springer, Heidelberg (1985)
4. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes 50, 179–211 (1991)
5. Bandura, A.: Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review 84, 191–215 (1977)
6. Bandura, A.: Self-efficacy. In: Ramachaudran, V.S. (ed.) Encyclopedia of Human Behavior (Vol, pp. 71–81. Academic Press, New York (1994); Friedman, H. (ed.): Encyclopedia of mental health. Academic Press, San Diego (1998),
http://www.des.emory.edu/mfp/BanEncy.html (last accessed: 11/05/2010)
7. Belsky, J.: Child maltreatment: An ecological model. American Psychologist 35(4), 320–335 (1980)
8. Belsky, J.: The determinants of parenting: A process model. Child Development 55, 83–96 (1984)
9. Hillson, J.M.C., Kuiper, N.A.: A stress and coping model of child maltreatment. Clinical Psychology Review 14, 261–285 (1994)
10. Hu, X., Puddy, P.: An Agent-Based Model for Studying Child Maltreatment and Child Maltreatment Prevention. In: Chai, S.-K., Salerno, J.J., Mabry, P.L. (eds.) SBP 2010. LNCS, vol. 6007, pp. 189–198. Springer, Heidelberg (2010)

11. Hu, X., Puddy, P.: Cognitive Modeling for Agent-based Simulation of Child Maltreatment, extended paper (2011),
    http://www.cs.gsu.edu/xhu/papers/SBP11_extended.pdf
12. Gist, M.E., Mitchell, T.R.: Self-Efficacy: A Theoretical Analysis of Its Determinants and Malleability. The Academy of Management Review 17(2), 183–211 (1992)
13. Milner, J.S.: Social information processing and physical child abuse. Clinical Psychology R&w 13, 275–294 (1993)
14. Milner, J.S.: Social information processing and child physical abuse: Theory and research. In: Hansen, D.J. (ed.) Nebraska Symposium on Motivation, vol. 45, pp. 39–84. University of Nebraska Press, Lincoln (2000)
15. Shumow, L., Lomax, R.: Parental efficacy: Predictor of parenting behavior and adolescent outcomes. Parenting, Science and Practice 2, 127–150 (2002)

Crowdsourcing Quality Control of Online Information: A Quality-Based Cascade Model

Wai-Tat Fu and Vera Liao

Department of Computer Science, University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
`wfu/liao28@illinois.edu`

Abstract. We extend previous cascade models of social influence to investigate how the exchange of quality information among users may moderate cascade behavior, and the extent to which it may influence the effectiveness of collective user recommendations on quality control of information. We found that while cascades do sometimes occur, their effects depend critically on the accuracies of individual quality assessments of information contents. Contrary to predictions of cascade models of information flow, quality-based cascades tend to reinforce the propagation of individual quality assessments rather than being the primary sources that drive the assessments. We found even small increase in individual accuracies will significantly improve the overall effectiveness of crowdsourcing quality control. Implication to domains such as online health information Web sites or product reviews are discussed.

Keywords: Web quality, user reviews, information cascades, social dynamics.

1 Introduction

The rapid growth of the WWW has led to increasing concern for the lack of quality control of online information. Many have suggested that quality assessments of online information could be "crowdsourced" to the public by harnessing the collective intelligence of online users by allowing them recommend contents and share the recommendations with others. Although this kind of crowdsourcing seems to work well in many domains [2, 3, 9], recent studies, however, show that aggregate behavior is often subject to effects of cascades, in which the social dynamics involved during the accumulation of recommendations may not be effective in filtering out low-quality information [8]. Indeed, previous models show that when people imitate choices of others, "bad cascades" may sometimes make low-quality information more popular than high-quality ones [1].

Previous cascade models often assume that users imitate choices of others without direct communication of quality information. We extend previous models by assuming that users can make multiple choices among web sites, and they can communicate their quality assessments to other users. Our goal is to understand how individual quality assessments may moderate effects of information cascades. We investigate how cascades may be influenced by factors such as accuracies of individual quality assessments, confidence levels of accepting other users' recommendations, and impact of

aggregate user recommendations on choices. The goal is to simulate effects at the individual level to understand how the social dynamics may lead to different aggregate patterns of behavior. Specifically, we investigate how individual quality assessments propagate in a participatory Web environment, and whether information cascades may moderate the effectiveness of aggregate user recommendations, such as whether they may lead to "lock in" to low-quality information; and if so, to what extent will they impact the overall effectiveness of crowdsourcing quality control of information.

2 Background

2.1 Quality Assessment of Online Information

While research has shown that people are in general good at choosing high quality Web sites, it is also found that people often browse passively by relying on what is readily available. For example, health information seekers are found to rely on search engines to choose online sources, and may utilize surface features (such as layout and structures of web pages) to infer quality [5]. Accuracies of quality assessments are also found to vary among individuals, and can be influenced by various factors such as background knowledge, Internet experience, cognitive resources available, etc [4]. Liao & Fu [7] also found that older adults, who tended to have declined cognitive abilities, were worse in assessment quality of online health information.

Research in the field of e-commerce shows that user reviews can often significantly influence choices of online information, and their impact may be biased by "group opinions" while ignoring their own preferences [3]. Research also shows that effects of cascades and social dynamics may influence effectiveness of user reviews [8] in guiding users to choose high-quality products. However, to our knowledge, none has studied how cascades may be influenced by accuracies of quality assessment at the individual level, which may lead to different emergent aggregate behavioral patterns. In fact, research has shown that in certain situations, even random accumulation of user recommendations may induce more users to follow the trend, creating a "rich gets richer" cascading effect that distorts the perception of quality. On the other hand, previous cascade models often do not allow users to pass their quality assessments to other users, and it is possible that the passing of quality information may moderate effects of cascades. The goal of the current model is to test the interactions between individual and aggregate quality assessments to understand the effectiveness of collective user recommendations on promoting high quality information.

2.2 Cascade Models

Many models of information cascades have been developed to understand collective behavior in difference contexts, such as information flow in a social network [6], choices of cultural products in a community [1], or effectiveness of viral marketing in a social network [10]. In the model by Bikhchandani et al [1], information cascades are modeled as imitative decision processes, in which users sequentially choose to accept or reject an option based on their own individual assessment and the actions of other users. Each user can then observe the choice made by the previous users and infer the individual assessments made by them. In most information cascade models,

the users can only observe the choices made by previous users and learn nothing from the users about the quality of the options. Given this limited information, a string of bad choices can accumulate, such that a cascade might lock future users on a low-quality option. Similarly, cascade models of information flow aim at characterizing how likely information will spread across a social network [6, 10] without any specific control on how likely one may assess the quality of the information and how the passing of the quality information to other users may influence how it spreads to other people in the network. It is therefore still unclear how the passing of quality assessments by users, in addition to their choices, might influence information cascades, and how the extent to which the aggregate quality assessments can be relied on to filter out low-quality information.

3   The Model

The general structure of the model is to have a sequence of users each deciding whether to select one of the 20 sites (or select nothing). When user i selects a site j, a recommendation will be given to the site based on the user's imperfect quality assessment. User $i+1$ will then interpret the recommendation by user i with a certain level of confidence to determine whether to select the same site j or not. If user i+1 does not select j, she will either select a site from a recommended list of sites based on aggregate recommendation by previous users or select nothing. The model therefore aims at capturing the interaction of "word of mouth" information from the neighboring users and the global assessment of quality by previous users.

3.1   The Web Environment

We simulated the environment by creating sets of 20 Web sites, with each set having different proportions of sites with good and bad quality. A user can freely pick any site (or not pick any) and give a positive or negative recommendation (or no recommendation) to it. This recommendation is available for the next user, acting as a local signal (i.e., word-of-mouth information) collected from neighboring users. Recommendations will also be accumulated and used to update the list of most recommended sites, which will be available for the next users. The choice of sites and recommendation are therefore influenced by three factors: (1) the perceived quality of the site by the user (the private signal), (2) the recommendation provided by the previous user (the local signal), and (3) the accumulated list of most recommended sites (the global signal).

3.2   A Quality-Based Cascade Model

We simulated the sequential choice and recommendation process by a set of 1000 users as they chose and gave recommendation to the set of 20 Web sites. When the simulation began, a user was randomly assigned to a site. The user then evaluated the site, and based on the user's perceived quality, it will assign a positive or negative recommendation to the site. Specifically, it was assumed that when the site had good quality, there was a probability Q (or $1-Q$) that the user would perceive that it had good (or bad) quality. When the perceived quality was good (or bad), there was a

probability *R* (or *1-R*) that the user would give a positive (or negative) recommendation. The next user could then interpret the recommendation given by the previous user to infer the quality of the site. If the recommendation was positive (or negative), there was a probability *C* (or *1-C*) that the next user would choose (or not choose) the same site. If the user did not choose the same site, there is a probability *L* (or *1-L*) that it would randomly select from the top recommended list (or not choose any site). Fig. 1 shows the decision tree of the model.
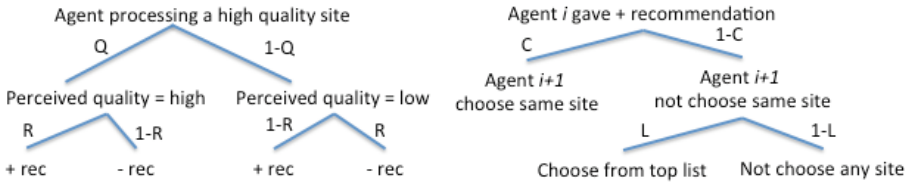
Fig. 1. The quality-based cascade model

3.3   Choice Accuracies

Fig. 2 shows the mean proportions of choice of high quality sites by the users with different values of Q (0.5, 0.55, 0.6) in a High (20% low), Medium (50% low), and Low (80% low) quality environments, with (a) low (C=0.5) and (b) high (C=0.9) confidence levels. In the "Top1" environment, when users decided to select from the recommendation list, she always picked the site with the highest recommendation; in the "Top10" environment, the user randomly picked a site among the best 10 sites.

   Results show significant interactions among the private, local, and global signals. Good individual quality assessments (i.e., Q) in general lead to better overall accuracies in choosing high quality sites, but its effect on choice is strongly magnified by both the global signal (i.e., the recommendation list) and the confidence level in neighboring users' recommendation (i.e., C). Even when the confidence level is low (C=0.5), good quality assessments can be propagated through the aggregate recommendation list, assuming that other users will select the top recommended sites (in the Top1 environment). However, when users randomly selected sites from a longer list (Top10 environment), this channel of propagation diminished quickly, as shown by the low accuracies in the Top10 environment (accuracies in Top5 were somewhere between Top1 and Top10). When the confidence level was high (C=0.9), good quality assessments can be propagated through "word of mouth" local information, which improved the overall accuracies even in the top-10 environment. In general, even a small increase of Q from 0.5 (random choice) to 0.55 will lead to significant increase in overall choice of the correct Web sites for the aggregate. *Results demonstrate how local and global signals can magnify individual assessment of quality to increase the overall effectiveness of crowdsourcing quality control.*

3.4   Effects of Information Cascades

Fig. 3 shows examples of the simulations that illustrate whether information cascades occur in different conditions. Each figure shows the aggregate (sum of positive and negative) recommendations of the 20 web sites (different color lines) in the Medium
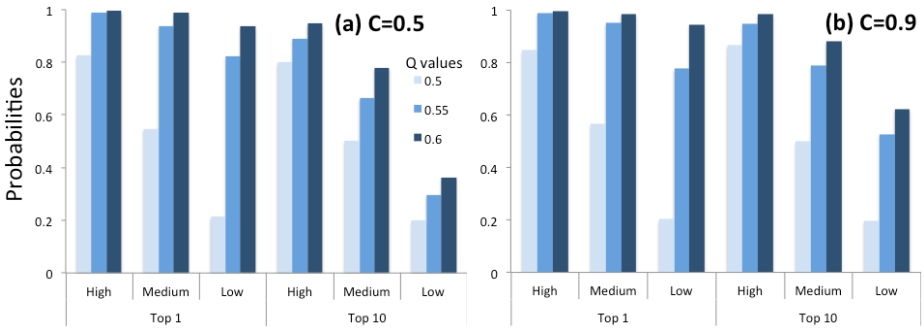
Fig. 2. Choice accuracies of high quality sites by users with different quality assessment accuracies (Q=0.5, 0.55, 0.6) and with low (a) and high (b) confidence levels (C) of other's recommendation, in environments with High (20%), Medium (50%) and Low (80%) ratio of low quality sites. Increasing Q from 0.5 to 0.55 leads to sharp increase in performance, especially in the Top1 environment and when C is high.

quality environment. Information cascades occur in all conditions, except when Q=0.5, C=0.5 in the Top10 environment, in which apparently there is no clear direction given by each user. In contrast, in the Top1 environment (similar results obtained for different values of C), information cascades do occur through the selection of the most recommended sites. When Q=0.5, because quality assessments are at chance level, information cascades occur whenever successive users choose and assign a positive recommendation on a random site and made the site the most recommended, which was enough to drive other users to keep selecting it. Higher Q value in Top1 environment shows even stronger information cascade, as consistent positive recommendations quickly lock on to a high quality site.

In the Top10 environment, when C=0.9, information cascade occur through the propagation of local "word-of-mouth" information. However, when Q=0.5, information cascade occurs much slower, as it requires successive random assessment of high quality of a particular site, but even when that happens, the effect is weak and tends to fade away quickly, as there is no reinforcement by the global signal. However, when Q is high, information cascades make the high quality sites quickly accumulate positive recommendation and increase their chance of being selected from the top-10 list. In other words, the pattern of results suggests that, when quality assessment is accurate, high quality sites tend to stand out quickly from the low-quality ones. Growth in positive recommendation for high quality sites quickly reinforces the flow of quality assessments by individuals. However, cascades themselves do not drive quality assessments to a sustainable level unless individual assessments of quality are low, and even so it requires a relatively rare sequence of random events to induce the cascades.

3.5   Effects of Initial Conditions

The above simulations assumed that all sites received no recommendation initially and users randomly choose among them to recommend to others. This may be different from actual situations as "word-of-mouth" information is often distributed
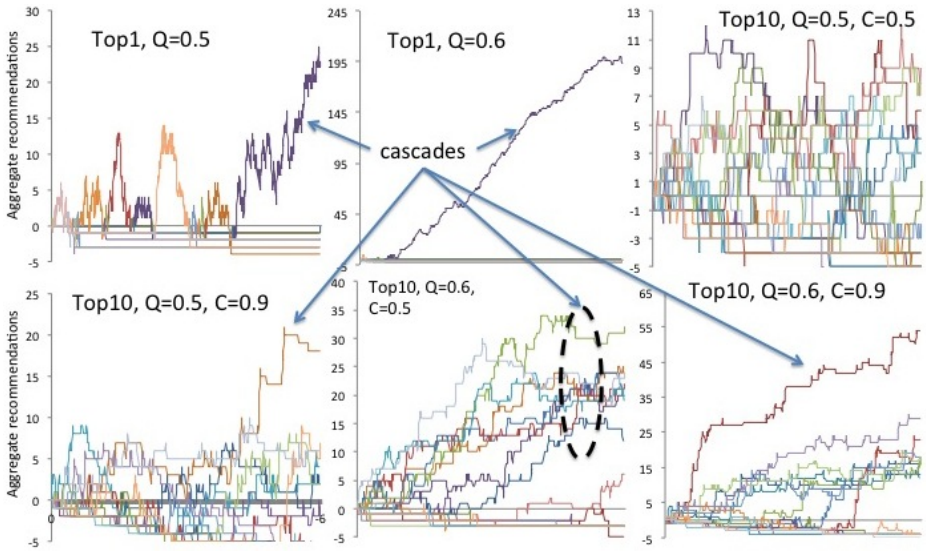
Fig. 3. Recommendations (y-axis) given to the sets of Web sites sequentially by 1000 users (x-axis) in the Medium environment. Color lines represent the net aggregate recommendations of 20 Web sites. Cascades occur when one (or more) Web site dominates others as more user recommendations are added. Except when Q=0.5, cascades tend to favor high-quality sites.

unevenly, or is subject to viral marketing as people are promoting their Web sites. We therefore simulated the effects of having positive recommendation on a subset of low quality sites and see how well these recommendations could hold up across time.

Fig. 4 shows the simulation results in the Medium quality environment with 2 of the low quality sites started off with 20 positive reviews (a higher number simply takes more cycles (users) to reach the same patterns). Consistent with previous simulations, when Q=0.5, the initially positively recommended low quality sites never come down, as users are not able to pass useful quality assessments to other users. In fact, bad cascades tend to drive up these low-quality sites. However, when Q is slightly higher (slightly more than random chance), recommendations of these sites all come down quickly as users collectively give negative recommendations on them. In the Top1 environment, sequential assignments of negative recommendations by multiple users quickly bring down the recommendations for the low quality sites and remove them from the top-1 list. When this happens, positive recommendations accumulate quickly on a high-quality site. In the Top10 environment, even when Q=0.5, when C=0.9, word-of-mouth information may accumulate and lead to cascades, although it is equally likely that the "winning" site has high or low quality. When Q is high, however, recommendations on low-quality sites come down quickly, while high-quality sites accumulate positive recommendations and become more likely to be selected once they appear on the top-10 list. This effect again was magnified by a higher C value, *showing that both local and global information significantly reinforces the collective quality assessments by multiple users.*
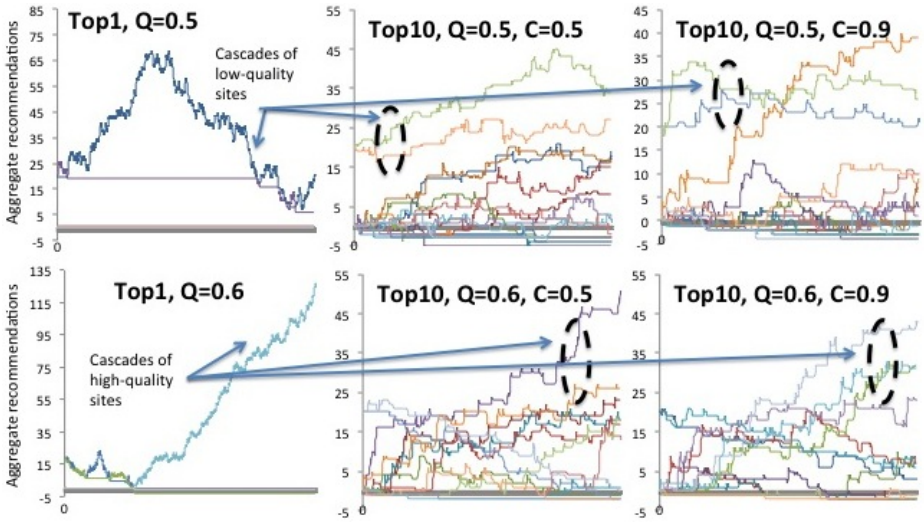
Fig. 4. Recommendations given to the sets of Web sites sequentially by 1000 users in the Medium quality environment. Two of the low-quality sites had 20 positive reviews when the simulation started. When Q=0.5, these low-quality sites tend to stay high, and "bad" cascades often drive them even higher; When Q=0.6, these low-quality sites quickly come down, and "good" cascades tend to magnify aggregate recommendations of high-quality sites.

4 Conclusions and Discussion

We extend previous cascade model to understand how quality assessments of information may be magnified by the effects of cascades at both the local ("word-of-mouth") and global (aggregate recommendation list) levels. We found that in general, even when individual assessments of quality is only slightly better than chance (e.g., p=0.55 or 0.6 that one can correct judge quality), local and global signals can magnify the aggregated quality assessments and lead to good overall quality control of information, such that users can more likely find high quality information. We also show that when users can pass quality assessments to others, *cascades tend to reinforce these assessments, but seldom drive the assessments to either direction*. The results at least partially support the effectiveness of crowdsourcing quality control: even when users are far from perfect quality assessment, so long as the overall quality assessments can flow freely in the social environment, effects of the aggregated quality assessments can be magnified and are useful for improving the choice of high-quality Web sites.

The current model allows users to choose among many options (or choose nothing), which seems to make cascades less likely to occur (compared to a binary choice), but also more unpredictable. The mixing of information cascades of many scales among competing choices results in a turbulent flow to outcomes that cannot be easily predicted [8], but requires a model to predict the multiplicative effects among different signals. Our model shows that cascades are not necessarily bad, they could be effective forces that magnify the aggregated effort of quality assessments

(even though individually they are far from perfect) to facilitate quality control of online information. Designs should therefore encourage more efficient flows of user recommendations to fully harness the potential of crowdsourcing of quality control.

References

1. Bikhchandani, S., Hirshleifer, D., Welch, I.: A theory of fads, fashion, custom, and cultural change in informational cascades. Journal of Political Economy 100(5), 992–1026 (1992)
2. Brossard, D., Lewenstein, B., Bonney, R.: Scientific knowledge and attitude change: The impact of a citizen science project. International Journal of Science Education 27(9), 1099–1121 (2005)
3. Cheung, M., Luo, C., Sia, C., et al.: Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. Int. J. Electron. Commerce 13(4), 9–38 (2009)
4. Czaja, S.J., Charness, N., Fisk, A.D., et al.: Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (create). Psychology & Aging 21(2), 333–352 (2006)
5. Eysenbach, G.: Credibility of health information and digital media: New perspective and implications for youth. In: Metzger, M.J., Flanagin, A.J. (eds.) The John d. And Catherine t. Macarthur Foundation Series on Digital Media and Learning, pp. 123–154. MIT Press, Cambridge (2008)
6. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D.C. (2003)
7. Liao, Q.V., Fu, W.-T.: Effects of cognitive aging on credibility assessments of online health information. In: Proceedings of the 28th Annual ACM Conference on Computer-human Interaction (CHI), Atlanta, GA (2010)
8. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science 311(5762), 854–856 (2006)
9. Sunstein, C.R.: Infotopia: How many minds produce knowledge. Oxford University Press, Oxford (2006)
10. Watts, D.J.: A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences of the United States of America 99(9), 5766–5771 (2002)

Consumer Search, Rationing Rules, and the Consequence for Competition

Christopher S. Ruebeck⋆

Department of Economics, Lafayette College
Easton PA, USA 18042
ruebeckc@lafayette.edu
http://sites.lafayette.edu/ruebeckc/

Abstract. Firms' conjectures about demand are consequential in oligopoly games. Through agent-based modeling of consumers' search for products, we can study the rationing of demand between capacity-constrained firms offering homogeneous products and explore the robustness of analytically solvable models' results. After algorithmically formalizing short-run search behavior rather than assuming a long-run average, this study predicts stronger competition in a two-stage capacity-price game.

Keywords: contingent demand, agent-based modeling, random arrival.

1 Introduction

The residual demand curve, or contingent demand curve, refers to the firm's conjectured demand given its assumptions about consumers and other firms' behaviors. Rather than discussing firms' beliefs about each other's actions, this paper investigates our more basic assumptions about customers' arrivals when firms are capacity-constrained. That is, it focuses on the allocation of customers between firms, the "rationing rule." The literature has shown that duopoly outcomes hinge on these assumptions, and we will see that previous characterizations of aggregate consumer behavior may have unrecognized ramifications.

When Bertrand [3] critiqued the Cournot [5] equilibrium in quantities, he asserted that two firms are enough to create the efficient market outcome: if the strategic variable is price rather than quantity, firms will undercut each other until price equals marginal cost. To instead reinforce the Cournot "quantity" equilibrium as abstracting from a "capacity" decision, Kreps and Scheinkman [7] showed that a two-stage game can support price competition with a capacity choice that is the same as the one-stage Cournot quantities.

This paper's discussion follows Davidson and Deneckere's [6] demonstration that the Kreps and Scheinkman result depends on the assumption that rationing is efficient. They characterized efficient rationing as one extreme, with Beckmann's [2] implementation of Shubik's [11] alternative, proportional rationing, occupying the

other extreme—although we will see behavior below that is outside those proposed extremes. Their analytic and simulation results cover rationing rules in general, and show that only the efficient rationing rule can provide the Kreps and Scheinkman result, arguing in addition that firms will not choose the efficient rationing rule if the two-stage game is extended to three stages that include the choice of rationing rule.

The theory discussed thus far was developed for homogeneous goods, and it has been expanded to include asymmetric costs (Lepore [8]) and differentiated goods (Boccard and Wauthy [4]), but the results below are not about differentiation or variation in costs. Instead, I model search in which the consumer will stop once a low enough price is found at a firm with available capacity. Although this abstracts from the reasons for consumers' need to search, doing so provides a baseline for the future both in terms of algorithms and results.

2   Rationing Rules

With two firms producing homogeneous goods in a simultaneous stage game, we study Firm 2's best response to Firm 1's decision of price and capacity. Both firms may be capacity constrained, but our discussion focuses on Firm 2's unconstrained price choice as it considers Firm 1's constrained decision. Because Firm 1 cannot satisfy the entire market at its price p_1, Firm 2 is able to reach some customers with a higher price $p_2 > p_1$. The rationing rule specifying which customers buy from each firm will be a stochastic outcome in the simulations to follow, but has not been treated that way in theoretical work thus far.

The shape of the efficient rationing rule, shown in Figure 1A, is familiar from exercises 'shifting the demand curve' that begin in Principles of Economics. This rationing rule's implicit assumption about allocating higher-paying customers to the capacity-constrained firm has been known at least since Shubik's discussion of the efficient rationing rule [12] [11]. The demand curve $D(p)$ shifts left by the lower-priced firm's capacity k_1, as depicted by the figure's darker lines. Recognizing that Firm 2 can also be capacity-constrained, residual demand when $p_2 > p_1$ is

$$D(p_2|p_1) = \min(k_2, \max(0, D(p_2) - k_1)) . \tag{1}$$

The rationing of customers between the two firms is indicated by the figure's lightly dashed lines, showing that those units of the good with highest marginal benefit are not available to the higher-priced firm.

Although the efficient rationing rule would appear to be an attractive characterization of a residual demand curve "in the long run," firms are typically making decisions and conjectures about each other's prices in the short run. The long run outcome occurs as a result of firms' shorter-term decisions, their long-run constraints, their anticipations of each other's decisions, and the evolution of the market, including both consumers' and firms' reactions to the shorter-term decisions. Game theoretic characterizations of strategic behavior in the short run are important in determining the long run, as the market's players not only react to each other but also anticipate each other's actions and reactions. Yet,
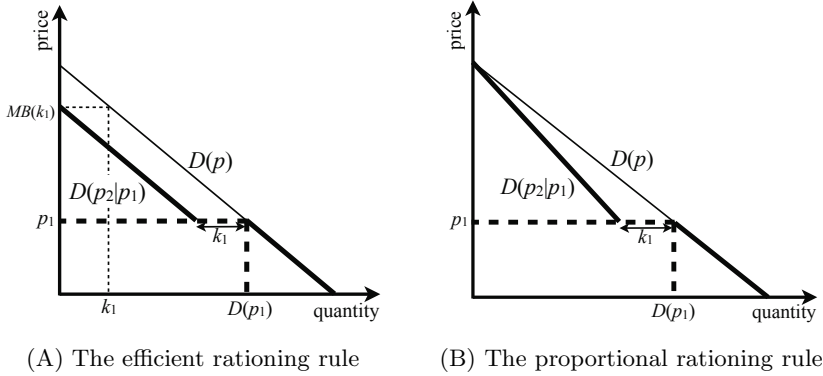
Fig. 1. Standard rationing rules $D(p_2|p_1)$ when Firm 1 chooses capacity k_1 and prices at p_1 in a homogeneous goods market described by demand curve $D(p)$

in assuming the efficient rationing rule, the level of abstraction may penetrate further in some dimensions of the problem than others, leading to an unattractive mismatch between the short run one-shot game and long run outcomes. After discussing proportional rationing's residual demand curve, the discussion and simulations below will show that even the short run abstraction may be mismatched with the descriptive reasoning behind the rationing rule.

Both the proportional residual demand curve depicted in Figure 1B and the efficient demand curve in Figure 1A match market demand $D(p)$ for $p_2 < p_1$: they both take Bertrand's assumption that all customers buy from the lower-priced firm if they can. The simulation results below will not have the same characteristic. In the case of a "tie", $p_2 = p_1$, the literature makes two possible assumptions on the discontinuity. Davidson and Deneckere assume a discontinuity on both the right and the left of (a.k.a. above and below) p_1 with the two firms splitting demand evenly, up to their production capacities. Allen and Hellwig [1] assume a discontinuity only to the left (below) p_1, pricing at equality by extending the upper part of the proportional rationing curve. It may be more accurate to characterize the upper part of the proportional rationing curve as an extension of the outcome when prices are equal, a point to which I will return below.

The key feature of the proportional rationing rule is that all buyers with sufficient willingness to pay have a chance of arriving at both firms. That is, now all consumers with willingness to pay at or above $p_2 > p_1$ may buy from Firm 2. Thus in Figure 1B the market demand curve is "rotated" around its vertical intercept (the price equal to the most any customer will pay) rather than "shifting" in parallel as in Figure 1A. Describing the realization of this contingent demand curve between p_1 and the maximum willingness to pay is not as intuitive as it is for the efficient contingent demand curve, although the highest point is easy: the maximum of of $D(p)$ must also be the maximum of $D(p_2|p_1)$ because both firms have a chance at every customer.

To justify the $D(p_2|p_1)$ proportional demand curve at other prices $p2 > p1$, the literature describes some specifics on consumers' arrival processes and then connects these two points with a straight line, so that "the residual demand at any price is proportional to the overall market demand at that price" (Allen and Hellwig), or as Beckmann first states it, "When selling prices of both duopolists are equal, total demand is a linear function of price. When prices of the two sellers differ, buyers will try as far as possible to buy from the low-price seller. Those who fail to do so will be considered a random sample of all demanders willing to buy at the lower price."

The proportion adopted draws directly from the efficient rationing rule: the fraction of demand remaining for Firm 2 if Firm 1 receives demand for its entire capacity (up to Firm 2's capacity and only greater than zero if market demand is greater than Firm 1's capacity). Formally, the proportional contingent demand for Firm 2 when $p_2 > p_1$ is

$$D(p_2|p_1) = \min \left[k_2, \max \left[0, \frac{D(p_1) - k_1}{D(p_1)} D(p_2) \right] \right]. \tag{2}$$

The key feature of this rationing rule is the constant fraction $\frac{D(p_1)-k1}{D(p_1)}$ that weights the portion of market demand Firm 2 receives at price p_2. First, this fraction is the demand firm 2 receives when prices are equal, and second this fraction of market demand $D(p_2)$ remains constant as Firm 2 increases price. The shape of the residual demand curve below will show that neither of these features need be true as we consider consumers' arrival.

3   Implementing Random Arrival

We now turn to implementing consumer arrival and search rules. For the efficient rationing rule, consumers would be sorted according to the discussion of Figure 1A. They would not arrive in random order: all customers with willingness-to-pay (WTP) greater than the marginal benefit at k_1, $MB(k_1)$, are sent to the capacity constrained, lower-price firm. Somehow these customers rather than others get the good, perhaps by resale (Perry [9]), but it is not the point of this investigation to model such a process. Instead, we will have consumers that arrive in random order—one of the assumptions of the conventional proportional rationing rule, but just part of it. The other part is that the customers that arrive first are always allocated to the lower-priced firm. Relaxing this second part of the assumption is what drives the new results below.

3.1   Random Arrival, Random First Firm

Buyers arrive in random order at two capacity-constrained sellers. As a buyer arrives he chooses randomly between sellers; if the first seller's price is higher than the buyer's WTP, the buyer moves on to the other one, or does not buy at all if both prices are higher than his WTP. As each later buyer arrives, she can only consider sellers that have not yet reached their capacity constraints.

As in the analytic models described above, we take the perspective of Firm 2 given Firm 1's chosen price and capacity. Labeling one firm as "1" does not indicate that buyers in the simulation arrive first at either location; each buyer has an equal chance of arriving first at either Firm 1 or Firm 2.

Figure 2A presents an example of the simulation results. Each of the lines in the plot (some overlapping each other) shows a statistic of 1300 runs; the statistics are min, max, mean, and median quantity for Firm 2 at each price p_2. The plot overlays six 1300-run sets, capturing the stochastic residual demand curve facing Firm 2 with no capacity constraint while the other firm sets price $p_1 = 9$ and capacity $k_1 = 7$. Market demand is $q = 24 - p$. Each of the 23 buyers has unitary demand at price $D_{23} = 23, D_{22} = 22, \ldots, D_1 = 1$. The proportional rationing rule is also shown; It is piecewise-linear, connecting the outcomes for $p = 0, 8, 9$, and 24. (The line segment connecting $p = 8$ and 9 reflects the discrete nature of demand in the simulation.) The efficient rationing rule is not shown, but it would maintain the slope of the lower line from $p_2 = p_1 = 9$ all the way to the vertical axis. The curved line in the plot is the mean residual demand received by Firm 2. As foreshadowed above, it is to the right not only of the efficient rationing rule, but the proportional rationing rule as well. The median is the series of stair steps around that line.
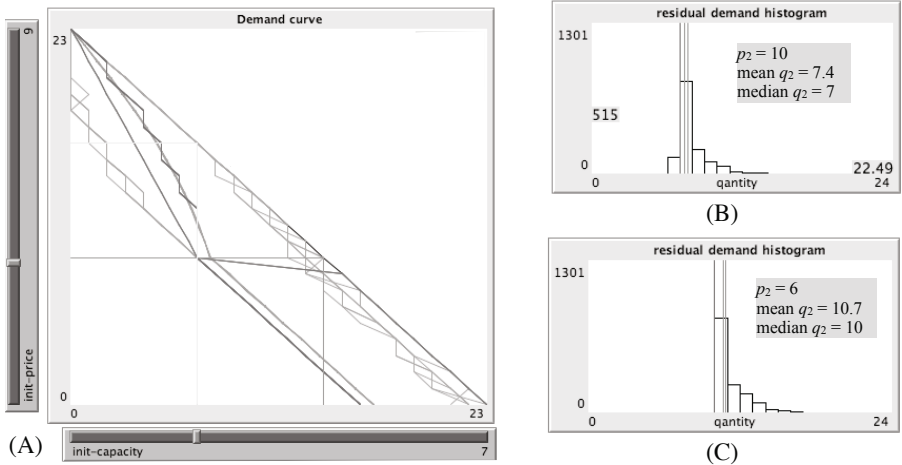
Fig. 2. Simulation results. Firm 2's residual demand curve with no capacity constraint and market demand $q = 23 - p$. Firm 1 has price $p_1 = 9$ and capacity $k_1 = 7$.

Consider first the plot's horizontal extremes. Across the six 1300-run sets, Firm 2's contingent demand has multiple maxima and minima q_2 (horizontal axis) for some p_2 (vertical axis) values. Because the 1300 simulated runs are not always enough to establish the true maxima and minima for all customer arrival outcomes, the distribution's tail is (i) thinner in the minima for higher p_2 in the upper left of the plot and (ii) thinner in the maxima for lower p_2 in the lower

right of the plot. Why? (i) The minimum q_2 at any p_2 is 7 units to the left of the market demand curve (but not less than 0), reflecting the capacity constraint of Firm 1. Thus the minimum contingent demand for Firm 2 is less likely to occur when its price p_2 is high because it is less likely that all high-value buyers arrive first at Firm 2 when there are few possible buyers. (ii) The theoretical maximum at each price p_2 is the entire demand curve: by chance all the customers may arrive first at Firm 2. It is more difficult to reach that maximum in a 1300-run sample at lower prices p_2, when there are more potential customers.

Returning to the central part of the distribution, the mean and median lines, we see that—unlike the assumptions behind Figure 1B—the residual demand curve's distribution is not always centered on the interval between its minimum and maximum. Figure 2B and Figure 2C illustrate the skewed distributions for two of Firm 2's prices, $p_2 = 6$ and 10. With these overall differences in hand, now consider, in the three areas of the graph, the differences in these simulations as compared to the theory pictured in Figure 1.

In the lower part of the curve, the median is actually equal to the minimum for all prices $p_2 < p_1$. The mean is pulled away from the median by the upper tail of the skewed distribution. The explanation for the observed behavior in the extremes applies here. Note, too, in the lower part of the curve that demand is smaller than the usual homogeneous goods assumption: The lower-priced firm does not capture the entire market. This is a consequence of maintaining the random arrival assumption for all prices in the simulation, not just for $p_2 > p_1$ as usually assumed (following Bertrand) in the analytical work described above.

At equal prices $p_2 = p_1$, the outcome is continuous in p_2, a point to consider again below. Finally, we saw that in the upper part of the curve $p_2 > p_1$, the mean residual demand curve no longer changes linearly with price: the two firms share the market with varying proportions rather than the usual constant proportion assumption. Note that the traditional proportional residual demand curve (Davidson and Deneckere) provides Firm 2 with a "better" result than that provided by the efficient residual demand curve because it provides a larger share of the market. The results in Figure 2 are even better for the higher-priced firm! Firm 2's proportion of the market grows as p_2 increases, rather than remaining a constant proportion, because Firm 1's capacity constraint improves the chance that more of the higher-value customers may happen to arrive first at Firm 1 instead of Firm 2.

3.2   Random Arrival, to the Lowest Price If Constraint Unmet

Next consider a specification of behavior that lies midway between the proportional rationing rule and the results in Figure 2. In this case, consumers will continue to search if they do not receive a large enough consumer surplus, $WTP - p$. This specification could be viewed as redundant because we have already specified both that consumers have a willingness-to-pay and that they do not know or have any priors on the firms' prices. It may make more sense to instead specify those priors or updating of priors; I will address this further in the closing section.

In Figure 3, the residual demand curve changes smoothly from the varying-proportion in Figure 2 to the proportional demand curve. When $p_2 > p_1$, the choice of some consumers to continue searching pulls the demand curve back towards the proportional prediction. When $p_2 < p_1$, the choice to continue searching pushes the demand curve out towards the usual homogeneous goods assumption. At $p_2 = p_1$, we have a discontinuity for any search threshold that is not zero (the case of Figure 2).
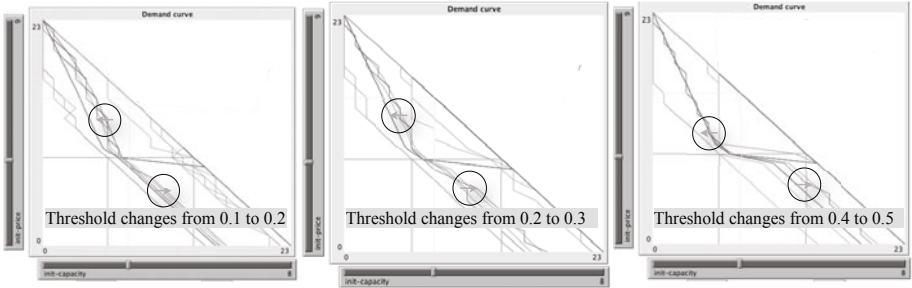
Fig. 3. Continue search when consumer surplus is not a large enough fraction of WTP

4   Conclusions and Further Work

Allen and Hellwig state that consumers go to firms on a "first-come first-served basis" and that, "Firm j with the higher price p_j meets the residual demand of those consumers—if any—who were unable to buy at the lower price." The results above focused directly on a first-come first-served rule, with the result that some consumers with $p_i < $ WTP $ < p_j$ may instead arrive too late to be served by the lower-priced firm and thus not buy. The resulting contingent demand curve provides a varying proportion of market demand at price p_j.

This approach recognizes that the searching consumers may know very little about the prices available before they arrive at the firms, and this assumption is unusual in search models. The limited literature that considers this type of search includes Rothschild [10] and Telser [13], but their analyses still require rather strong assumptions to arrive at a model amenable to analytic methods. Telser's investigation finishes with the conclusion (emphasis added), "If the searcher is ignorant of the distribution, then acceptance of the first choice drawn at random from the distribution *confers a lower average cost* than more sophisticated procedures for a wide range of distributions. In most cases these experiments show that it simply does not pay to discover and patronize the lower price sellers . . . [and] we face the problem of explaining how a seller wishing to stress lower prices than his rivals can attract customers."

Thus, one direction to continue this work is to investigate the effect of this demand curve on equilibrium outcomes, moving from a one-shot static analysis to a dynamic interaction. Davidson and Deneckere show that any contingent

demand curve different from the efficient demand curve must provide more market power (a larger quantity effect) to the higher-priced firm when the other firm is capacity-constrained. Thus firms in the first stage have greater incentive to avoid being capacity-constrained and choose capacities greater than the Cournot quantity, and leading to an equilibrium price support in the second-stage game that is closer to the Bertrand outcome of marginal cost pricing.

The results of these simulations show that the competitive pressure is even stronger when we take seriously the assumptions underlying uninformed consumer serach. Setting a price above one's competitor leads to less loss of demand, so that the market power that remains for the higher-priced firm is even larger than in the case of proportional rationing. At the same time, undercutting is less profitable: the lower-priced unconstrained firm does not receive all demand as assumed generally in the literature on homogeneous products. Thus we can expect that further analysis in a dynamic setting will provide outcomes that are even more competitive than found in existing analytical models.

References

1. Allen, B., Hellwig, M.: Bertrand-Edgeworth duopoly with proportional residual demand. International Economic Review 34(1), 39–60 (1993)
2. Beckmann, M.: Edgeworth-Bertrand duopoly revistied. In: Henn, R. (ed.) Operations Research Verfahren III, Meisenhein. Verlag Anton Hain, Meisenhein (1965)
3. Bertrand, J.: Théorie mathématique de la richesse sociale (review). Journal des Savants 48, 499–508 (1883)
4. Boccard, N., Wauthy, X.: Equilibrium vertical differentiation in a Bertrand model with capacity precommitment. International Journal of Industrial Organization 28(3), 288–297 (2010)
5. Cournot, A.: Researches into the Mathematical Principles of Wealth. Macmillan, London (1838)
6. Davidson, C., Deneckere, R.: Long-run competition in capacity, short-run competition in price, and the Cournot model. The RAND Journal of Economics 17(3), 404–415 (1986)
7. Kreps, D., Scheinkman, J.: Precommitment and bertrand competition yield Cournot outcomes. The Bell Journal of Economics 14(2), 326–337 (1983)
8. Lepore, J.: Consumer rationing and the Cournot outcome. B.E. Journal of Theoretical Economics, Topics 9(1), Article 28 (2009)
9. Perry, M.: Sustainable positive profit multiple-price strategies in contestable markets. Journal of Economic Theory 32(2), 246–265 (1984)
10. Rothschild, M.: Searching for the lowest price when the distribution of prices is unknown. The Journal of Political Economy 82(4), 689–711 (1974)
11. Shubik, M.: A comparison of treatments of a duopoly problem. Econometrica 23(4), 417–431 (1955)
12. Shubik, M.: Strategy and Market Structure: Competition, Oligopoly, and the Theory of Games. John Wiley & Sons, New York (1959)
13. Telser, L.: Searching for the lowest price. American Economic Review 63(2), 40–49 (1973)

Pattern Analysis in Social Networks with Dynamic Connections

Yu Wu[1] and Yu Zhang[2]

[1] Department of Computer Science
Standford University
`ywu2@standford.edu`
[2] Department of Computer Science
Trinity University
`yzhang@trinity.edu`

Abstract. In this paper, we explore how decentralized local interactions of autonomous agents in a network relate to collective behaviors. Most existing work in this area models social network in which agent relations are fixed; instead, we focus on dynamic social networks where agents can rationally adjust their neighborhoods based on their individual interests. We propose a new connection evaluation rule called the Highest Weighted Reward (HWR) rule, with which agents dynamically choose their neighbors in order to maximize their own utilities based on the rewards from previous interactions. Our experiments show that in the 2-action pure coordination game, our system will stabilize to a clustering state where all relationships in the network are rewarded with the optimal payoff. Our experiments also reveal additional interesting patterns in the network.

Keywords: Social Network, Dynamic Network, Pattern Analysis.

1 Introduction

Along with the advancement of modern technology, computer simulation is becoming a more and more important tool in today's researches. The simulation of large scale experiments which originally may take people months or even years can now be run within minutes. Computer simulation has not only saved researchers a great amount of time but also enabled them to study many macro topics that were impossible to study experimentally in the past. In this situation, Multi-Agent System (MAS) has emerged as a new area that facilitates the study of large scale social networks.

Researchers have invented many classic networks, such as random networks, scale-free networks (Albert and Barbási 2002), small world networks (Watts 1999), to list a few. Although these networks have successfully modeled many social structures, they all share the weakness of being static. In today's world, many important virtual networks, such as e-commerce networks, social network services, etc., have much less stable connections among agents and thus the network structures will constantly change. Undoubtedly, the classic networks will fail to capture the dynamics of these networks. Out of this concern, (Zhang and Leezer 2009) develops the network model HCR

(Highest Cumulative Reward) that enables agents to update their connections through a rational strategy. In this way, any classic network can be easily transformed into a dynamic network through the HCR rule. In their model, agents evaluate their neighbors based on their performance history, and all interactions in history are equally weighted. However, people may argue that the evaluation function is not very realistic since in the real world recent events may have a greater influence on people than long past ones.

Motivated by this fact, we extend HCR to a new rule called HWR (Highest Weighted Reward). The difference is that HWR allows the agents to use a discount factor to devalue past interactions with their neighbors. The more recent the interaction is, the more heavily the corresponding reward is weighted in the evaluation function. In our research, we identified certain patterns from the simulation results and then demonstrated the existence of the pattern empirically.

2   Related Work

To our view, the social simulation conducted on MAS can go into two directions: One is to build a model as realistic as possible for a specific scenario, such as the moving crowds in an airport. The other one is to build an abstract model that captures the essence of human activities and can be used to model many different social networks of the same kind. Our work here belongs to the latter category.

Research in this area usually focuses on the study of social norms. One of the most basic questions that can be asked about the social norms is whether the agents in the social network will eventually agree on one solution. Will the whole network converge on one solution? The results vary drastically in different networks and under different algorithms. The agents can converge into various patterns and with different time period. Generally, there are two categories in this area studying the emergence of social norms. The first category studies social norms in static networks. The second studies the evolving or dynamic network.

In the first category of static network, one of the most significant findings was by (Shoham and Tennenholtz 1997). They proposed the HCR (Highest Current Reward) rule. In each timestep, an agent adopting HCR rule switches to a new action if and only if the total rewards obtained from that action in the last time step are greater than the rewards obtained from the currently chosen action in the same time period. The authors simulated various networks under different parameter configurations (memory update frequency, memory restart frequency, and both) to experiment on an agent's effectiveness in learning about their environment. They also theoretically proved that under HCR rule the network will always converge to one social norm.

Another important work done in the first category is the Generalized Simple Majority (GSM) rule proposed by (Delgado 2002). Agents obeying the GSM rule will change to an alternative strategy if they have observed more instances of it on other agents than their present action. This rule generalizes the simple majority rule. As the randomness $\beta \to \infty$, the change of state will occur as soon as more than half of the agents are playing a different action. In that case, GSM will behave exactly as the simple majority rule does.

The second category of research on agents' social behavior is the study of evolutionary networks or dynamic networks, such as (Borenstein 2003, Zimmermann 2005). Here researchers investigate the possible ways in which an agent population

may converge onto a particular strategy. The strong assumptions usually make the experiments unrealistic. For example agents often have no control over their neighborhood. Also, agents do not employ rational selfish reward maximizing strategies but instead often imitate their neighbors. Lastly, agents are often able to see the actions and rewards of their neighbors, which is unrealistic in many social settings.

In the second category of the research, (Zhang and Leezer 2009) proposed the Highest Rewarding Neighborhood (HRN) rule. The HRN rule allows agents to learn from the environment and compete in networks. Unlike the agents in (Borenstein 2003), which can observe their neighbors' actions and imitate it, the HRN agents employ selfish reward maximizing decision making strategy and are able learn from the environment. Under the HRN rule, cooperative behavior emerges even though agents are selfish and attempt only to maximize their own utility. This arises because agents are able to break unrewarding relationships and therefore are able to maintain mutually beneficial relationships. This leads to a Pareto-optimum social convention, where all connections in the network are rewarding.

This paper proposes the HWR rule, which is extended from the HRN rule. In HRN rule, the agent values its neighbors based on the all rewards collected from them in the past, and all interactions are weighted equally throughout the agent's history. However, this will cause the agent to focus too much on the history of the neighbor and fail to respond promptly to the neighbor's latest action. Therefore, we introduce the HWR rule, which introduces a time discount factor that helps the agents to focus on the recent history.

3   Highest Weighted Reward (HWR) Neighbor Evaluation Rule

3.1   HWR Rule

The HWR rule is based on a very simple phenomenon we experience in our everyday lives: human beings tend to value recent events more than events which happened a long time ago. To capture this feature, we introduce a time discount factor, which is usually smaller than 1, in order to linearly devalue the rewards collected from past interactions. When time discount factor equals 1, the HWR rule will behave in the same way as the HRN rule does. Notice that both HWR and HRN are neighbor evaluation rules, which means the purpose of these rules is to help the agents value their connections with neighbors more wisely. In other words, these rules can only make decisions regarding connection choosing, but will not influence the agent's action choosing decisions.

According to the HWR rule, an agent will maintain a relationship if and only if the weighted average reward earned from that relationship is no less than a specified percentage of the weighted average reward earned from every relationship. Next we carefully go through one cycle of HWR evaluation step-by-step.

1. For each neighbor, we have a variable named *TotalReward* to store the weighted total reward from that neighbor. In each turn, we update the *TotalReward* through the following equation in which *c* represents the discount factor:

$$TotalReward = TotalReward \times C + RewardInThisTurn \qquad (1)$$

2. In a similar way, we keep a variable named *GeneralTotalReward* to store the weighted total reward the agent got from all neighbors in the history. In each turn, we update the *GeneralTotalReward* through the following equation:

$$GeneralTotalReward = GeneralTotalReward \times C +$$
$$TotalRewardInThisTurn / NumOfNeighbors \quad (2)$$

Here *TotalRewardInThisTurn* needs to be divided by the *NumOfNeighbors* in each turn in order to find the average reward per connection. Since we want to calculate the average reward of all interactions, the total reward in each round needs to be divided by the number of interactions in that turn.

3. When we start to choose non-rewarding neighbors, we first calculate the average reward for every agent. The *AvgReward* is calculated in the following way: suppose the agent has been playing with the neighbor for n turns, then:

$$AvgReward = \frac{TotalReward}{1 + c + c^2 + ... + c^{n-1}} = \frac{TotalReward}{\dfrac{1 - c^{n-1}}{1 - c}} \quad (3)$$

4. Calculate the average total reward. Similar to step 3:

$$AvgTotalReward = \frac{GeneralTotalReward}{1 + c + c^2 + ... + c^{n-1}} = \frac{GeneralTotalReward}{\dfrac{1 - c^{n-1}}{1 - c}} \quad (4)$$

5. Compare the ratio of *AvgReward/AvgTotalReward* with the *threshold*. If the former is greater than the later, then we keep the neighbor; if not, we regard that agent as a bad neighbor. In this way, the agent can evaluate its neighbor with an emphasis on recent history.

The above explanation has assumed an ideal environment in which agents have infinity amount of memory for ease of exposition. In reality and our experiments, agents have a limited amount of memory and can only remember the interactions of a certain number of turns.

3.2   Discount Factor Range

Generally speaking, the HWR rule can work with different games and in various networks. Since this rule just adds a dynamic factor into the network, it does not contradict other aspects of the network. However, in some special cases, a certain experiment setup may make some discount factor value meaningless. Assume we are playing a pure coordination game (Guillermo and Kuperman 2001) with payoff table (C: Cooperate; D: Defect):

Table 1. Two-Person Two-Action Pure Coordination Game

	C	D
C	1	-1
D	-1	1

Also assume the agents only want to keep the neighbors whose average reward is positive for them. Then in such a case, no matter how many rounds an agent has been playing with a neighbor, its most recent reward will dominate the whole reward sequence. To be more specific, if an agent receives a reward of 1 from a neighbor in the last interaction but has constantly received -1 in the previous n trials, the cumulative reward will be $1 - c - c^2 - \ldots - c^n$. If c is too low, the cumulative reward will still be bigger than 0. We can find the range of c through the following derivation:

$$1 - c - c^2 - \ldots - c^n > 0 \ \Rightarrow \ 1 > c + c^2 + \ldots + c^n \Rightarrow 1 > c(1 + c^n)/(1 - c) \tag{5}$$

When $n \to \infty$, $c^n \to 0$, then

$$\Rightarrow 1 > c/(1 - c) \Rightarrow c < 0.5 \tag{6}$$

Therefore, we usually limit c in range of 0.5 to 1.

In fact, for most cases c will take a value much higher than 0.5. Since the past rewards were devalued exponentially, a small c value will cause the past rewards to be practically ignored very soon. In our common experiment setup, we let the agent have a memory size of 30. In this case, if $c=0.5$, then we see the oldest action will be devalued by a factor of $0.5^{29} \approx 1.86E\text{-}9$, which makes the reward totally insignificant. In fact, even for a relatively high c value such as 0.9, the last factor will still be as small as $0.9^{29} \approx 0.047$. Therefore, just in case the past rewards will be devalued too heavily, we usually keep the c in the range of 0.8 to 1.

3.3   Pattern Predictions

Based on the HWR rule, we make some predictions about the outcome of the network pattern.

> **Argument:** *Given a pure coordination game, placing no constraints on the initial choices of action by all agents, and assuming that all agents employ the HWR rule, then the following holds:*
> - *For every $\varepsilon > 0$ there exists a bounded number M such that if the system runs for M iterations then the probability that a stable clustering state will be reached is greater than $1 - \varepsilon$.*
> - *If a stable clustering state is reached then the all agents are guaranteed to receive optimal payoff from all connections.*

In order to further clarify the argument, the following definitions are given:

> **Stable Clustering State:** A network reaches a stable clustering state when all the agents in the network belong to one and only one closed cluster.
> **Closed Cluster:** A set of agents forms a closed cluster when there does not exist any connection between any agent in the set and another agent outside of the set, and also all the agents in the set are playing the same action.

Basically, what the argument claims is that the whole network will eventually converge into one single cluster of same action or two clusters with different actions, and once in such a state, the rewards collected from the network will be maximized.

4 Experiment

The operating procedure can be shown in the flow chart.

Agents play with each other	Update actions based on the current neighborhood	Evaluate each connection based on the rewards just received	Break up with bad neighbors and update neighborhood
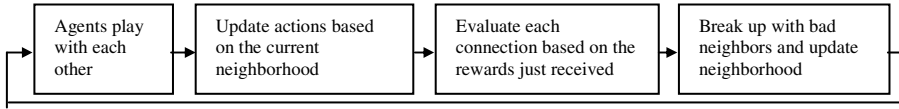
All experiments are conducted in an environment that has the following properties:

1. All trials are run in a random network having 1000 to 10000 agents.
2. Agents play with each other the pure coordination game defined in Table 1.
3. The number of connections in the network remains the same throughout the trial.
4. Every time a connection is broken, both agents have a 50% chance to gain the right to connect to a new neighbor. Exactly one of the two agents will find a new neighbor. This restriction guarantees the number of connections remains the same.
5. All agents have a limited memory size.
6. All agents adopt obey the simple majority rule, which means the agent will always choose the action that the majority of its neighbors used in the last turn. If there is the same number of neighbors adopting different actions, the agent will not change its current action.

4.1 Two-Cluster Stable Clustering State

First, we show the emergence of the stable clustering state. Since the final outcome can have two different patterns with either one or two final stable clusters, we present the patterns in two parts. First, let's check the situation where the network eventually converges into one single cluster. The following experiments are conducted with the parameters: *Number of Runs: 50, Time Steps: 30, Network Size: 5000, Neighborhood Size: 10, Reward Discount: 0.9, Threshold: 0.9.*
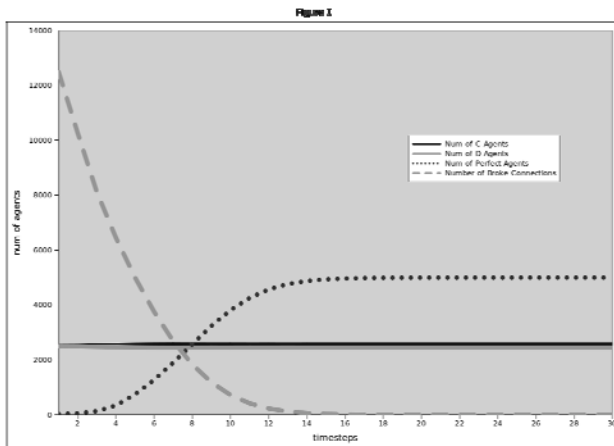
Fig. 1. Network Converges into Two Clusters

Fig. 1 shows the average results collected from 50 trials. We see that at the end of each trial, all agents have split into two camps and entered into a unique stable clustering state. First, we notice that the lines that represent the numbers of C and D agents almost merge together in the middle of the plot. Since the pay-off matrix for C and D actions is symmetric and there is no difference between these two actions except different symbolic names, the average numbers of C and D agents should be both around 2500, just as the plot shows. Here in order to capture the dynamics of the network, we introduce a new term named perfect agent. An agent will be counted as a perfect agent only when all of his neighbors plays the same action as he does. In Figure 1, while the numbers of C and D agents remain almost constant, the number of perfect agents rises steadily from 0 to 5000 throughout the trial. The increasing number of perfect agents shows the network is evolving from a chaotic state towards a stable clustering state. At the same time, we see the number of broken connections is steadily dropping towards 0. As there are more and more agents becoming perfect agents, connections between neighbors are broken less and less frequently.

4.2 One-Cluster Stable Clustering State

Now we present the experiment results where the agents all adopt the same action at last and merge into one cluster. In order to show the results more clearly, we have run 50 trials and selected 23 trials where C eventually becomes the dominant action. The experiments are carried out with the following parameters: *Number of Runs: 50, Time Steps: 30, Network Size: 5000, Neighborhood Size: 100, Reward Discount: 0.9, Threshold: 0.9.* In Fig. 2, we see number of C agents is steadily rising while number of D agents dropping until the whole network converges into a single cluster. The number of perfect agents and broken connections behaves in a manner similar to that in Fig. 1.
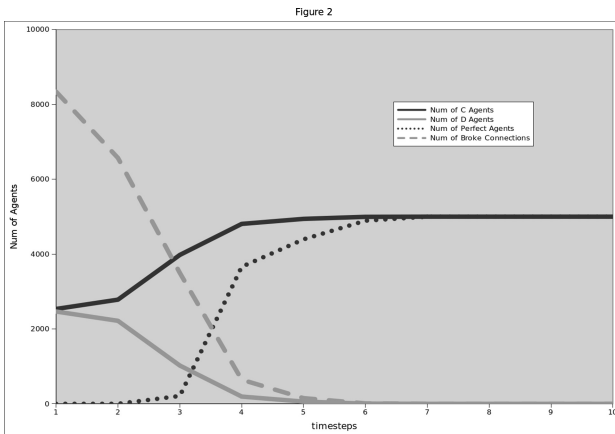
Fig. 2. Network Converges into One Single Cluster

4.3 Pattern Possibilities with Various Network Sizes

The above experiments demonstrate that the network sometimes converges into one single cluster and sometimes into two clusters. After a series of simulations, we found

out that the size of neighborhood has a significant influence over the probability that a network will converge into one cluster or not. In order to show this, we created 16 networks with different sizes and ran 50 trials with each network. Here are the parameters: *Number of Runs: 50, Time Steps: 30, Network Size: 1000, Reward Discount: 0.9,* Threshold: 0.9. Fig. 3 shows the number of one-cluster trials and two-cluster trials with each network.
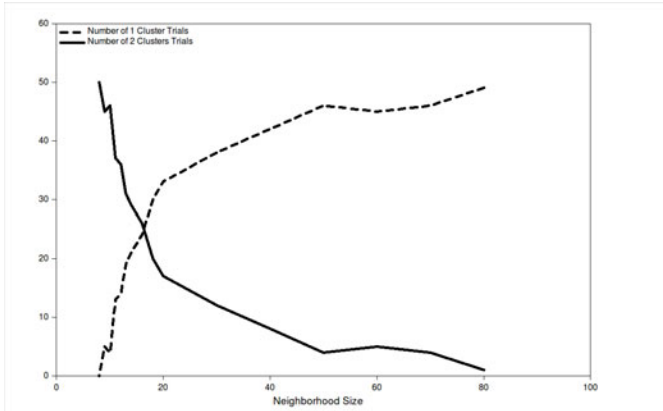
Fig. 3. Network Converges into One Single Cluster

5 Conclusion

The HWR rule enables agents to increase their rewards not only through switching actions but also through updating neighborhoods. As the experiment results show, agents adopting the HWR rule are able to change the network structure to maximize their own interests. Even though the agents behave selfishly, the network still reaches a Pareto-optimum social convention at last.

Acknowledgement

References

1. Abramson, G., Kuperman, M.: Social Games in a Social Network. Physical Review (2001)
2. Albert, R., Barbási, A.L.: Statistical Mechanics of Complex Networks. Modern Physics, 47–97 (2002)
3. Shoham, Y., Tennenholtz, M.: On the Emergence of Social Conventions: Modeling, Analysis and Simulations. Artificial Intelligence, 139–166 (1997)
4. Delgado, J.: Emergence of Social Conventions in Complex Networks. Artificial Intelligence, 171–175 (2002)

5. Borenstein, E., Ruppin, E.: Enhancing Autonomous Agents Evolution with Learning by Imitation. Journal of Artificial Intelligence and Simulation of Behavior 1(4), 335–348 (2003)
6. Zimmermann, M., Eguiluz, V.: Cooperation, Social Networks and the Emergence of Leadership in a Prisoners Dilemma with Adaptive Local Interactions. Physical Review (2005)
7. Watts, D.J.: Small Worlds. Princeton University Press, Princeton (1999)
8. Zhang, Y., Leezer, J.: Emergence of Social Norms in Complex Networks. In: Symposium on Social Computing Applications (SCA 2009), The 2009 IEEE International Conference on Social Computing (SocialCom 2009), Vancouver, Canada, August 29-31, pp. 549–555 (2009)

Formation of Common Investment Networks by Project Establishment between Agents

Jesús Emeterio Navarro-Barrientos

School of Mathematical and Statistical Sciences
Arizona State University, Tempe, AZ 85287-1804, USA

Abstract. We present an investment model integrated with trust and reputation mechanisms where agents interact with each other to establish investment projects. We investigate the establishment of investment projects, the influence of the interaction between agents in the evolution of the distribution of wealth as well as the formation of common investment networks and some of their properties. Simulation results show that the wealth distribution presents a power law in its tail. Also, it is shown that the trust and reputation mechanism proposed leads to the establishment of networks among agents, presenting some of the typical characteristics of real-life networks like a high clustering coefficient and short average path length.

Keywords: complex network, trust-reputation, wealth distribution.

1 Introduction

Recently, different socio-economical problems have been modeled and investigated using agent-based simulations, this approach presents more flexible, realistic and simple conceptual modeling perspectives. Many important contributions to this field are provided by the research group called *agent-based computational economics (ACE)* [9]. Different ACE models have been proposed to study for example the loyalty from buyers to sellers [2], investors and brokers in financial markets [3], and to understand the emergence of networks between agents [9], among others. The topology of the networks emerging from the interaction between agents is usually analyzed using statistical methods [1].

The main goal of this paper is to improve the understanding of two main components in ACE models: (i) the economical component that describes the dynamics of the wealth distribution among agents; and (ii) the social component that describes the dynamics of loyalty, trust and reputation among agents. For this, we integrate a wealth distribution model where agents invest a constant proportion of their wealth, and a network formation model where agents interact with each other to establish investment projects.

2   The Model

2.1   Wealth Dynamics

Consider an *agent-based system* populated with N agents, where each agent posses a *budget* $x_k(t)$ (measure of its "wealth" or "liquidity") that evolves over time given the following dynamic:

$$x_k(t+1) = x_k(t)\Big[1 + r_m(t)\,q_k(t)\Big] + a(t), \tag{1}$$

where $r_m(t)$ denotes the return on investment (RoI) that the agent k receives from its investment $q_k(t)$ in project m, $q_k(t)$ denotes a *proportion of investment*, i.e the fraction or ratio of the budget of agent k that the agent prefers to invest in a market and $a(t)$ denotes an external income, which for simplicity we assume to be constant, $a(t) = 0.5$.

In this model, agent k invests a portion $q_k(t)x_k(t)$ of its total budget at every time step t, yielding a gain or loss in the market m, expressed by $r_m(t)$. Similar wealth models have been presented in [4,5,7] where the dynamics of the investment model are investigated using some results from the theory of multiplicative stochastic processes [8]. Note that this approach assumes that the market, which acts as an *environment* for the agent, is not influenced by its investments, i.e. the returns are exogenous and the influence of the market on the agent is simply treated as random. This is a crucial assumption which makes this approach different from other attempts to model real market dynamics [3].

2.2   Trust-Reputation Mechanisms and Project Establishment

In order to launch a particular investment project m at time t, a certain minimum amount of money I_{thr} needs to be collected among the agents. The existence of the investment threshold I_{thr} is included to enforce the interaction between agents, as they need to *collaborate* until the following condition is reached:

$$I_m(t) = \sum_k^{N_m} q_k(t)\,x_k(t) \geq I_{\text{thr}}, \tag{2}$$

where N_m is the number of agents collaborating in the particular investment project m. For simplicity, it is assumed that each agent participates in only one investment project at a time.

The first essential feature to be noticed for the formation of common investment networks is the establishment of preferences between agents. It is assumed that the decision of an agent to collaborate in a project will mainly depend on the previous history it has gained with other agents. Consider an agent k which accepts to collaborate in the investment project m initiated by agent j. Thus, agent k receives the following payoff at time t:

$$p_{kj}(t) = x_k(t)\,q_k(t)\,r_m(t). \tag{3}$$

Reiterated interactions between agent k and agent j lead to different payoffs over time that are saved in a trust weight:

$$w_{kj}(t+1) = p_{kj}(t) + w_{kj}(t)\, e^{-\gamma}, \tag{4}$$

where γ represents the memory of the agent. The payoffs obtained from previous time steps t may have resulted from the collaborative action of different agents, however, these are unknown to agent k, i.e agent k only keeps track of the experience with the initiator of the project, agent j. Furthermore, in order to mirror reality, it is assumed that from the population of N agents only a small number J are initiators, i.e. $J \ll N$, where the reputation of an initiator j can be calculated as follows (for more on trust and reputation models see [6]):

$$W_j(t+1) = \sum_{k=0}^{N} w_{kj}(t). \tag{5}$$

At every time step t an initiator is chosen randomly from the population and assigned with an investment project. The initiator randomly tries to convince other agents to invest in the project until an amount larger than the threshold I_{thr} has been collected. For this, we use a Boltzmann distribution to determine the probability that the contacted agent k may accept the offer of agent j:

$$\tau_{kj}(t) = \frac{e^{\beta w_{kj}(t)}}{\sum_{i=1}^{J} e^{\beta w_{ki}(t)}}, \tag{6}$$

where in terms of the trust weight w_{kj}, the probability $\tau_{kj}(t)$ considers the good or bad previous experience with agent j with respect to the experience obtained with other initiators; and β denotes the greediness of the agent, i.e. how much importance does the agent give to the trust weight w_{kj}. In order to make a decision, agent k uses a technique analogous to a roulette wheel where each slice is proportional in size to the probability value $\tau_{kj}(t)$. Thus, agent k draws a random number in the interval $(0, 1)$ and accepts to invest in the project of agent j if the segment of agent j in the roulette spans the random number. Finally, an initiator j stops to contact other agents if either the investment project has reached the threshold I_{thr} or if all agents in the population have been asked for collaboration. Now, if the project could be launched it has to be evaluated. The evaluation should in general involve certain "economic" criteria that also reflects the nature of the project. However, for simplicity we assume that the failure or success of an investment project m is randomly drawn from a uniform distribution, i.e. $r_m(t) \sim U(-1, 1)$.

3   Results of Computer Simulations

We performed some simulations using the parameter values in Table 1, for simplicity, we assume that the initial budget $x_k(0)$ is the same for all agents and the

Table 1. Parameter values of the computer experiments for the model of formation of common investment networks

Parameter	Value	Parameter	Value
Num. of agents	$N = 10^4$	Initial budget	$x_k(0) = 1$
Num. of initiators	$J = 100$	Initial trust	$w_{kj}(0) = 0$
Num. of time steps	$t = 10^5$	Memory	$\gamma_k = 0.1$
Investment threshold	$I_{\text{thr}} = 9$	Greediness	$\beta_k = 1$

proportion of investment is assumed to be constant and the same for all agents, i.e. $q_k(t) = q = const$.

Fig. 1 (left) shows the evolution of the budget distribution over time for $q = 0.5$. Note that the probability distribution of the budget converges to a stationary distribution with a power law in the tail, a known property of investment models based on multiplicative processes repelled from zero [8,5]. Fig. 1 (right) shows the distribution of the budget at time step $t = 10^5$ for different constant proportions of investment q. Note that even for a large number of time steps, the budget distribution for agents with a proportion of investment of $q = 0.1$ has not yet converged to a stationary distribution, whereas for $q = 0.5$ and $q = 0.9$, the distribution has reached a stationary state.

We examine the evolution over time of the budget and reputation of the initiators to understand their role in the dynamics of the investment model. For the sake of clarity, the rank-size distribution of the budget of the initiators is shown in Fig. 2 (left), note that richer initiators would grow ever richer. It is also interesting to examine the evolution of the budget of the initiator with the largest and the smallest budget at the end of the simulation. This is shown in the inset of Fig. 2 (left), note that the budget of the best agent was not always increasing over time, which shows the relevance of stochastic returns in the dynamics. Fig. 2 (right) shows the rank-size distribution of the reputation of the initiators, Eq. (5), note that the distribution does not change over time, meaning that only for a small number of agents there is a significant increase or decrease on reputation over time. Moreover, it can be shown that the average value of the reputation has a shift to larger positive values over time. This occurs due to an aggregation of external incomes $a(t)$, Eq. (1), into the dynamics of the trust weights, Eq. (4). Moreover, the inset in Fig. 2 (right) shows the reputation of the best and the worst initiator indicating the presence of no symmetrical positive/negative reputation values.

The influence of other parameters in the dynamics of the model was also analyzed, however, for brevity, we discuss only the role of the number of initiators J in the dynamics. It can be shown that, if the number of initiators is small, then more investors will be willing to invest in their projects. This leads to a larger amount of investment that can be collected by the initiators. It was shown in Fig. 1 that the tail of the wealth distribution has a power law distribution, it can be shown that the larger the number of initiators J, the larger the slope

of the power law. The reason for this is that a small number of initiators collect more money from the investors leading to larger profits and looses which over time lead to wider distributions than for a large number of initiators.
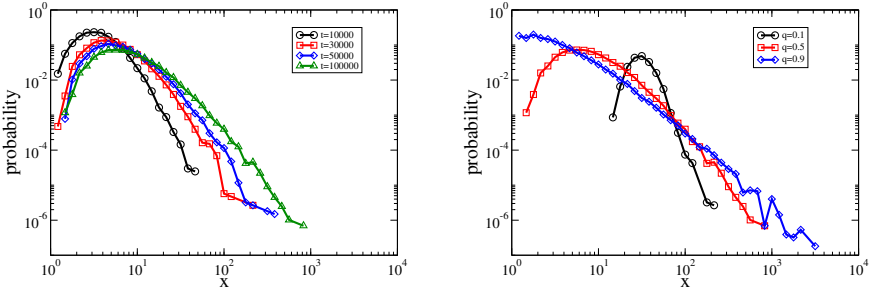
Fig. 1. Probability distribution of the budget: (left) evolution over time for a constant proportion of investment $q = 0.5$; (right) for different proportions of investment q after $t = 10^5$ time steps. Additional parameters as in Table 1.
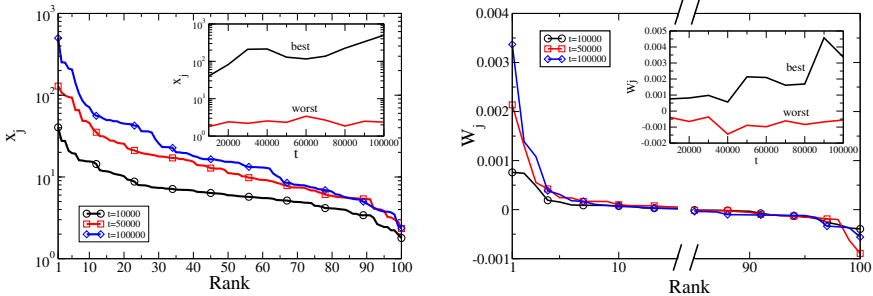
Fig. 2. For initiators, evolution over time of the rank-size distribution of (left) budget and (right) reputation. Insets show the best and the worst budget and reputation, respectively. Additional parameters as in Table 1.

4   Structure of Common Investment Networks

In this section, we analyze the topology of the networks. For simplicity and brevity, we focus our analysis on the influence of different constant proportions of investment in the emerging structure of the common investment networks. For this, we run different computer experiments for a small population of agents $N = 10^3$ (N is also the number of nodes in the network) and other parameter values as in Table 1. The first experiment investigates the influence of the proportion of investment in the properties of the network. Fig. 3 shows the networks emerging from the common investment model for different proportion of investment q at time step $t = 10^3$. Note that these networks have two types of nodes, the black nodes represent investors and the gray nodes represent initiators. Based on visual

impression, the density of the network decreases with respect to the proportion of investment. This occurs because agents investing more also tend to loose more, which leads to more mistrust. However, from the visual representation of the network it is not possible to draw many conclusions from the dynamics of the networks. Thus, we need to rely on other methods to characterize the networks, typically the following statistical properties of the networks are used: number of links V, maximal degree $k_{\max}$ (the degree of the highest degree vertex of the network), average path length l (the average shortest distance between any pair of nodes in the network), and clustering coefficient C (measures the transitivity of the network).
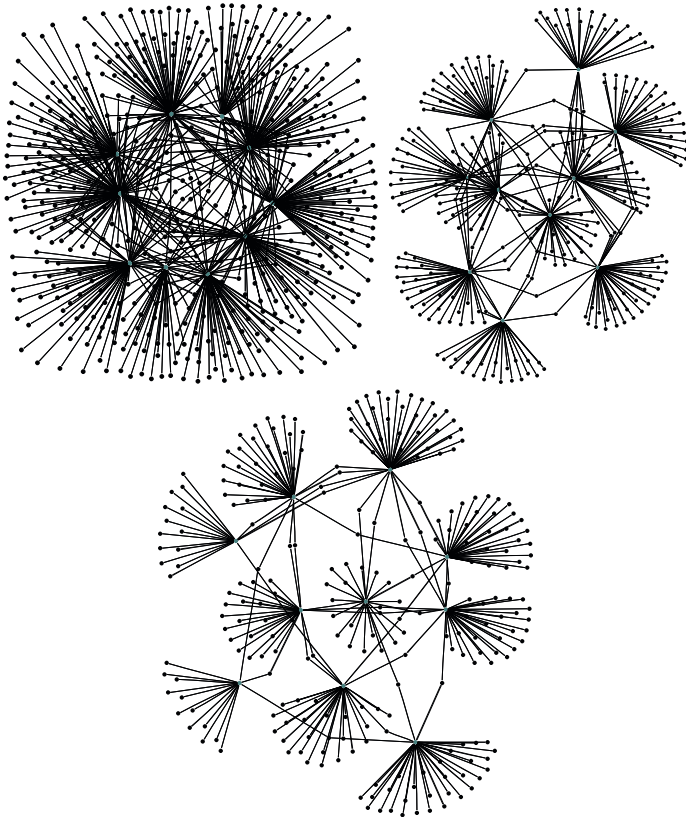
Fig. 3. Common investment networks for different proportions of investment at time step $t = 10^3$: (left) $q = 0.1$, (right) $q = 0.5$ and (bottom) $q = 0.9$. A link between agents represents a positive decision weight, i.e $w_{kj} > 0$. For $N = 10^3$ investors (black nodes) and $J = 10$ initiators (gray nodes). Additional parameters as in Table 1.

On the following we present the most important properties of the networks emerging from our model. First, it can be shown that both the number of links

V as well as the maximal degree $k_{\max}$ of the network increase over time proportionally to a power-law function, where the slope of the power law decreases if the proportion of investment increases. This basically says that the density of the networks increase over time and can be characterized. Moreover, it can be shown that the clustering coefficient C is large for small proportions of investment. This occurs because a small proportion of investment leads to a higher clustering in the network due to the mistrust that large losses generate in the investors. Table 2 shows some of the most important characteristics for different number of investors N and initiators J for a large number of time steps, i.e. $t = 10^5$. For each network we indicate the average degree $\langle k \rangle$ (the first moment of the degree distribution), the average path length l and the clustering coefficient C. For comparison reasons we include the average path length $l_{\mathrm{rand}} = \log(N)/\log(\langle k \rangle)$ and the clustering coefficient $C_{\mathrm{rand}} = \langle k \rangle / N$ that can be obtained from a random network with the same average degree $\langle k \rangle$ of the investment networks. Note that the average degree $\langle k \rangle$ increases with respect to the system size. In general, the networks show a small average path length $l \approx 2$, meaning that any investor or initiator in the network is on average connected to each other by two links. The clustering coefficient of the investment networks is larger than the clustering coefficient of a random network, this indicates the presence of transitivity in our networks. This occurs mainly because of the large number of investors connected to initiators. Finally, note that the clustering coefficient of the networks decreases with respect to N, this is in qualitative agreement with properties of small-world networks [1].

Table 2. Properties of the networks for different number of investors N and initiators J. The properties measured are: V the number of links, $k_{\max}$ the maximal degree, $\langle k \rangle$ the average degree, l the average path length and C the clustering coefficient. For proportion of investment $q = 0.5$, $t = 10^5$ and additional parameters as in Table 1.

N	J	V	$k_{\max}$	$\langle k \rangle$	l	C	l_{rand}	C_{rand}
1000	10	4847	517	0.9694	2.05766	0.74557	-	0.0009694
2000	20	19972	1050	3.9944	1.99365	0.71337	5.488	0.0019972
3000	30	41073	1475	8.2146	1.99314	0.71130	3.8018	0.0027382
10000	100	134279	1477	26.86	2.15630	0.24136	2.7989	0.0026860

5   Conclusions and Further Work

This article presents the initial investigations on a model for formation of common investment networks between agents. The model integrates two models, an investment model where agents place investments proportional to their budget, and a trust and reputation model where agents keep track of their experience with initiators of projects, establishing a common investment network. The most important conclusions that can be drawn from the model here presented are the following: (i) the budget of the agents reaches a stationary distribution after some time steps and presents a power law distribution on the tail, a property

discussed in other investment models [5,7,8]; and (ii) the topology of the investment networks emerging from the model present some of the typical characteristics of real-life networks like a high clustering coefficient and short average path length.

We focused our investigations on the feedback that describes the establishment and reinforcement of relations among investors and initiators. This is considered a "social component" of the interaction between agents and it was shown how this feedback process based on positive or negative experience may lead to the establishment of networks among agents. Furthermore, we noted that the external income sources play an important role on the dynamics, leading to non-symmetrical distributions of trust and reputation among agents. We note also that further experiments are needed for different memory γ and greediness β values to understand their influence in the dynamics of the networks.

Acknowledgements

References

1. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97 (2002)
2. Kirman, A., Vriend, N.J.: Evolving market structure: An ACE model of price dispersion and loyalty. J. Econ. Dyn. Control 25, 459–502 (2001)
3. LeBaron, B.: Agent-based computational finance: Suggested readings and early research. J. Econ. Dyn. Control 24, 679–702 (2000)
4. Navarro, J.E., Schweitzer, F.: The investors game: A model for coalition formation. In: Czaja, L. (ed.) Proceedings of the Workshop on Concurrency, Specification & Programming, (CS & P 2003), vol. 2, pp. 369–381. Warsaw University, Czarna (2003)
5. Navarro-Barrientos, J.E., Cantero-Alvarez, R., Rodrigues, J.F.M., Schweitzer, F.: Investments in random environments. Physica A 387(8-9), 2035–2046 (2008)
6. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artificial Intelligence Review 24(1), 33–60 (2005)
7. Solomon, S., Richmond, P.: Power laws of wealth, market order volumes and market returns. Physica A 299(1-2), 188–197 (2001)
8. Sornette, D., Cont, R.: Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. Journal of Physics 1(7), 431–444 (1997)
9. Tesfatsion, L., Judd, K. (eds.): Handbook of Computational Economics, vol. 2. Elsevier, Amsterdam (2006)

Constructing Social Networks from Unstructured Group Dialog in Virtual Worlds

Fahad Shah and Gita Sukthankar

Department of EECS
University of Central Florida
4000 Central Florida Blvd, Orlando, FL
sfahad@cs.ucf.edu, gitars@eecs.ucf.edu

Abstract. Virtual worlds and massively multi-player online games are rich sources of information about large-scale teams and groups, offering the tantalizing possibility of harvesting data about group formation, social networks, and network evolution. However these environments lack many of the cues that facilitate natural language processing in other conversational settings and different types of social media. Public chat data often features players who speak simultaneously, use jargon and emoticons, and only erratically adhere to conversational norms. In this paper, we present techniques for inferring the existence of social links from unstructured conversational data collected from groups of participants in the Second Life virtual world. We present an algorithm for addressing this problem, Shallow Semantic Temporal Overlap (SSTO), that combines temporal and language information to create directional links between participants, and a second approach that relies on temporal overlap alone to create undirected links between participants. Relying on temporal overlap is noisy, resulting in a low precision and networks with many extraneous links. In this paper, we demonstrate that we can ameliorate this problem by using network modularity optimization to perform community detection in the noisy networks and severing cross-community links. Although using the content of the communications still results in the best performance, community detection is effective as a noise reduction technique for eliminating the extra links created by temporal overlap alone.

Keywords: Network text analysis, Network modularity, Semantic dialog analysis.

1 Introduction

Massively multi-player online games and virtual environments provide new outlets for human social interaction that are significantly different from both face-to-face interactions and non-physically-embodied social networking tools such as Facebook and Twitter. We aim to study group dynamics in these virtual worlds by collecting and analyzing public conversational patterns of Second Life users.

Second Life (SL) is a massively multi-player online environment that allows users to construct and inhabit their own 3D world. In Second Life, users control avatars, through which they are able to explore different environments and interact with other avatars in a variety of ways. One of the most commonly used methods of interaction in Second

Life is basic text chat. Users are able to chat with other users directly through private instant messages (IMs) or to broadcast chat messages to all avatars within a given radius of their avatar using a public chat channel. Second Life is a unique test bed for research studies, allowing scientists to study a broad range of human behaviors. Several studies on user interaction in virtual environments have been conducted in SL including studies on conversation [1] and virtual agents [2].

The physical environment in Second Life is laid out in a 2D arrangement, known as the SLGrid. The SLGrid is comprised of many regions, with each region hosted on its own server and offering a fully featured 3D environment shaped by the user population. The current number of SL users is estimated to be 16 million, with a weekly user login activity reported in the vicinity of 0.5 million [3].

Although Second Live provides us with rich opportunities to observe the public behavior of large groups of users, it is difficult to interpret who the users are communicating to and what they are trying to say from public chat data. Network text analysis systems such as Automap [4] that incorporate linguistic analysis techniques such as stemming, named-entity recognition, and n-gram identification are not effective on this data since many of the linguistic preprocessing steps are defeated by the slang and rapid topic shifts of the Second Life users. This is a hard problem even for human observers, and it was impossible for us to unambiguously identify the target for many of the utterances in our dataset. In this paper, we present an algorithm for addressing this problem, Shallow Semantic Temporal Overlap (SSTO), that combines temporal and language information to infer the existence of directional links between participants. One of the problems is that using temporal overlap as a cue for detecting links can produce extraneous links and low precision. To reduce these extraneous links, we propose the use of community detection. Optimizing network modularity reduces the number of extraneous links generated by overly generous temporal co-occurrence assumption but does not significantly improve the performance of SSTO.

There has been previous work on constructing social networks of MMORPG players, e.g., [5] looks at using concepts from social network analysis and data mining to identify tasks. Recent work [6] has compared the relative utility of different types of features at predicting friendship links in social networks; in this study we only examine conversational data and do not include information about other types of Second Life events (e.g., item exchanges) in our social networks.

2   Method

2.1   Dataset

We obtained conversation data from eight different regions in Second Life over fourteen days of data collection; the reader is referred to [7] for details. To study user dialogs, we examined daily and hourly data for five randomly selected days in the eight regions. In total, the dataset contains 523 hours of information over the five days (80,000 utterances) considered for the analysis across all regions. We did a hand-annotation of one hour of data from each of the regions to serve as a basis for comparison.

While there are corpora like [8], there has not been any body of work with an online chat corpus in a multi-user, open-ended setting — the characteristics of this dataset. In

such situations it is imperative to identify conversational connections before proceeding to higher level analysis like topic modeling, which is itself a challenging problem. We considered several approaches to analyzing our dialog dataset, ranging from statistical NLP approaches using classifiers to corpus-based approaches using tagger/parsers; however we discovered that there is no corpus available for group-based online chat in an open-ended dialog setting. It is challenging even for human labelers to annotate the conversations themselves due to the large size of the dataset and the ambiguity in a multi-user open-ended setting. Furthermore, the variability of the utterances and the nuances such as emoticons, abbreviations and the presence of emphasizers in spellings (e.g., "Yayyy") makes it difficult to train appropriate classifiers. Parser/tagger-based approaches perform poorly due to the lack of corpus and inclusion of non-English vocabulary words.

Consequently, we decided to investigate approaches that utilize non-linguistic cues such as temporal co-occurrence. Although temporal co-occurrence can create a large number of false links, many aspects of the network group structure are preserved. Hence we opted to implement a two-pass approach: 1) create a noisy network based solely on temporal co-occurrence, 2) perform modularity detection on the network to detect communities of users, and 3) attempt to filter extraneous links using the results of the community detection.

2.2   Modularity Optimization

In prior work, community membership has been successfully used to identify latent dimensions in social networks [9] using techniques such as eigenvector-based modularity optimization [10] which allows for both complete and partial memberships. As described in [10], modularity (denoted by Q below) measures the chances of seeing a node in the network versus its occurrence being completely random; it can be defined as the sum of the random chance $A_{ij} - \frac{k_i k_j}{2m}$ summed over all pairs of vertices i, j that fall in the same group, where s_i equals 1 if the two vertices fall in the same group and -1 otherwise:

$$Q = \frac{1}{4m} \sum (A_{ij} - \frac{k_i k_j}{2m}) s_i s_j. \tag{1}$$

If B is defined as the modularity matrix given by $A_{ij} - \frac{k_i k_j}{2m}$, which is a real symmetric matrix and s column vectors whose elements are s_i then Equation 1 can be written as $Q = \frac{1}{4m} \sum_{i=1}^{n} (u_I^T s)^2 \beta_i$, where β_i is the eigenvalue of B corresponding to the eigenvector u (u_i are the normalized eigenvectors of B so that $s = \sum_i^n a_i u_i$ and $a_i = u_i^T s$). We use the leading eigenvector approach to spectral optimization of modularity as described in [11] for the strict community partitioning (s being 1 or -1 and not continuous). For the maximum positive eigenvalue we set $s = 1$ for the corresponding element of the eigenvector if its positive and negative otherwise. Finally we repeatedly partition a group of size n_g in two and calculate the change in modularity measure given by $\Delta q = \frac{1}{4m} \sum_{i,j \epsilon g} [B_{ij} - \delta_{ij} \sum_{k \epsilon g} B_{ik}] s_i s_j$, where δ_{ij} is the Kronecker δ symbol, terminating if the change is not positive and otherwise choosing the sign of s (the partition) in the same way as described earlier.

2.3   Shallow Semantics and Temporal Overlap Algorithm (SSTO)

Because of an inability to use statistical machine learning approaches due to the lack of sufficiently labeled data and absence of a tagger/parser that can interpret chat dialog data, we developed a rule-based algorithm that relies on shallow semantic analysis of linguistic cues that commonly occur in chat data including mentions of named entities as well as the temporal co-occurrence of utterances to generate a to/from labeling for the chat dialogs with directed links between users. Our algorithm employs the following types of rules:

salutations: Salutations are frequent and can be identified using keywords such as "hi", "hello", "hey". The initial speaker is marked as the *from* user and users that respond within a designated temporal window are labeled as *to* users.

questions: Question words (e.g., "who", "what", "how") are treated in the same way as salutations. We apply the same logic to requests for help (which are often marked by words such as "can", "would").

usernames: When a dialog begins or ends with all or part of a username (observed during the analysis period), the username is marked as *to*, and the speaker marked as *from*.

second person pronouns: If the dialog begins with a second person pronoun (i.e., "you", "your"), then the previous speaker is considered as the *from* user and the current speaker the *to* user; explicit mentions of a username override this.

temporal co-occurrences: Our system includes rules for linking users based on temporal co-occurrence of utterances. These rules are triggered by a running conversation of 8–12 utterances.

This straightforward algorithm is able to capture sufficient information from the dialogs and is comparable in performance to SSTO with community information, as discussed below.

2.4   Temporal Overlap Algorithm

The temporal overlap algorithm consists of using the temporal co-occurrence to construct the links. It exploits the default timeout in Second Life (20 minutes) and performs a lookup for 20 minutes beginning from the occurrence of a given username and constructs an undirected link between the speakers and this user. This process is repeated for all users within that time window (one hour or day) in 20 minute periods. This algorithm gives a candidate pool of initial links between the users without considering any semantic information. Later, we show that incorporating community information from any source (similar time overlap or SSTO based) and on any scale (daily or hourly) enables us to effectively prune links, showing the efficacy of mining community membership information.

2.5   Incorporating Community Membership

Our dataset consists of 5 randomly-chosen days of data logs. We separate the daily logs into hourly partitions, based on the belief that an hour is a reasonable duration for social interactions in a virtual world. The hourly partitioned data for each day is used to

generate user graph adjacency matrices using the two algorithms described earlier. The adjacency matrix is then used to generate the spectral partitions for the communities in the graph, which are then used to back annotate the tables containing the to/from labeling (in the case of the SSTO algorithm). These annotations serve as an additional cue capturing community membership. Not all the matrices are decomposable into smaller communities so we treat such graphs of users as a single community.

There are several options for using the community information — we can use the community information on an hourly- or daily basis, using the initial run from either SSTO or the temporal overlap algorithms. The daily data is a long-term view that focuses on the stable network of users while the hourly labeling is a fine-grained view that can enable the study of how the social communities evolve over time. The SSTO algorithm gives us a conservative set of directed links between users while the temporal overlap algorithm provides a more inclusive hypothesis of users connected by undirected links.

For the SSTO algorithm, we consider several variants of using the community information:

SSTO: Raw SSTO without community information;

SSTO+LC: SSTO (with loose community information) relies on community information from the previous run only when we fail to make a link using language cues.

SSTO+SC: SSTO (with strict community information) always uses language cues in conjunction with the community information.

For the temporal overlap algorithms, we use the community information from the previous run.

TO: Raw temporal overlap algorithm without community information;

TO+DT Temporal overlap plus daily community information;

TO+HT Temporal overlap plus hourly community information.

3   Results

In this section we summarize the results from a comparison of the social networks constructed from the different algorithms. While comparing networks for similarity is a difficult problem [12], we restrict our attention to comparing networks as a whole in terms of the link difference (using Frobenius norm) and a one-to-one comparison for the *to* and *from* labelings for each dialog on the ground-truthed subset (using precision and recall).

3.1   Network Comparison Using the Frobenius Norm

We constructed a gold-standard subset of the data by hand-annotating the to/from fields for a randomly-selected hour from each of the Second Life regions. It is to be noted that there were instances where even a human was unable to determine the person addressed due to the complex overlapping nature of the dialogs in group conversation in an open ended setting (Table 2).

To compare the generated networks against this baseline, we use two approaches. First we compute a Frobenius norm [13] for the adjacency matrices from the corresponding networks. The Frobenius norm is the matrix norm of an $M \times N$ matrix A and is defined as:

$$\|A\| = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} |a_{ij}|^2}. \tag{2}$$

The Frobenius norm directly measures whether the two networks have the same links and can be used since the networks consists of the same nodes (users). Thus, the norm serves as a measure of error (a perfect match would result in a norm of 0). Table 1 shows the results from this analysis.

Table 1. Frobenius norm: comparison against hand-annotated subset

	SSTO	SSTO+LC	SSTO+SC	TO	TO+DT	TO+HT
Help Island Public	35.60	41.19	46.22	224.87	162.00	130.08
Help People Island	62.23	60.50	66.34	20.29	20.29	54.88
Mauve	48.45	45.11	51.91	58.44	58.44	49.89
Morris	24.67	18.92	20.76	43.12	37.54	38.98
Kuula	32.12	30.75	32.66	83.22	73.15	77.82
Pondi Beach	20.63	21.77	21.56	75.07	62.62	71.02
Moose Beach	17.08	18.30	21.07	67.05	53.64	50.97
Rezz Me	36.70	39.74	45.78	38.72	39.01	41.10
Total error	277.48	**276.28**	306.30	610.78	507.21	514.74

3.2 Direct Label Comparisons

The second quantitative measure we present is the head-to-head comparison of the to/from labelings for the dialogs using any of the approaches described above (for SSTO) against the hand annotated dialogs. This gives us the true positives and false positives for the approaches and allows us to see which one is performing better on the dataset, and if there is an effect in different Second Life regions. Table 2 shows the results from this analysis.

(a) Hand labeled network   (b) SSTO labeled network   (c) TO labeled network
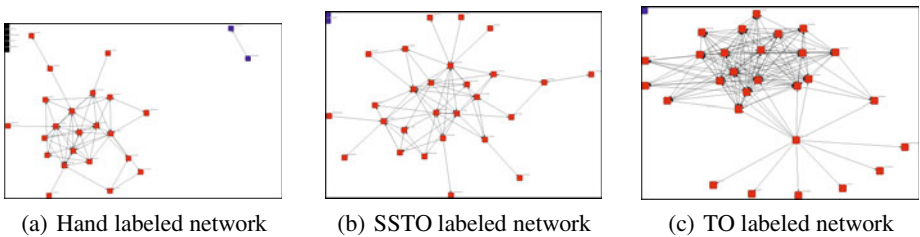
Fig. 1. Networks from different algorithms for one hour in the Help Island Public region

Table 2. Precision/Recall values for one-to-one labeling comparison

		Help Island Public	Help People Island	Mauve	Morris	Kuula	Pondi Beach	Moose Beach	Rezz Me
Total Dialogs		360	184	128	179	227	144	128	97
Hand Labeled	recall	0.6278	0.9076	0.9453	0.6983	0.8370	0.6944	0.6797	0.8866
	total	226	167	121	125	190	100	87	86
SSTO+SC	match	61	59	49	43	63	27	12	23
	precision	0.2607	0.6629	0.6364	0.4216	0.4632	0.3971	0.2105	0.4600
	recall	0.2699	0.3533	0.4050	0.3440	0.3316	0.2700	0.1379	0.2674
	F-Score	0.2652	0.4609	0.4204	0.3789	0.3865	0.3214	0.1667	0.3382
	total	234	89	77	102	136	68	57	50
SSTO+LC	match	61	51	37	39	52	26	12	15
	precision	0.3005	0.6456	0.6607	0.4643	0.4561	0.4194	0.2667	0.4688
	recall	0.2699	0.3054	0.3058	0.3120	0.2737	0.2600	0.1379	0.1744
	F-Score	0.2844	0.4146	0.4181	0.3732	0.3421	0.3210	0.1818	0.2542
	total	203	79	56	84	114	62	45	32
SSTO	match	76	68	51	45	66	30	20	27
	precision	0.3065	0.7083	0.6145	0.4500	0.4748	0.4225	0.3077	0.4576
	recall	0.3363	0.4072	0.4215	0.3600	0.3474	0.3000	0.2299	0.3140
	F-Score	0.3207	0.5171	0.5000	0.3617	0.4012	0.3509	0.2299	0.3724
	total	248	96	83	100	139	71	65	59

4   Conclusion

For the temporal overlap algorithm (TO), the addition of the community information reduces the link noise, irrespective of the scale — be it hourly or daily. This is shown by the decreasing value of the Frobenius norm in all the cases as compared to the value obtained using temporal overlap algorithm alone. In general shallow semantic approach (SSTO) performs the best and is only improved slightly by the loose incorporation of community information. For the SSTO algorithm, the daily or hourly community partition also does not affect the improvement. Table 2 shows how the dialog labeling generated from various algorithms agrees with the ground truth notations produced by a human labeler. Since TO only produces undirected links, we do not include it in the comparison. Plain SSTO generally results in a better precision and recall than SSTO plus either strict or loose community labeling. This is further confirmed from the visualizations for one hour of data in a day (both chosen randomly from the dataset) for each of the ground-truth, SSTO and TO as shown in figure 1, where the network from SSTO more closely resembles the ground-truth as compared to the on from TO.

The challenging nature of this dataset is evident in the overall low precision and recall scores, not only for the proposed algorithms but also for human labelers. We attribute this largely to the inherent ambiguity in the observed utterances. Among the

techniques, SSTO performs best, confirming that leveraging semantics is more useful than merely observing temporal co occurrence. We observe that community information is not reliably informative for SSTO but does help TO, showing that link pruning through network structure is useful in the absence of semantic information.

Acknowledgments

References

1. Weitnauer, E., Thomas, N.M., Rabe, F., Kopp, S.: Intelligent agents living in social virtual environments – bringing Max into Second Life. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 552–553. Springer, Heidelberg (2008)
2. Bogdanovych, A., Simoff, S.J., Esteva, M.: Virtual institutions: Normative environments facilitating imitation learning in virtual agents. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 456–464. Springer, Heidelberg (2008)
3. Second Life: Second Life Economic Statistics (2009), http://secondlife.com/whatis/economy_stats.php (retrieved July 2009)
4. Carley, K., Columbus, D., DeReno, M., Bigrigg, M., Diesner, J., Kunkel, F.: Automap users guide 2009. Technical Report CMU-ISR-09-114, Carnegie Mellon University, School of Computer Science, Institute for Software Research (2009)
5. Shi, L., Huang, W.: Apply social network analysis and data mining to dynamic task synthesis to persistent MMORPG virtual world. In: Proceedings of Intelligent Virtual Agents (2004)
6. Kahanda, I., Neville, J.: Using transactional information to predict link strength in online social networks. In: Proceedings of the Third International Conference on Weblogs and Social Media (2009)
7. Shah, F., Usher, C., Sukthankar, G.: Modeling group dynamics in virtual worlds. In: Proceedings of the Fourth International Conference on Weblogs and Social Media (2010)
8. Mann, W.C.: The dialogue diversity corpus (2003), http://www-bcf.usc.edu/~billmann/diversity/DDivers-site.htm (retrieved November 2010)
9. Tang, L., Liu, H.: Relational learning via latent social dimensions, in kdd 2009. In: Proceedings di of the 15th ACM SIGKDD International ti Conference on Knowledge, pp. 817–826. ACM, New York (2009)
10. Newman, M.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103, 8577–8582 (2006)
11. Newman, M.: Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 036104 (2006)
12. Prulj, N.: Biological network comparison using graphlet degree distribution. Bioinformatics 23(2) (2007)
13. Golub, G.H., Loan, C.F.V.: Matrix Computations, 3rd edn. JHU Press (1996)

Effects of Opposition on the Diffusion of Complex Contagions in Social Networks: An Empirical Study

Chris J. Kuhlman[1], V.S. Anil Kumar[1], Madhav V. Marathe[1],
S.S. Ravi[2], and Daniel J. Rosenkrantz[2]

[1] Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA
{ckuhlman,akumar,mmarathe}@vbi.vt.edu
[2] Computer Science Department, University at Albany – SUNY,
Albany, NY 12222, USA
{ravi,djr}@cs.albany.edu

Abstract. We study the diffusion of complex contagions on social networks in the presence of nodes that oppose the diffusion process. Unlike simple contagions, where a person acquires information or adopts a new ideology based on the influence of a single neighbor, the propagation of complex contagions requires multiple interactions with different neighbors. Our empirical results point out how network structure and opposing perspectives can alter the widespread adoption of social behaviors that can be modeled as complex contagions.

Keywords: social influence, opposition, complex contagions, networks.

1 Introduction and Motivation

Understanding how populations of interacting individuals change their views, beliefs, and actions has many applications. Among these are encouraging peaceful social changes, stopping or inciting mob behavior, spreading of information or ideologies, and sowing the seeds of distrust within a group [3, 5, 11].

Human behaviors in these situations are often complicated by significant risks. For example, people have been jailed for participating in peaceful demonstrations [4]. Also, in changing one's ideology or acting against prevailing norms, one risks reprisals to oneself and family [10]. In these types of situations, people do not change their attitudes in a flippant manner. Granovetter [5] studied such phenomena and introduced the concepts of simple and complex contagions to describe how one's social acquaintances can influence a person's choices.

A social network is a graph in which vertices represent people and an edge between two nodes indicates a certain relationship (e.g. the two corresponding persons are acquaintances of each other). **A simple contagion** is an entity such as a belief, information, rumor, or ideal that can propagate through a network such that a person acquires the contagion and is willing to pass it on based on a *single* interaction with *one* neighbor. A **complex contagion**,

in contrast, requires *multiple* interactions with (different) sources in order to accept the contagion and propagate it to others. Associated with these ideas is the concept of a **threshold** t, which is the number of neighbors with a particular view that is required for a person to change her mind to the same point of view. Thus, a simple contagion corresponds to $t = 1$ while a complex contagion corresponds to $t > 1$. Actions with greater consequences are generally modeled as complex contagions because one requires affirmation to change her beliefs [3].

Resistance offered by opponents is another complicating factor in the spread of a contagion. For example, if one seeks to change a person's view, there will often be those with opposite opinions that seek to prevent the change. One approach to model this counter-influence is to treat these network nodes as having a fixed (old) opposing view that will not change, thus inhibiting diffusion.

In this work, we carry out a simulation-based study of the diffusion of complex contagions on synthetic social networks. A diffusion instance (called an **iteration**) consists of the following. We specify a subset of nodes as initially possessing the contagion (corresponding to nodes in state 1), while all other nodes (which are in state 0) do not possess the contagion. We also specify a fraction f of **failed** nodes (following the terminology in [1]) that are in state 0 initially and will remain in state 0 throughout the diffusion process. Failed nodes are determined in one of two ways, following [1]: **random** failed nodes are chosen uniformly at random from the entire set of nodes; **targeted** failed nodes are nodes with the greatest degrees. These failed nodes represent the counter-influence that opposes the propagation. The diffusion process progresses in discrete time steps. At each time step, a node changes from state 0 to state 1 when at least t of its neighbors are in state 1, but nodes never change from state 1 to 0; hence, this is a **ratcheted** system [8, 9]. A diffusion instance ends when no state changes occur.

We examine two types of networks. Scale-free (`SF`) networks possess hub nodes with high connectivity to other nodes, and are representative of many social networks [6]. Exponential decay (`ED`) networks have a tighter range of node degrees; their usefulness in modeling social behavior is discussed in [1].

Our empirical results provide insights for two perspectives: (1) how the diffusion process can spread widely even in the presence of failed nodes, and (2) how diffusion can be resisted by choosing the number and type of failed nodes. The significance of our results in modeling social behavior is discussed in Section 3, following discussions of the model and related work. Due to space limitations, a number of details regarding our experimental procedures and results are not included in this paper; these appear in a longer version [8].

2   Dynamical System Model and Related Work

2.1   System Model and Associated Definitions

For studying diffusion in networks, we use the **synchronous dynamical system** (SyDS) model detailed in [8, 9]. Here we provide a brief account of it.

In this model, each node of a social network represents an individual and each undirected edge $\{u, v\}$ indicates that the contagion may propagate from u to v

or from v to u. Each node has a state value which may be 0 indicating that the node is **unaffected** (i.e., the contagion has not propagated to that node) or 1 indicating that the node has been **affected** (i.e., the contagion has propagated to that node). Once a node's state changes to 1, it remains in that state for ever.

State changes at the nodes of a social network are determined by **local transition functions**. Each of these is a Boolean t-threshold function which can be described as follows. The function f_v at a node v has value 1 if the state of v is 1 or at least t of the neighbors of v are in state 1; else $f_v = 0$. At each time step, all the nodes compute their local transition function values *synchronously*; the value of the function gives the new state of the node. We use the term *t*-**threshold system** to denote a SyDS in which the local transition function at each node is the t-threshold function. (The threshold t is the same for all the nodes.) When no node changes state, the system is said to have reached a **fixed point**. It is known that every such ratcheted SyDS reaches a fixed point within n steps, where n is the number of nodes in the underlying graph [8]. In the context of contagions, a fixed point marks the end of the diffusion process.

For any SyDS, nodes which are in state 1 at time 0 are called **seed** nodes. The **spread size** is the number (or fraction) of nodes in state 1 (i.e., the number of affected nodes) at the end of the diffusion process. A **cascade** occurs when the spread size is a large fraction of all possible nodes; that is, most of the nodes that can be affected are indeed affected. An example to show that node failures can significantly affect the time needed to reach a cascade is presented in [8].

2.2   Related Work

The average time required to reach a fixed point in ratcheted dynamical systems with relative thresholds (i.e., thresholds normalized by node degrees) was examined in [2, 3, 12]. The work presented in [2, 3] considered small-world and random networks generated by rewiring network edges. The focus was on measuring the average spread size as a function of time and the average number of time steps to cascade as a function of the rewiring probability. We know of no work on complex contagions that examines the time histories of individual diffusion instances to assess variations in behavior with respect to the time to cascade, with or without failed nodes.

The effect of failed nodes on the ability of SF and ED graphs to propagate complex contagions using ratcheted dynamics was investigated by Centola [1]. His work used one average degree value $d_{ave} = 4$ and one threshold value ($t = 2$). We extend that work by examining comparable networks with twice the average degree ($d_{ave} = 8$) to examine the influence of this parameter. Other works have examined the thwarting of diffusion by specifying a set of blocking nodes tailored for a *particular* set of seed nodes (e.g., [9] and references therein). The focus here, in contrast, is network dynamics for blocking or failed nodes that are *fixed* a priori for *all* sets of seed nodes. Watts [12] primarily examines Erdős-Rényi random graphs, which are different from ours, and his generating function approach cannot be extended directly to analyze the dynamics with targeted failed nodes.

3   Results and Their Significance

We now provide a summary of our results and their significance in modeling social behavior.

(a) For the diffusion of desirable social changes that can be modeled as complex contagions (e.g. ideology), the required duration of commitment to achieve the goal can vary significantly depending on the sets of individuals that initiate the behavior and those who oppose it; in particular, the variation can be by a factor of 3 or more (Section 5.1).

(b) Sustained resistance may be required to inhibit undesirable complex contagions. Vigilant monitoring may be necessary because a contagion may initially propagate very slowly, and only at a much later time reach a critical mass whereupon spreading is rapid. Furthermore, once the rate of diffusion increases, there is relatively little time for countermeasures (Section 5.1).

(c) In critical applications where diffusion must be thwarted, the requisite number of opposing individuals can be much larger (e.g. by three orders of magnitude) compared to the number of individuals who initiated the behavior. Moreover, the opposing population may need to be 10% or more of all the nodes in the social network, particularly if the region from where a contagion may emerge is uncertain (Section 5.2).

(d) Whether fostering or blocking diffusion, network structure matters. Swaying the view of a well-connected node in an SF network has a greater impact than one in an ED network – this effect is accentuated in graphs with higher average degree (Section 5.2).

(e) Owing to the nature of ratcheted dynamical systems, it is instructive to separate the (1) probability of a cascade occurring and (2) the average cascade size, given that a cascade occurs. These single-parameter measures provide greater insight than computing a normalized average spread size (as done in [1]) over all diffusion instances (Section 5.3).

4   Networks and Experiments

One ED network and one SF network are evaluated in this study (these names were chosen to be consistent with [1]). Both are composed of 10000 nodes. The ED and SF networks have, respectively, 39402 and 39540 edges, average degrees of 7.88 and 7.91, and average clustering coefficients of 0.244 and 0.241. The rationale for choosing these networks is as follows. First, we wanted to investigate exponential decay and scale-free networks to be consistent with Centola's work [1] so that we can compare the two sets of results. Centola used ED and SF networks with an average degree of 4. We chose networks with an average degree close to 8 for two reasons. First, we want to study the resilience properties of such networks when the average degree differs significantly from the threshold value. Second, there are social networks in the literature that have degree close to 8 [6], so that as part of a larger effort, we are planning to compare these synthetic networks to real social networks of the same average degree. We chose

an average clustering coefficient of about 0.24 to compare with the value 0.25 used in Centola's work [1]. Network details are given in [8].

A diffusion instance, called an iteration, was described in Section 1. A **simulation** consists of 400 iterations, which is comprised of 20 failed node sets and 20 seed node sets for each failed node set. All iterations in a simulation utilize the same network, threshold, failed node determination method, and fraction f of failed nodes. Table 1 contains the parameters and their values used in this experimental study. Simulations were run with all combinations of the parameters in this table. Failed nodes are determined first, and then seed node sets are generated randomly from the remaining nodes. Further details regarding the choices of seed nodes and failed nodes are given in [1, 8].

Table 1. Parameters and values of parametric study

Networks	Failed Node Methods	Seeding Method	Threshold, t	Failed Node Fraction, f
ED, SF	random, targeted	set one node and all neighbors to state 1	2	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 2×10^{-1}

5   Experimental Results

This section presents the empirical results from our simulations on ED and SF networks. The results point out the impact of failures on the time needed to reach a cascade, the size of cascades and the probability of cascades.

5.1   Time to Reach a Cascade with a Fixed Number of Failures

Figures 1(a) and 1(b) depict the cumulative fraction of affected nodes as a function of time for the ED and SF networks, respectively, with random failures and $f = 0.2$. (Throughout this section, the captions for figures give the parameters used in generating the corresponding plots.) For each simulation, each of the 400 iterations is represented by a gray curve, and the average over the 400 diffusion instances is represented by the black curve. The ED results show significant scatter among iterations in the times for the diffusion to pick up speed (i.e., when the curves turn upwards) and for a fixed point to be reached (i.e., when the curves plateau). The time variation is roughly a factor of three. Note that some iterations have small spread sizes. It can be seen that the times to reach a fixed point are much shorter in Figure 1(b) for the SF network. The hub nodes of high degree, when not specified as failed nodes, greatly assist the spreading process. Essentially, hubs reduce the diameter of the SF network, compared to that of the ED network, so that if cascades occur, they occur quickly. The targeted failed node results for ED and $f = 0.2$ are qualitatively the same as those in Figure 1(a), but targeted failures in the SF network completely thwart all diffusion.

Figure 1(c) shows that as f increases, variability in times to reach a fixed point increases for the ED network. From Figures 1(a) and 1(b), where the time

to reach a fixed point follows soon after diffusion speed up, the data in Figure 1(c) also reflect the variation in time to speed up as a function of f. If a goal is to spread an ideology or fad, Figures 1(a) and 1(b) also show that depending on seed and failed node sets and network structure, the required duration of commitment can vary by a factor of 10, with commensurate variations in costs.

(a)   (b)   (c)
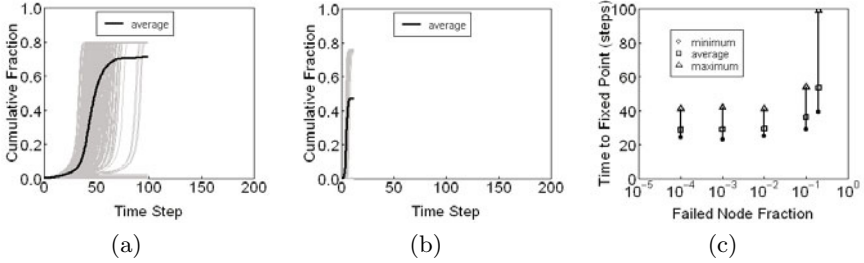
Fig. 1. Cumulative fraction of affected nodes as a function of time for $t = 2$ and $f = 0.2$: (a) ED network and random failed nodes, and (b) SF network and random failed nodes. 400 iterations are shown in gray in each plot. In (c), the minimum, average, and maximum times to reach a fixed point for iterations with large spread sizes are plotted for the ED network and random failed nodes.

5.2 Effect of Failures on Cascade Size

In Figure 2 we combine our results with those of Centola [1] to examine the average normalized spread size for both types of networks as a function of random and targeted failures and average degree. The average normalized spread size $\langle S \rangle$, a measure used in [1, 3], is given by $\langle S \rangle = \sum_{k=1}^{n_i} S_k/n_i$ where the sum is over the $n_i = 400$ iterations and S_k is the spread size for the kth iteration. (Other measures are used later in this section.) Figure 2(a) shows results for the ED network. The dashed curves are for $d_{ave} = 4$, taken from Centola [1], while the solid curves are for $d_{ave} = 8$ and were generated in this work. For a given f value, the $d_{ave} = 8$ curves (solid lines) are above those for $d_{ave} = 4$. This is to be expected, since a greater average degree provides more neighbors from which a node can receive the contagion.

The relationship between the magnitudes of d_{ave} and t is important. With $d_{ave} = 4$ and $t = 2$, a node v with degree only one less than the average needs every neighbor to acquire the contagion and pass it on (v requires two neighbors to receive the contagion and a third neighbor to which it can transmit). Hence, one would expect a significant difference when d_{ave} increases to 8. In this case, each node on average has 8 neighbors from which only 3 are needed to contract the contagion and help in passing it on. This is why the curves differ even with f as small as 10^{-5}, where there are no failed nodes: the network structure plays a role irrespective of failed nodes. We confirm these observations in Figure 2(b) for the SF network. The effect of a greater average degree on random failures (solid gray curve as compared to the dashed gray curve) is to drive the curve

significantly upward to the point that it is moving toward the shape of the solid gray curve for the ED network in Figure 2(a). These results for increasing d_{ave} are significant because real social and contact networks can have average degrees much larger, on the order of 20 or 50 or more [6, 7]. Further, targeted degree based failures have much more impact than random failures. Hence, from a practical standpoint, if there are relatively few people with a great number of neighbors (i.e., highly influential people), it is important to convert them to one's point of view, for if they are in the opposition, they can markedly decrease $\langle S \rangle$, and as we shall see in the next section, the probability of achieving widespread diffusion.

(a)   (b)
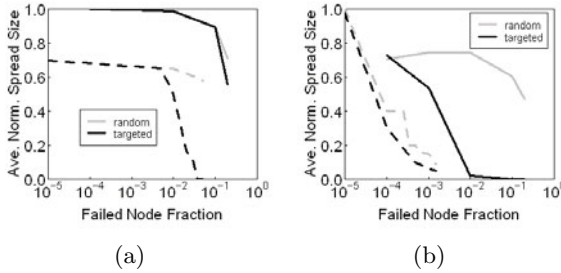
Fig. 2. The average normalized spread size $\langle S \rangle$ as a function of failed node fraction f for $t = 2$: (a) ED network and (b) SF network. Dashed and solid lines are for average degrees of 4 and 8, respectively. Dashed curves are taken from [1].

5.3   Effect of Failures on Cascade Probability

In discussing the previous results, we used the $\langle S \rangle$ measure defined in [1]. Here, we evaluate our simulation results in a different manner in order to gain greater insights into the network dynamics. Figure 3(a), generated for the SF network and targeted failed nodes, is constructed in the following fashion. In one simulation, say for $f = 10^{-3}$, 400 iterations are performed with 20 failed node sets and 20 seed sets. For each iteration, the fraction of affected nodes (at the fixed point) is obtained. These 400 values are arranged in increasing numerical order and are plotted as the black curve in Figure 3(a). The black curve lies along the horizontal axis from abscissa value of 0 until it rises sharply to 1.0 at an abscissa value of 0.45. The gray curve for $f = 10^{-2}$ remains along the x-axis until it extends vertically at an abscissa value of 0.98. In this case, however, the curve turns horizontal at a final fraction of nodes affected of 0.9. The features of these curves are representative of the behavior of deterministic ratcheted SyDSs across a wide range of networks, both synthetic like those evaluated here and social networks generated from, for example, online data, and for diffusion processes with and without "blocking" nodes such as the failed nodes examined here (e.g., [8, 9]). Plots of this nature seem to be fundamental to ratcheted systems.

The point of transition of a curve in Figure 3(a) is the probability of a cascade and the y-axis value after the transition is simply the cascade size for iterations that cascade. Although not apparent from Figure 3(a), this cascade size is

essentially constant. Other curves from this study and other types of studies [8, 9] also demonstrate this behavior. This means that we can plot the probability of a cascade and the average cascade size—each of which is a one-parameter characterization—as a function of f (one f value corresponding to each curve in Figure 3(a)). This is done in Figures 3(b) and 3(c). The black curves in these latter two figures were generated from the data in Figure 3(a) for targeted failures; the gray curves were generated from random failure data. If we compare these last two figures with the solid curves of Figure 2(b), it is obvious that the dominant driver in the change of $\langle S \rangle$ is the change in the probability of a cascade. As an example, consider $f = 5 \times 10^{-3}$ and targeted failures for $d_{ave} = 8$ in Figure 2(b), which gives an *average* normalized spread size of 0.3. If our goal is to prevent a cascade, it appears that this f is satisfactory. However, this conclusion is misleading because as Figure 3(c) illustrates, *if* widespread diffusion is achieved, then the contagion will reach over 90% of nodes.

(a)   (b)   (c)
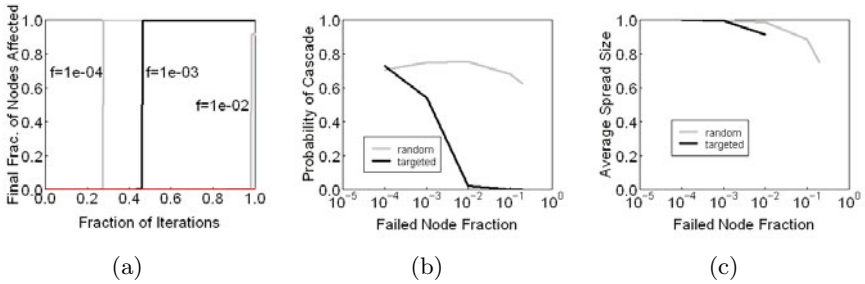
Fig. 3. (a) For the SF network, ordered spread size for 400 iterations, for $t = 2$ for each curve. (b) Probability of a cascades as a function of f. (c) Average spread size for iterations that cascade as a function of f.

6   Conclusions and Future Work

Our results show that time to cascade for particular diffusion instances can vary significantly from mean behavior, that the effects of a threshold must be examined in light of network parameters such as degree distribution, and that for ratcheted systems, it is often possible to isolate the size of a cascade from the probability of a cascade. Future work includes study of realistic social networks.

References

1. Centola, D.: Failure in Complex Social Networks. Journal of Mathematical Sociology 33, 64–68 (2009)
2. Centola, D., Eguiluz, V., Macy, M.: Cascade Dynamics of Complex Propagation. Physica A 374, 449–456 (2007)
3. Centola, D., Macy, M.: Complex Contagions and the Weakness of Long Ties. American Journal of Sociology 113(3), 702–734 (2007)
4. Euchner, C.: Nobody Turn Me Around. Beacon Press (2010)
5. Granovetter, M.: Threshold Models of Collective Behavior. American Journal of Sociology 83(6), 1420–1443 (1978)
6. Leskovec, J.: website, http://cs.stanford.edu/people/jure/
7. Khan, M., Kumar, V., Marathe, M., Zhao, Z., Dutta, T.: Structural and Relational Properties of Social Contact Networks with Applications to Public Health Informatics, NDSSL Technical Report No. 09-066 (2009)
8. Kuhlman, C., Kumar, V., Marathe, M., Ravi, S., Rosenkrantz, D.: Effects of Opposition on the Diffusion of Complex Contagions in Social Networks, NDSSL Technical Report No. 10-154 (2010)
9. Kuhlman, C.J., Anil Kumar, V.S., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J.: Finding Critical Nodes for Inhibiting Diffusion of Complex Contagions in Social Networks. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010. LNCS, vol. 6322, pp. 111–127. Springer, Heidelberg (2010)
10. Nafisi, A.: Reading Lolita in Tehran: A Memoir in Books. Random House (2003)
11. Nicholls, W.: Place, Networks, Space: Theorising the Geographies of Social Movements. Transactions of the Institute of British Geographers 34, 78–93 (2009)
12. Watts, D.: A Simple Model of Global Cascades on Random Networks. PNAS 99, 5766–5771 (2002)

Promoting Coordination for Disaster Relief – From Crowdsourcing to Coordination

Huiji Gao, Xufei Wang, Geoffrey Barbier, and Huan Liu

Computer Science and Engineering
Arizona State University
Tempe, AZ 85281
{Huiji.Gao,Xufei.Wang,Geoffrey.Barbier,Huan.Liu}@asu.edu

Abstract. The efficiency at which governments and non-governmental organizations (NGOs) are able to respond to a crisis and provide relief to victims has gained increased attention. This emphasis coincides with significant events such as tsunamis, hurricanes, earthquakes, and environmental disasters occuring during the last decade. Crowdsourcing applications such as Twitter, Ushahidi, and Sahana have proven useful for gathering information about a crisis yet have limited utility for response coordination. In this paper, we briefly describe the shortfalls of current crowdsourcing applications applied to disaster relief coordination and discuss one approach aimed at facilitating efficient collaborations amongst disparate organizations responding to a crisis.

Keywords: Disaster Relief, Crisis Map, Crowdsourcing, Groupsourcing, Response Coordination, Relief Organization.

1   Introduction

Natural disasters have severe consequences including causalities, infrastructure and property damages, and civilian displacement. The 2004 Indian Ocean earthquake and tsunami killed 230,000 people in 14 countries. Hurricane Katrina occurred in 2005 killed at least 1,836 people and resulted in approximately $81 billion [4] in economical loss. The catastrophic magnitude 7.0 earthquake in Haiti on January 12, 2010 resulted in more than 230,000 deaths, 300,000 injuries, and one million homeless.

Quality data collection from disaster scenes is a challenging and critical task. Timely and accurate data enables government and non-governmental organizations (NGOs) to respond appropriately. The popularity and accessibility of social media tools and services has provided a new source of data about disasters. For example, after the devastating Haiti earthquake in January 2010, numerous texts and photos were published via social media sites such as Twitter, Flickr, Facebook, and blogs. This is a type of data collection and information sharing that strongly leverages participatory social media services and tools known as crowdsourcing [3]. Crowdsourcing is characterized by collective contribution and has been adopted in various applications. For instance, Wikipedia depends

on crowds to contribute wisdom without centralized management. Customer reviews are helpful in buying products from Amazon.

Collaboration is a process to share information, resources, and cooperate among various organizations. Collaboration in disaster relief operations is usually decentralized because of the relative independence of relief organizations. Attempts have been made to help the relief community to enhance cooperation. *Haiti Live*[1], which leverages web 2.0 technologies to collect data from various sources such as phones, the web, e-mail, Twitter, Facebook, etc., provides an up-to-date publicly available crisis map that is in turn available to relief organizations. However, without centralized control, it is difficult to avoid conflicts such as responding to the same request repeatedly by different organizations. An approach to enable efficient collaboration between disparate organizations during disaster relief operations is imperative in order for relief operations to be successful in meeting the needs of the people impacted by the crisis.

This paper is organized as follows: we summarize related work in disaster relief management systems. We discuss the need for a collaboration system and highlight our approach.

2   Related Work

Dynes et al. [1] propose a disaster zone model which consists of five spatial zones of *Impact*, *Fringe*, *Filter*, *Community*, and *Regional* and eight socio-temporal disaster stages. Goolsby et al. [2] demonstrate the possibility in leveraging social media to generate community crisis maps and introduce an *inter-agency* map not only to allow organizations to share information, but also to collaborate, plan, and execute shared missions. The inter-agency map is designed to share (some) information between organizations if the organizations share the same platform or have similar data representation formats. In [5], Sophia and Leysia summarize 13 crisis-related mashups to derive some high-level design directions of next generation crisis support tools.

Ushahidi is a crisis map platform created in 2007. The platform has been deployed in Kenya, Mexico, Afghanistan, Haiti, and other locations. Ushahidi can integrate data from various sources: phones, a web application, e-mail, and social media sites such as Twitter and Facebook. This platform uses the concept of crowdsourcing for social activism and public accountability to collectively contribute information, visualize incidents, and cooperate among various organizations.

Sahana[2] is an open source disaster management system which was started just after the 2004 Indian Ocean tsunami. It provides a collection of tools to help manage coordination and collaboration problems resulting from a disaster. The major functions are supporting the search for missing persons, coordinating relief efforts, matching donations to needs, tracking the status of shelters, and the reporting of timely information. Additionally, Sahana facilitates the management

---

[1] http://haiti.ushahidi.com/
[2] http://www.sahanafoundation.org/

of volunteers by capturing their skills, availability, allocation, etc. This system has been deployed in several countries [6].

3   Why Crowdsourcing Falls Short for Disaster Relief

Although crowdsourcing applications can provide accurate and timely information about a crisis, there are three reasons why current crowdsourcing applications fall short in supporting disaster relief efforts. First, and most importantly, current applications do not provide a *common mechanism* specifically designed for collaboration and coordination between disparate relief organizations. For example, microblogs and crisis maps do not provide a mechanism for apportioning response resources. Second, current crowdsourcing applications do not have adequate *security features* for relief organizations and relief operations. For example, crowdsourcing applications that are publicly available for reporting are also publicly available for viewing. While this is important for providing information to the public it can create conflict when decision must be made about where and when relief resources are needed. Additionally, in some circumstances, relief workers themselves are targeted by nefarious groups. Publicizing the details of relief efforts can endanger relief workers. Third, data from crowdsourcing applications, while providing useful information, does not always provide all of the *right information* needed for disaster relief efforts. There are often duplicate reports and information essential for relief coordination is not readily available or easily accessible such as lists of relief resources or communication procedures and contact information for relief organizations.

4   How to Facilitate Disaster Relief

In order to facilitate efficient coordination during relief efforts, relief organizations need to leverage the information available from the group. Supplementing the crowdsourcing information that is available through social media, the relief organizations can contribute to a unified source of information customized for the group or "groupsourced." We define *groupsourcing* as intelligently using information provided by a sanctioned group comprised of individuals with disparate resources, goals, and capabilities. Essentially, the response group is performing crowdsourcing but specialized for the entire group and taking it a few steps further. Using information provided from the crowd and the sanctioned members of the group, a relief response can be coordinated efficiently.

In order to be most efficient and most effective in helping to resolve the crisis, the members of the response group should subscribe to the centralized administrative control of an information management system to ensure data integrity, data security, accuracy, and authentication will be realized for each member of the response group. One member of the response group can field the response coordination system ensuring equal access to other members of the response group. This data management system can leverage the latest social media technology and supplement functionality customized for the group.

5   What an Application Needs

A disaster relief information system designed for better collaboration and coordination during a crisis should contain four technical modules: request collection, response, coordination, and statistics, as shown in Figure 1. User requests are collected via crowdsourcing and groupsourcing and stored in a requests pool after preprocessing. The system visualizes the requests pool on its crisis map. Organizations respond to requests and coordinate with each other through the crisis map directly. A statistics module runs in the background to help track relief progress. ACT (ASU Coordination Tracker) is an open disaster relief coordination system that implements the groupsourcing system architecture.
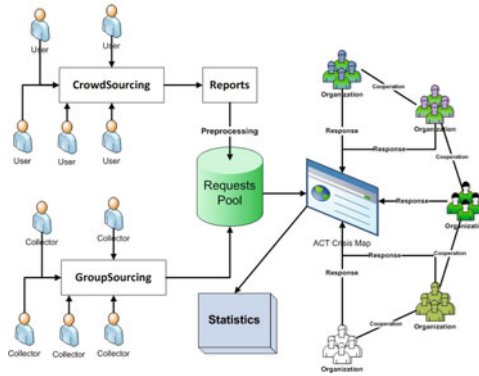
Fig. 1. Groupsourcing System Architecture

5.1   Request Collection

Relief organizations need two types of requests: requests from crowds (crowdsourcing) and requests from groups (groupsourcing). Crowdsourcing refers to requests submitted by people (e.g. victims, volunteers) who are not from certified organizations. Crowdsourcing data from disaster scenes are valuable for damage assessment and initial decision making. However, these data are usually subjective, noisy, and can be inaccurate. The groupsourcing requests originate from responding organizations such as United Nation, Red Cross, etc. The key difference between these two types of requests is the level of trustworthiness. Requests from the relief organizations are more likely objective and accurate compared to those from crowds. Sample requests are shown in Figures 2 and 3.

Figure 4 shows an example crisis map (left picture) viewing fuel requests. Each green node represents an individual fuel request and each red node represents a *request cluster*, which consists of multiple individual requests (or additional request clusters) that are geographically close to each other. The number and the node size of each individual request represents the requested quantity, for example, 113 (left picture) means this request asks for 113 k-gallon fuels in that region. Individual requests in each cluster can be further explored by zooming-in (right picture).

Fig. 2. Crowdsourcing

Fig. 3. Groupsourcing
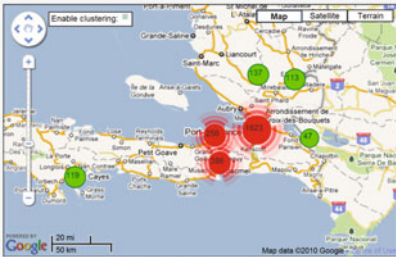
Fig. 4. ACT Crisis Map

5.2   Request Response

A key aspect of efficient coordination is enabling relief organizations to respond
in an organized and collaborative manner. Response actions need to be supported
by the information system such that all of the relief organizations can contribute,
receive, and understand the response options and planned response actions. The
system must be available on heterogeneous hardware and software platforms
(i.e. web services and contemporary commercially available portable devices).
Figure 4 shows how a requests pool can be visualized on a crisis map. As the
map resolution changes, nodes indicating the same type of requests will merge
or split based on their location closeness.

Here is how a response transaction may flow. Once an organization decides
to respond to a request, it selects the node and adds it into a relief package.
Responders are allowed to add multiple nodes of various categories into the re-
lief package. The requests that are being fulfilled will be removed from the relief

Fig. 5. ACT Response Process

Fig. 6. ACT Relief Package

package and invisible on the crisis map. The response process and a sample relief package are demonstrated in Figures 5 and 6.

5.3   Response Coordination

Relief organizations that respond independently can complicate relief efforts. If independent organizations duplicate a response, it will draw resources away from other needy areas that could use the duplicated supplies, delay response to other disaster areas, and/or result in additional transportation and security requirements. We base our approach to response coordination on the concept of "inter-agency [2]" to avoid response conflicts while maintaining the centralized control.

To support response coordination, all available requests (i.e. requests that have not been filled) are displayed on the crisis map. Relief organizations look at available requests and selectively respond to the requests they can support. To avoid conflicts, relief organizations are not able to respond to requests another
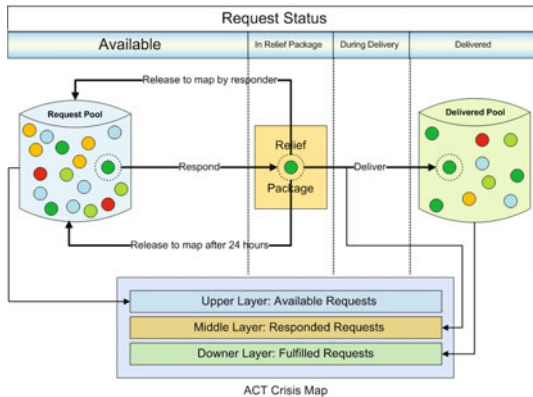
Fig. 7. Request States Transition in ACT

organization is fulfilling. Claimed requests that are not addressed for 24 hours become visible again on the crisis map.

Each request is in one of four states: *available*, *in process*, *in delivery*, or *delivered*. After necessary preprocessing, submitted requests are put into the requests pool. Requests become *available* and visualized in the crisis map. If a request is selected by an organization, the state changes to *in process*. The request becomes *available* again if it is not addressed in 24 hours. The request state becomes *in delivery* if it is on the way to the final destination. Finally, the state becomes *delivered* if the request is fulfilled. A detailed state transition is summarized in Figure 7.

5.4   Statistics for Improved Communication

Relief organizations are usually not informed of other responders' actions since they are only aware of the available requests. Statistics such as organizational contribution during relief operations are helpful in evaluating the relief progress, adjusting relief strategies and making further decisions. The following statistics give insights into the relief effort:

– Current request delivery status. Information about current fulfillment status for each type of requests.
– Spatio-temporal distribution of requests. *Where* and *when* information about requests and responses.
– Distribution of contributed resource types of each organization. Each organization has its own specialization and strength in relief operations as well as different budgets on various resources.

An example of distribution statistical reporting is shown in Figure 8. The left pie graph represents an example of detailed water requests, note that 10% of the requests are not fulfilled. The right bar graph represents the water delivery at a 10-day resolution in an earthquake for the Red Cross and United Nations.

Fig. 8. Statistics on water fulfillment and organizational contribution

6 Conclusion and Future Work

Social media is being used as an efficient communications mechanism during disasters. In this work, we discussed the challenges with coordination for disaster relief and present an architecture aimed at improving disaster response beyond current practice. We believe this approach will enable organizations to cooperate with each other more efficiently.

We are developing an open system, ACT, with the primary goal of providing relief organizations the means for better collaboration and coordination during a crisis. The system implementation is ongoing and additional work needs to be done to test our approach. We are investing approaches to enable collaboration and provide appropriate security to relief organizations and workers. We are designing reporting functions that are insightful for disaster relief.

By both leveraging crowdsource information and providing the means for a groupsource response, organizations will be able to assist persons affected by a crisis more effectively in the future.

Acknowledgments

References

1. Dynes, R.R.: Organized Behavior in Disaster: Analysis and Conceptualization. US Dept. of Commerce, National Bureau of Standards (1969)
2. Goolsby, R.: Social media as crisis platform: The future of community maps/crisis maps. ACM Trans. Intell. Syst. Technol. 1(1), 1–11 (2010)
3. Howe, J.: The rise of crowdsourcing. Wired Magazine 14(6), 1–4 (2006)
4. Knabb, R.D., Rhome, J.R., Brown, D.P.: Tropical cyclone report: Hurricane katrina. National Hurricane Center 20 (August 23-30, 2005)
5. Liu, S.B., Palen, L.: Spatiotemporal mashups: A survey of current tools to inform next generation crisis support. In: Proceedings of the 6th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2009 (2009)
6. Samaraweera, I., Corera, S.: Sahana victim registries: Effectively track disaster victims. In: Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2007 (2007)

Trust Maximization in Social Networks

Justin Zhan and Xing Fang

Department of Computer Science
North Carolina A&T State University
{Zzhan,xfang}@ncat.edu

Abstract. Trust is a human-related phenomenon in social networks. Trust research on social networks has gained much attention on its usefulness, and on modeling propagations. There is little focus on finding maximum trust in social networks which is particularly important when a social network is oriented by certain tasks. In this paper, we propose a trust maximization algorithm based on the task-oriented social networks.

Keywords: Social Networks; Trust; Maximization.

1 Introduction

With the advent of web 2.0 technologies, social networks are propagating rapidly. Instead of granting users passive browsing web content, it allows users to have personal accounts. The users may post their own words, pictures, and movies, etc. to the social network sties. Users on the social networks can also write comments, share personal preferences, and make friends with people they want. Examples of such networks are Facebook, MySpace, Twitter, and other blogs or forums. While more and more people are registering accounts on social networks and enjoying the new technologies [1], problems inherent with this not-well developed network environment finally came into reality. Hogben [2] pointed out that Trust (Trustworthy Users) will be one major security issue for the future social networks.

Trust has been being a topic that intrigues intensive research work from numerous computational scientists. Owing to its human-related characteristics, trust is hard to be modeled or even uniformly defined for computational systems. Marsh introduced a trust computing model [3]. Unfortunately, this frequently-referenced work is highly theoretical and difficult to apply, particularly on social networks [4]. Sztompka [5] defined trust as a bet about the future contingent actions of others. Grandison and Sloman [6] addressed that trust is the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context. Mui et al. claimed that trust is a subjective expectation an agent has about another's future behavior based on the history of their encounters [7]. Olmedilla et al. presented that trust of a party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period within a specified context (in relation to service X) [8].

As an important property, trust has also been involved in different types of networks. Online shopping networks such as Amazon and eBay allow buyers to evaluate sellers by posting comments and trust ratings after every purchase. Online

P2P lending network, like prosper.com [9], is based on mutual trust [10]. Trust is also served as a vital factor for selecting helpers in social authentication [11]. In addition to focusing on direct trust, trust inference mechanisms are introduced to infer an accurate trust value between two people who may not have direct connections (communications) on social networks [12]. Kamvar et al. proposed the Eigen Trust Algorithm, where the idea is to eliminate malicious peers on P2P networks through reputation management [13]. The algorithm is able to calculate a global trust value for each peer on the networks based on each peer's historical behavior. The ones with relatively low values will be identified as malicious peers and will then be isolated. Similar approaches were adopted in [14, 15] to calculate trust values for the users on the semantic web and the trust-based recommendation system, respectively.

The remainder of this paper is organized as follows: in section 2, we reviewed existing trust inference mechanisms. In section 3, we present the trust maximization algorithm based on task-oriented social networks.

2   Trust Inference Mechanisms

Trust Inference Mechanism is designed to calculate an accurate trust value between two persons who may not have direct connections on social network. Liu et al. [16] categorize existing Trust Inference Mechanisms into three different types: Multiplication, Average, and Probabilistic mechanisms.

2.1   Multiplication Mechanism

Trust transitivity and asymmetry are two major attributes for trust computation [4]. If person A highly trusts person B with value T_{AB} and person B highly trusts person C with value T_{BC}, A then is able to trust person C, to some extent, with value T_{AC} [17, 18]. Thus trust transitivity is represented as: $T_{AC} = T_{AB} * T_{BC}$. While trust asymmetry describes the truth that A trusts B (C) with value T_{AB} (T_{AC}) does not indicate B (C) trusts A with the same value. Titled as multiplication, the first type of trust inference mechanism is theoretically originated from trust transitivity as well as asymmetry. Ziegler and Lausen [19] propose Appleseed, a local group trust computation metric, to examine trust propagation via local social networks. Bhuiyan et al. [20] present a framework integrated with trust among community members and public reputation information for recommender systems. Zuo et al. [21] introduce a transitive trust evaluation framework to evaluate trust value for social network peers.

2.2   Average Mechanism

Golbeck and Hendler [4] present an average trust inference mechanism. Trust value between two indirect connected nodes, source and sink, can be calculated via averaging the trust ratings to the sink that belong to the good nodes in between of the two. They believe in a graph with n nodes, probability of the bulk of the nodes will precisely rate the sink is following a binomial distribution, $\sum_{i=\frac{n}{2}}^{n} \binom{n}{i} a^i (1-a)^{n-i}$, where n denotes the number of the nodes and a denotes the accuracy that the nodes can correctly rate the sink. In this case, the magnitude of trust value actually depends on n and a.

2.3 Probabilistic Mechanism

The third type of trust inference mechanism belongs to the probabilistic method. Kuter and Golbeck [22] proposed a Probabilistic Confidence Model for the trust value computing. The structure of this confidence model is based on a Bayesian Network in which the confidence values, which are used to estimate the trust values, are introduced and their bounds are calculated. Liu et al. [16] propose a trust inference mechanism based on trust-oriented social networks. In their trust inference mechanism, they introduce two variables r ($0 \leq r \leq 1$) and ρ ($0 \leq \rho \leq 1$). r is an intimate degree value of social relationships between participants in trust-oriented social networks. If $r_{AB} = 0$, it indicates that A and B have no social relationship. $r_{AB} = 1$ indicates A and B have the most intimate social relationship. ρ is a role impact value. It illustrates the impact of the participants' recommendation roles (expert or beginner) in a specific domain in trust-oriented social networks. If $\rho_A = 0$, it indicates that A has no knowledge in the domain. If $\rho_A = 1$, it indicates A is a domain expert.

3 Task-Oriented Social Network and Trust Maximization Alg

A task-oriented social network is able to accomplish certain tasks based on its formation. A general social network is defined as a connected graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of n nodes, and $E \subseteq V \times V$ is a set of edges [23]. For each $v_i \in V$, it denotes an individual located on the social network. For each pair of $(u, v) \in E$, it denotes the connection (communication) between two nodes. $A = \{a_1, a_2, \ldots, a_m\}$ is a universe of m skills. By denoting $a_j \in v_i$, we claim that the individual i has the skill j. And each individual is associated with a small set of skills, where $A(v_i) \subset A$. A task $\mathcal{T}$ is defined as a subset of skills required to accomplish the task, that is $\mathcal{T} \subseteq A$. For a certain task, that can be performed by a group of individuals, V', where $V' \subseteq V$. Hence, this group of individuals forms a task-oriented social network, G', where $G' \subseteq G$ and $G' = (V', E')$.

3.1 Task-Oriented Social Network Generation (TOSNG)

PROBLEM 1: /*TOSNG*/ Given a general social network $G = (V, E)$ and a task $\mathcal{T}$, find a task-oriented social network $G' = (V', E')$ that $A(V') \supseteq A(\mathcal{T})$.

The *TOSNG* problem is essentially a Team Formation (*TF*) problem where a formed team aims to accomplish certain tasks or missions [23]. Several conditions that are able to affect the formation process are considered. The conditions may include individuals' drive, temperament [24] and their personalities [25]. Moreover, communication-cost is introduced as a constraint of the *TF* problem [23] where Lappas et al. explored the computational complexity of solving the *TF* problem which has been shown NP-complete.

3.2 Maximum Trust Routes Identification (MTRI)

PROBLEM 2: /*MTRI*/ Given a task-oriented social network $G' = (V', E')$, E' is a collection of edges consisting of the maximum trust routes.

The aforementioned trust inference mechanisms provide fundamentals for solving the *MTRI* problem. However, one concern is their time complexity. We adopt the simplest multiplication method as our solution for the *MTRI* problem. Recall that trust routes are directed and trust values are vectors. In order to identify the maximum trust routes for a certain social network, we need to map the network to trust realm.
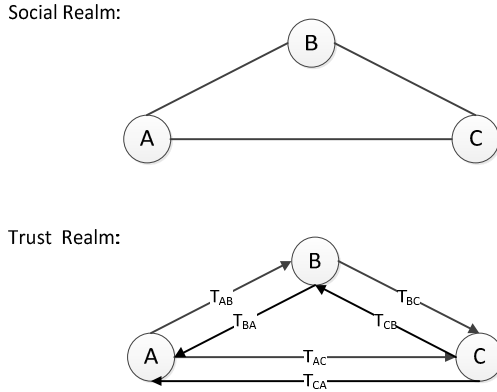
Fig. 1. Social Realm and Trust Realm

Figure 1 depicts a small social network consisting of three individuals Alice (A), Bob (B), and Carol (C). In social realm, A, B, and C are three nodes connected by undirected (normal) edges indicating that they have communications with each other. According to our previous research, trust value can be deduced based on the historical communications [12]. Hence, in a trust realm, the directed edges are weighted with certain trust values to demonstrate the trust relationships. For instance, $|T_{AB}| = |\vec{T}_{AB}|$, where $T_{AB} \in [0,1]$, in which one means 100% trust and 0% means no trust at all.
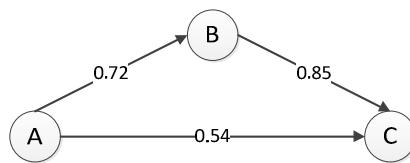
Fig. 2. An Example of Trust Edges

In Figure 2, suppose that we want to identify the maximum trust route, based on the perspective of Alice (node A), and the identified route must contain Carol (node C). To tackle the problem, we first sort out all of the possible routes from A to C. Two routes, $\overrightarrow{R_{A(B)C}}$ and $\overrightarrow{R_{AC}}$, are identified. Second, trust value for each route is inferred by the multiplication method:

$$T_{A(B)C} = T_{AB} * T_{BC} = 0.72 * 0.85 \approx 0.61$$
$$T_{AC} = 0.54$$

Given $(T_{A(B)C} = 0.61) > (T_{AC} = 0.54)$, $\overrightarrow{R_{A(B)C}}$ is then identified as the maximum trust route with removed directions showing in Figure 3. This example also reveals a very important property of trust routes. That is trust routes are not necessary the shortest routes.
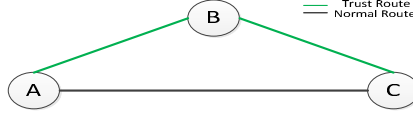
Fig. 3. Maximum Trust Route in Social Realm

Trust-Maximization Algorithm

Algorithm 1. The Trust-Maximization algorithm

Input: $G = (V, E)$ and $\mathcal{T} \subseteq A$
Output: Team $V' \subseteq V$ and $G' = (V', E')$
1: **for** $a_k \in \mathcal{T}$ **do**
2:     $S(a_k) = \{v_i \mid a_k \in v_i\}$
3: $a_{\text{rare}} \leftarrow \text{argmin}|S(a_k)|$
4: **for** $v_i \in S(a_{\text{rare}})$ **do**
5:     **for** $a_j \in \mathcal{T}$ and $a_j \neq a_{\text{rare}}$ **do**
6:         $\vec{\mathcal{R}}(v_i) \leftarrow \overrightarrow{\text{MaxTrustRoute}}(v_i, S(a_j))$
7:     $\text{Tr}(v_i) \leftarrow \sum_{v_i} \text{TrustValue}(\vec{\mathcal{R}}(v_i))$
8: $v_{\text{Tr-max}} \leftarrow \text{argmax}|\text{Tr}(v_i)|$
9: $G' = (V', E') \leftarrow \mathcal{R}(v_{\text{Tr-max}})$

Algorithm 1 is the pseudocode of the Trust-Maximization algorithm. The algorithm is a variation of the RarestFirst algorithm in [23], for the purpose of team formation. In this sense, for each skill, a_k, required by a given task, $\mathcal{T}$, a support group $S(a_k)$ is computed. The support group with the lowest cardinality is denoted by $S(a_{\text{rare}})$. $\overrightarrow{\text{MaxTrustRoute}}(v_i, S(a_j))$ function returns the maximum trust route from v_i to $S(a_j)$, one support group. TrustValue function identifies the trust values of the maximum trust routes. $v_{\text{Tr-max}}$ is one of the nodes belonging to $S(a_{\text{rare}})$ that the routes starts from $v_{\text{Tr-max}}$ totally have the maximum magnitued of trust. Suppose that all routes between any pair of nodes have been pre-identified. We also assume that each route's trust value, on all directions, has been pre-computed. The running time complexity of the Trust-Maximization algorithm is $\mathcal{O}(n^2)$ at the worst case.

PROPOSITION: The output of the Trust-Maximization algorithm, $G' = (V', E')$ consists of maximum trust routes, from the perspective of $S(a_{\text{rare}})$.

PROOF: G' is a task oriented social network that is able to accomplish the given task, $\mathcal{T}$. a_{rare} is a required skill from $\mathcal{T}$ that is possessed by the least number of individuals comparing with any other required skills from $\mathcal{T}$. $S(a_{\text{rare}})$ is a group of individuals

who possess a_{rare}. It is intuitively appropriate to deploy trust routes from the perspective of $S(a_{\text{rare}})$ because of its indispensability. v_i is, at least, one individual from $S(a_{\text{rare}})$. $S(a_j)$ are groups with individuals possessing the task-required skills except a_{rare}. There may exist multiple routes, from v_i to every $S(a_j)$. $\overrightarrow{\text{MaxTrustRoute}}$ only returns the maximum trust route, from v_i to every $S(a_j)$. We then have $\vec{\mathcal{R}}(v_i)$ which is a collection of all the maximum trust routes from v_i to $S(a_j)$. Recall that, for each route, its trust value is located in the domain of $[0,1]$. $\sum_{v_i} \text{TrustValue } (\vec{\mathcal{R}}(v_i))$ is meant to add all of the trust values of the maximum trust routes from v_i, the result will be a maximized trust value of v_i. If v_i is one individual, the output graph is the collection of v_i's maximum trust routes to $S(a_j)$. If v_i is multiple individuals, the output graph is also the collection of the maximum trust routes from the one who possess the highest maximized trust value.   □

References

1. Schonfeld, E.: Facebook Connect + Facebook Ads = A Social Ad Network, `http://techcrunch.com/2009/03/03/` `facebook-connect-facebook-ads-a-social-ad-network/`
2. Hogben, G.: Security Issues in the Future of Social Networking. In: Proceedings of W3C Workshop on the Future of Social Networking, Barcelona, Spain (January 2009)
3. Marsh, S.: Formalising Trust as a Computational Concept, PhD thesis, Department of Mathematics and Computer Science, University of Stirling (1994)
4. Golbeck, J., Hendler, J.: Inferring Trust Relationships in Web-based Social Networks. ACM Transactions on Internet Technology 6(4), 497–529 (2006)
5. Sztompka, P.: Trust: A Sociological Theory. Cambridge University Press, Cambridge (1999)
6. Grandison, T., Sloman, M.: A Survey of Trust in Internet Applications. Communications Surveys and Tutorials 4(4), 2–16 (2000)
7. Mui, L., Mohtashemi, M., Halberstadt, A.: A Computational Model of Trust and Reputation for E-business. In: Proceedings of the 35th International Conference on System Science, Big Island, Hawaii, pp. 280–287 (January 2002)
8. Olmedialla, D., Rana, O., Matthews, B., Nejdl, W.: Security and Trust Issues in Semantic Grids. In: Proceedings of the Dagsthul Seminar, Semantic Grid: The Convergence of Technologies, vol. 05271
9. Prosper, `http://www.prosper.com/`
10. Shen, D., Krumme, C., Lippman, A.: Follow the Profit or the Herd? Exploring Social Effects in Peer-to-Peer Lending. In: Proceedings of The Second International Conference on Social Computing, Minneapolis, MN, USA, pp. 137–144 (August 2010)
11. Brainard, J., Juels, A., Rivest, R., Szydlo, M., Yung, M.: Fourth-Factor Authentication: Somebody You Know. In: Proceedings of the 13th ACM Conference on Computer and communications security, Virginia (October 2006)
12. Zhan, J., Fang, X.: A Computational Trust Framework for Social Computing. In: Proceedings of The Second International Conference on Social Computing, Minneapolis, MN, USA, pp. 264–269 (August 2010)
13. Kamvar, S., Schlosser, M., Garcia-Molina, H.: The Eigen Trust Algorithm for Reputation Management in P2P Networks. In: Proceedings of the 12th International World Wide Web Conference, Budapest, Hungary, pp. 640–651 (May 2003)

14. Richardson, M., Agrawal, R., Domingos, P.: Trust Management for the Semantic Web. In: Proceedings of the 2nd International Semantic Web Conference, Sanibel, Finland, pp. 351–368 (October 2003)
15. Walter, F., Battistion, S., Schweitzer, F.: A Model of A Trust-based Recommendation System. Auton Agent Multi-Agent System 16(1), 57–74 (2007)
16. Liu, G., Wang, Y., Orgun, M.: Trust Inference in Complex Trust-oriented Social Networks. In: Proceedings of the International Conference on Computational Science and Engineering, Vancouver, Canada, pp. 996–1001 (August 2009)
17. Jonsang, A., Gray, E., Kinaterder, M.: Simplification and Analysis of Transitive Trust Networks. Web Intelligence and Agent Systems 4(2), 139–161 (2006)
18. Golbeck, J.: Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering. In: Proceedings of the International Provenance and Annotation Workshop, Chicago, Illinois, USA, pp. 101–108 (May 2006)
19. Ziegler, C., Lausen, G.: Propagation Models for Trust and Distrust in Social Networks. Information Systems Frontiers 7(4-5), 337–358 (2005)
20. Bhuiyan, T., Xu, Y., Josang, A.: Integrating Trust with Public Reputation in Location-based Social Networks for Recommendation Making. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, Australia, pp. 107–110 (December 2008)
21. Zuo, Y., Hu, W., Keefe, T.: Trust Computing for Social Networking. In: Proceedings of the 6th International Conference on Information Technology New Generations, Las Vegas, Nevada, USA, pp. 1534–1539 (April 2009)
22. Kuter, U., Golbeck, J.: SUNNY: A New Algorithm for Trust Inference in Social Networks Using Probabilistic Confidence Models. In: Proceedings of the 22nd AAAI Conference on Artificial Intelligence, British Columbia, Canada, pp. 1377–1382 (July 2007)
23. Lappas, T., Liu, K., Terzi, E.: Finding a Team of Experts in Social Networks. In: Proceedings of ACM International Conference on Knowledge Discovery and Data Mining (KDD 2009), Paris, France, pp. 467–475 (June-July 2009)
24. Fitzpatrick, E., Askin, R.: Forming Effective Worker Teams with Multi-Funcitional Skill Requirements. Computers and Industrial Engineering 48(3), 593–608 (2005)
25. Chen, S., Lin, L.: Modeling Team Member Characteristics for the Formation of a Multifunctional Team in Concurrent Engineering. IEEE Transactions on Engineering Management 51(2), 111–124 (2004)

Towards a Computational Analysis of Status and Leadership Styles on FDA Panels

David A. Broniatowski and Christopher L. Magee

Engineering Systems Division, Massachusetts Institute of Technology,
E38-450, 77 Massachusetts Ave., Cambridge, MA 02141, USA
{david,cmagee}@mit.edu

Abstract. Decisions by committees of technical experts are increasingly impacting society. These decision-makers are typically embedded within a web of social relations. Taken as a whole, these relations define an implicit social structure which can influence the decision outcome. Aspects of this structure are founded on interpersonal affinity between parties to the negotiation, on assigned roles, and on the recognition of status characteristics, such as relevant domain expertise. This paper build upon a methodology aimed at extracting an explicit representation of such social structures using meeting transcripts as a data source. Whereas earlier results demonstrated that the method presented here can identify groups of decision-makers with a contextual affinity (i.e., membership in a given medical specialty or voting clique), we now can extract meaningful status hierarchies, and can identify differing facilitation styles among committee chairs. Use of this method is demonstrated on the transcripts of U.S. Food and Drug Administration (FDA) advisory panel meeting transcripts; nevertheless, the approach presented here is extensible to other domains and requires only a meeting transcript as input.

Keywords: Linguistic analysis, Bayesian inference, committee decision-making.

1 Introduction

This paper presents a computational methodology designed to study decision-making by small groups of technical experts. Previously, we showed how a computational approach could be used to generate social networks from meeting transcripts [1]. The advantage of using a computational approach over more traditional social science content analysis methods include the repeatability and consistency of the approach, as well as time- and labor-saving advances in automation. Quinn et al. [2] provide a compelling justification for the adoption of a computational analysis by scholars of agenda-setting in political science. We feel that this argument may equally be extended to other domains of interest to social scientists and to social computing generally.

2 Data Source

The empirical analysis mentioned above requires data in the form of committee meeting transcripts. The ideal data source must involve a negotiation (e.g., analysis or

evaluation of a technological artifact); include participation of representatives from multiple groups (e.g., multiple experts from different fields or areas of specialization); contain a set of expressed preferences per meeting (such as a voting record); and multiple meetings must be available, so as to enable statistical significance. These requirements are met by the Food and Drug Administration's medical device advisory panels – bodies that review the most uncertain medical devices prior to their exposure to the American market. A device's approval and future market diffusion often rests upon the panel's assessment of the device's safety. These panels are aimed at producing a recommendation, informed by the expertise and knowledge of panel members, which can supplement the FDA's "in-house" decision process. Multiple experts are consulted so that the group decision's efficacy can take advantage of many different institutional perspectives. Panel members' values and institutional contexts may differ, leading to different readings of the evidence, and therefore different recommendations. In health care, Gelijns et al. [3] note that strictly evidence-based decisions are often not possible due to different interpretations of data, uncertain patterns of technological change, and differing value systems among experts.

3   Methodology

We use Bayesian topic models to determine whether actors within a committee meeting are using similar terminology to discuss the common problem to be solved. The choice of Bayesian topic models is driven by a desire to make the assumptions underlying this analysis minimal and transparent while maintaining an acceptable level of resolution of patterns inferred from the transcript data.

   We assume that each speaker possesses a signature in his or her word choice that can inform the selection of topics. We therefore use a variant of the Author-Topic (AT) Model [4], which creates probabilistic pressure to assign each author to a specific topic. Shared topics are therefore more likely to represent common jargon. Briefly, a social network is created by first generating a term-document matrix from a meeting transcript. Words are stemmed using Porter's Snowball algorithm [5], and function-words are removed. The AT Model is then used to assign each word token to a topic. Following the procedure outlined in [1], each utterance by a committee voting member is assigned to both that individual and to a "false author". Practically, this has the effect of removing procedural words from the analysis, and focusing instead on a speaker's unique language. A collapsed Gibbs sampler, using a variant of the algorithm outlined in [6], is used to generate samples from the AT Model's posterior distribution. The number of topics is chosen independently for each transcript as follows: 35 AT Models are fit to the transcript for $T = 1 \ldots 35$ topics. An upper limit of 35 topics was chosen based upon empirical observation (see Broniatowski and Magee [1] for more details). For each model, 20 samples are generated from one randomly initialized Markov chain after an empirically-determined burn-in of 1000 iterations. We find the smallest value, t_0, such that the 95th percentile of all samples for all larger values of T is greater than the 5th percentile of t_0. We set the value of $T = t_0 + 1$ so as to ensure that the model chosen is beyond the knee in the curve.

Once the number of topics has been chosen, a T-topic AT Model is again fit to the transcript. Following the procedure used by Griffiths and Steyvers [6], ten samples are taken from 20 randomly initialized Markov chains, such that there are 200 samples in total. These form the basis for all subsequent analysis. A network is generated for each sample by examining the joint probability distribution for each author-pair in that sample:

$$P(X_1 \cap X_2) = \frac{\sum_{i}^{T} P(Z = z_i \mid X_1) * P(Z = z_i \mid X_2)}{\sum_{i,j}^{T,T} P(Z = z_i \mid X_1) * P(Z = z_j \mid X_2)}$$

If the joint probability distribution exceeds 1/T, then we say that this author-pair is *linked* in this sample. Using the Bonferroni criterion for a binomial distribution with family-wise error rate of 5%, and assuming ~15 voting committee members, two authors are connected by an edge in the transcript-specific social-network, Δ, if they are linked more than 125 times out of 200 samples. Full details are given by Broniatowski and Magee [1], where we showed that this approach can be used to identify groups of speakers who share similar identities (e.g., members of the same medical specialty or voting clique).

We extend this method such that directed edges are overlaid on top of the social network. Directionality was determined by examining the order in which each speaker discussed a particular topic. For example, if two speakers, X and Y, share a topic and X consistently speaks after Y does, then we say that Y *leads* X on this topic. Conversely, X *lags* Y on this topic. These directions are then averaged over all topics in proportion to the amount that each speaker discusses each topic. In particular, consider a sample, s, from the posterior distribution of the AT model. Within this sample, choose a pair of speakers, x_1 and x_2, and a topic z. Given that utterances are temporally ordered, this defines two separate time-series. These time series can be used to generate the *topic-specific cross correlation* for speakers x_1 and x_2, in topic z:

$$(f_{i,z}^{s} * f_{j,z}^{s})[\delta] = \sum_{d=-\infty}^{\infty} f_{i,z}^{s}{}^{*}[d] f_{j,z}^{s}[\delta + d]$$

where $f_{i,t}^{s}$ (d) is the number of words spoken by author i and assigned to topic z in document d, in sample s. For each sample, s, from the AT Model's posterior distribution, we examine the cross-correlation function for each author pair, $\{x_i, x_j\}$, in topic z. Let there be k peaks in the cross-correlation function, such that

$$m_k = \arg\max_{\delta} (f_i^{t} * f_j^{t})[\delta]$$

For each peak, if $m_k > 0$, we say that author i *lags* author j in topic z, at point m_k (i.e., $l_{i,j,z,m_k}^{s} = 1$). Similarly, we say that author i *leads* author j in topic z at point m_k (i.e., $l_{i,j,z,m_k}^{s} = -1$) if $m_k < 0$. Otherwise, $l_{i,j,z,m_k}^{s} = 0$. For each sample, s, we define the *polarity* of authors i and j in topic z to be the median of the $l_{i,j,zt}^{s}$.

$$p_{i,j,t}^{s} = median(l_{i,j,z}^{s})$$

If most of the peaks in the cross-correlation function are greater than zero, then the polarity = 1; if most of the peaks are less than zero, then the polarity = -1; otherwise, the polarity = 0.

We are particularly interested in the topic polarities for author-pairs who are linked in the graph methodology outlined above – i.e., where $\Delta_{i,j} = \Delta_{j,i} = 1$. Using the polarity values defined above, we are interested in determining directionality in Δ. For each sample, s, we define the *direction* of edge $e_{i,j}$ in sample s as:

$$d^s(e_{i,j}) = \sum_{t=1}^{T}(p^s_{i,j,t} * P^s(Z = z_i \mid x_i) * P^s(Z = z_i \mid x_j))$$

This expression weights each topic polarity by its importance in the joint probability distribution between x_i and x_j, and is constrained to be between -1 and 1 by definition. The set of 200 $d^s(e_{i,j})$ defines a distribution such that the *net edge direction,* $d(e_{i,j})$ is determined by partition of the unit interval into three segments of equal probability. In particular, this is done by examining the proportion of the $d^s(e_{i,j})$ that are greater than 0. If more than 66% of the $d^s(e_{i,j}) > 0$ then $d(e_{i,j}) = 1$ (the arrow points from j to i). If less than 33% of $d^s(e_{i,j}) > 0$ then $d(e_{i,j}) = -1$ (the arrow points from i to j). Otherwise, $d(e_{i,j}) = 0$ (the arrow is bidirectional). The result is a directed graph, an example of which is shown in Figure 1.

It is interesting to note that these graphs may change their topology on the basis of which nodes are included. In particular, inclusion of the committee chair may or may not "flatten" the graph by connecting the sink nodes (i.e., nodes on the bottom of the graph) to higher nodes. In other meetings, the chair may act to connect otherwise disparate groups of voters or serve to pass information from one set of nodes to another.
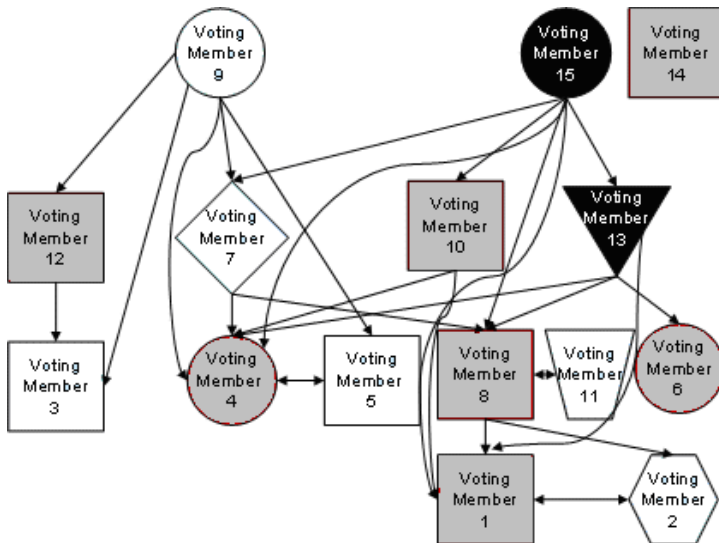
Fig. 1. Directed Graph representation of meeting held on June 23, 2005. Hierarchy metric = 0.35 [10]. Node shape represents medical specialty and note color represents vote (white is approval; grey is non-approval; black is abstention).
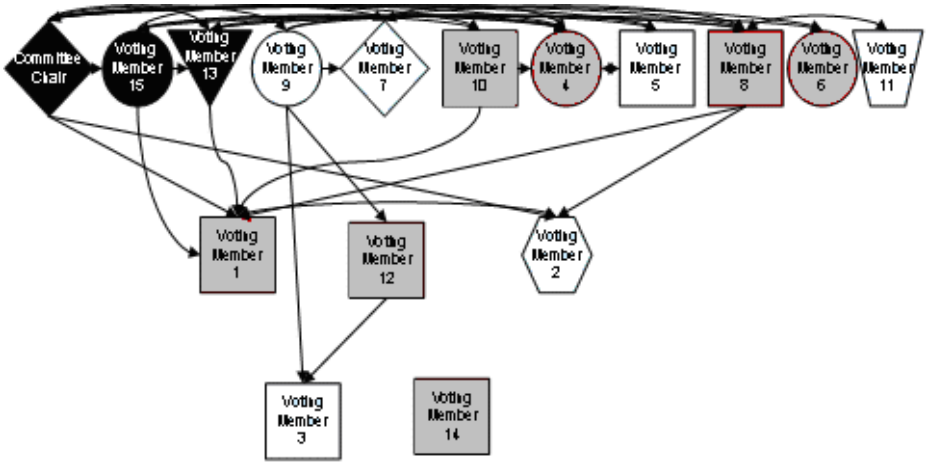
Fig. 2. Directed Graph representation of meeting held on June 23, 2005, with the committee chair included. Hierarchy metric = 0.78. Node shape represents medical specialty and note color represents vote (white is approval; grey is non-approval; black is abstention).
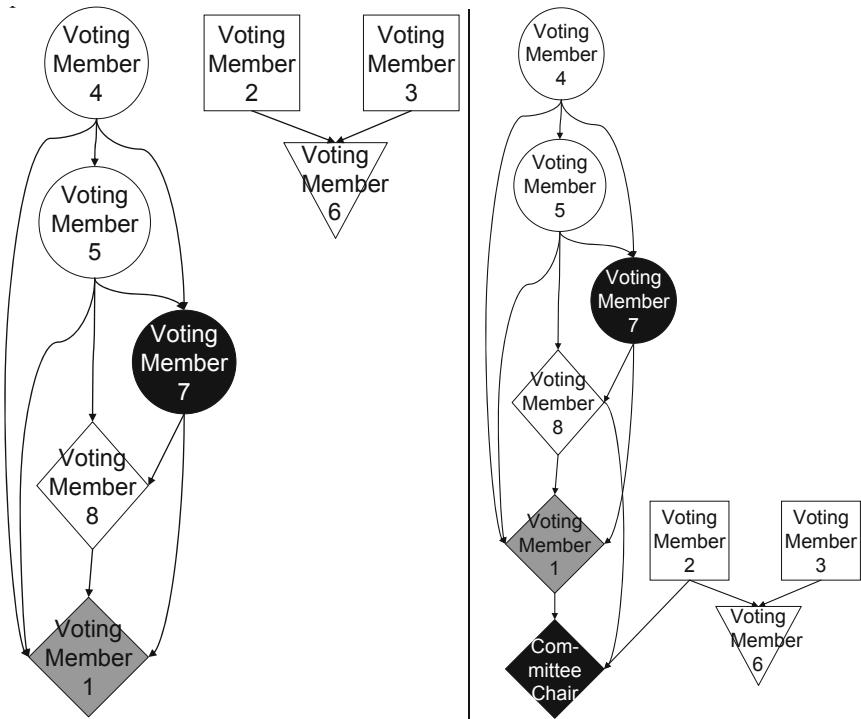
Fig. 3. Directed Graph representation of meeting held on October 27, 1998 without (left) and with (right) the committee chair. Hierarchy metric = 0 for both cases. Node shape represents medical specialty and note color represents vote (white is approval; grey is non-approval; black is abstention).

We may quantify the impact that the committee chair has upon the meeting by determining, for each graph, which proportion of edges is part of a cycle. This is a metric of the hierarchy in the graph [10]. The difference in this metric between graphs with and without the chair therefore quantifies impact of the chair on the meeting. We display this metric for the graph without the chair (e.g., Figure 2) and with the chair (e.g., Figure 3).

The difference in this metric between graphs with and without the chair therefore quantifies the impact of the chair on the meeting. For the meeting held on June 23, 2005, this value is 0.78-0.35 = 0.43. This suggests that the chair is significantly changing the topology of the meeting structure – in particular, it seems that the chair is connecting members at the "bottom" of the graph to those at the "top". Other meetings display different behavior by the chair. Consider the meeting held on October 27, 1998 (in Figure 3).

In this meeting, the committee chair served to connect the two disparate clusters on the panel. Nevertheless, the chair is not creating any new cycles on the graph. This is reflected in the fact that the hierarchy metric for both of these meetings is equal to 0.
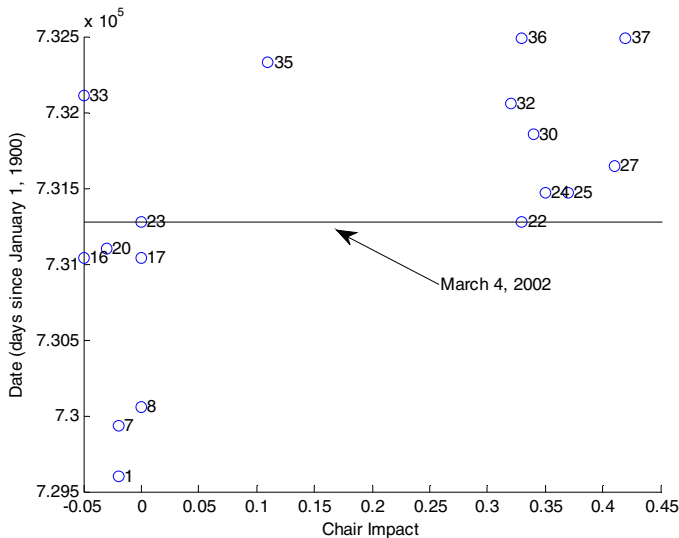
Fig. 4. Impact of chair vs. meeting date for each of the 17 meetings in which there was a voting minority. Each meeting is labeled by its corresponding ID from the sample of all 37. Meetings with a chair impact greater than zero are significantly more likely to have occurred later than are meetings with a chair impact of zero or less (p=0.005, using a non-parametric Mann-Whitney U test).

The chair's role in "flattening" the structure of a given meeting's graph could suggest a particular facilitation strategy, wherein the committee chair tries to elicit the opinions of voting members who speak later, and might otherwise be less influential. When the chair does not act to significantly change the graph structure, the chair may be taking on the role of a synthesizer – gathering the opinions of the other voters to answer FDA questions, but not causing any of the other members to repeat what has

already been said. This suggests that there may be at least two different strategies that committee chairs could use during a meeting. We note that the impact of the chair seems to be increase markedly around March 2002 (see Figure 4). Furthermore, meetings with a chair impact greater than zero are significantly more likely to have occurred later than are meetings with a chair impact of zero or less (p=0.005, using a non-parametric Mann-Whitney U test).

4   Conclusions

The marked change in strategy across multiple chairs has many potential explanations, e.g., a change in FDA policy regarding committee operations, e.g., a change in conflict of interest reporting procedures, but there is no obvious connection between these two events. Alternatively, the types of devices that came to the Circulatory Systems Devices Panel seem to have changed around this time; perhaps there was an increase in the difficulty of the devices that the panel reviewed (e.g., concurrent with the entrance on the market of drug-eluting stents, Left Ventricular Assist Devices and other potentially risky devices). Finally, we might hypothesize some combination of these ideas – that a change in FDA policy might have some way impacted upon chair style, and that this change in policy might have been driven by changing market conditions. Testing these hypotheses requires looking across multiple panels and committees, which we leave to future work. Nevertheless, if these findings were cross-validated with other metrics of committee chair efficacy, the method presented here suggests the possibility for a fruitful analysis of leadership dynamics.

References

1. Broniatowski, D., Magee, C.L.: Analysis of Social Dynamics on FDA Panels using Social Networks Extracted from Meeting Transcripts. In: 2nd IEEE Conference on Social Computing, 2nd Symposium on Social Intelligence and Networking (2010)
2. Quinn, K.M., et al.: How to Analyze Political Attention with Minimal Assumptions and Costs. American Journal of Political Science 54, 209–228 (2010)
3. Gelijns, A.C., et al.: Evidence, Politics, And Technological Change. Health Affairs 24, 29–40 (2005)
4. Rosen-Zvi, M., et al.: The author-topic model for authors and documents. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 487–494 (2004)
5. Porter, M.F.: Snowball: A language for stemming algorithms (October 2001)
6. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America 101, 5228–5235 (2004)
7. Luo, J., et al.: Measuring and Understanding Hierarchy as an Architectural Element in Industry Sectors. SSRN eLibrary (2009),
   `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1421439`

Tracking Communities in Dynamic Social Networks

Kevin S. Xu[1], Mark Kliger[2], and Alfred O. Hero III[1]

[1] EECS Department, University of Michigan, 1301 Beal Avenue
Ann Arbor, MI 48109-2122 USA
{xukevin,hero}@umich.edu
[2] Medasense Biometrics Ltd., P.O. Box 633, Ofakim, 87516 Israel
mark@medasense.com

Abstract. The study of communities in social networks has attracted considerable interest from many disciplines. Most studies have focused on static networks, and in doing so, have neglected the temporal dynamics of the networks and communities. This paper considers the problem of tracking communities over time in dynamic social networks. We propose a method for community tracking using an adaptive evolutionary clustering framework. We apply the method to reveal the temporal evolution of communities in two real data sets. In addition, we obtain a statistic that can be used for identifying change points in the network.

Keywords: dynamic, social network, community, tracking, clustering.

1 Introduction

Traditionally, social network data have been collected through means such as surveys or human observation. Such data provide a view of a social network as a static entity over time. However, most social networks are dynamic structures that evolve over time. There has been recent interest in analyzing the temporal dynamics of social networks, enabled by the collection of dynamic social network data by electronic means such as cell phones, email, blogs, etc. [2, 6, 10].

A fundamental problem in the analysis of social networks is the detection of communities [7]. A community is often defined as a group of network members with stronger ties to members within the group than to members outside of the group. Previous studies on the community structure of social networks have typically focused on static networks. In doing so, the temporal dynamics of the networks and communities have been neglected. The natural extension of community detection to dynamic networks is community tracking, which makes it possible to observe how communities grow, shrink, merge, or split with time.

In this paper, we propose a method for tracking communities in dynamic social networks. The proposed method makes use of an evolutionary clustering framework that detects communities at each time step using an adaptively weighted combination of current and historical data. The result is a set of communities at each time step, which are then matched with communities at other time steps so

that communities can be tracked over time. We apply the proposed method to reveal the temporal evolution of communities in two real data sets. The detected communities are more accurate and easier to interpret than those detected by traditional approaches. We also obtain a statistic that appears to be a good identifier of change points in the network.

2 Methodology

The objective of this study is to track the evolution of communities over time in dynamic social networks. We represent a social network by an undirected weighted graph, where the nodes of the graph represent the members of the network, and the edge weights represent the strengths of social ties between members. The edge weights could be obtained by observations of direct inter-action between nodes, such as physical proximity, or inferred by similarities between behavior patterns of nodes. We represent a dynamic social network by a sequence of time snapshots, where the snapshot at time step t is represented by $W^t = \left[w_{ij}^t\right]$, the matrix of edge weights at time t. W^t is commonly referred to as the adjacency matrix of the network snapshot.

The problem of detecting communities in static networks has been studied by researchers from a wide range of disciplines. Many community detection methods originated from methods of graph partitioning and data clustering. Popular community detection methods include modularity maximization [7] and spectral clustering [12, 14]. In this paper, we address the extension of community detection to dynamic networks, which we call *community tracking*. We propose to perform community tracking using an adaptive evolutionary clustering framework, which we now introduce.

2.1 Adaptive Evolutionary Clustering

Evolutionary clustering is an emerging research area dealing with clustering dynamic data. First we note that is possible to cluster dynamic data simply by performing ordinary clustering at each time step using the most recent data. However this approach is extremely sensitive to noise and produces clustering results that are inconsistent with results from adjacent time steps. Evolutionary clustering combines data from multiple time steps to compute the clustering result at a single time step, which allows clustering results to vary smoothly over time. Xu et al. [13] recently proposed an evolutionary clustering framework that adaptively estimates the optimal weighted combination of current and past data to minimize a mean-squared error (MSE) criterion. We describe the framework in the following.

Define a smoothed adjacency matrix at time t by

$$\bar{W}^t = \alpha^t \bar{W}^{t-1} + \left(1 - \alpha^t\right) W^t \tag{1}$$

for $t \geq 1$ and by $\bar{W}^0 = W^0$. α^t can be interpreted as a *forgetting factor* that controls the amount of weight given to past data. We treat each network snapshot W^t as a realization from a nonstationary random process and define the

expected adjacency matrix $\Psi^t = \left[\psi_{ij}^t\right] = \mathrm{E}\left[W^t\right]$. If we had access to the expected adjacency matrix Ψ^t, we would expect to see improved clustering results by clustering on Ψ^t rather than the noisy realization W^t. However, Ψ^t is unknown in real applications so the goal is to estimate it as accurately as possible. If we take the estimate to be the convex combination defined in (1), it was shown in [13] that the optimal choice of α^t that minimizes the MSE in terms of the Frobenius norm $\mathrm{E}\left[\|\bar{W}^t - \Psi^t\|_F^2\right]$ is given by

$$
\left(\alpha^t\right)^* = \frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{var}\left(w_{ij}^t\right)}{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(\bar{w}_{ij}^{t-1} - \psi_{ij}^t\right)^2 + \mathrm{var}\left(w_{ij}^t\right)\right\}}, \tag{2}
$$

where n denotes the number of nodes in the network. In a real application, ψ_{ij}^t and $\mathrm{var}\left(w_{ij}^t\right)$ are unknown so $\left(\alpha^t\right)^*$ cannot be computed. However, it can be approximated by replacing the unknown means and variances with sample means and variances. The communities at time t can then be extracted by performing ordinary community detection on the smoothed adjacency matrix $\bar{W}^t$.

Any algorithm for ordinary community detection can be used with the adaptive evolutionary clustering framework. In this paper, we use Yu and Shi's normalized cut spectral clustering algorithm [14]. It finds a near global-optimal separation of the nodes into k communities, where k is specified by the user. The algorithm involves computing the eigenvectors corresponding to the k largest eigenvalues of a normalized version of $\bar{W}^t$, then discretizing the eigenvectors so that each node is assigned to a single community. We refer readers to [13] for additional details on the adaptive evolutionary spectral clustering algorithm.

2.2 Tracking Communities over Time

There are several additional issues that also need to be addressed in order to track communities over time. The communities detected at adjacent time steps need to be matched so that we can observe how any particular community evolves over time. This can be achieved by finding an optimal permutation of the communities at time t to maximize agreement with those at time $t - 1$. If the number of communities at time t is small, it is possible to exhaustively search through all such permutations. This is, however, impractical for many applications. We employ the following heuristic: match the two communities at time t and $t - 1$ with the largest number of nodes in agreement, remove these communities from consideration, match the two communities with the second largest number of nodes in agreement, remove them from consideration, and so on until all communities have been exhausted.

Another issue is the selection of the number of communities k at each time. Since the evolutionary clustering framework involves simply taking convex combinations of adjacency matrices, any heuristic for choosing the number of communities in ordinary spectral clustering can also be used in evolutionary spectral

clustering by applying it to $\bar{W}^t$ instead of W^t. In this paper we use the eigengap heuristic [12] of selecting the number of communities k such that the gap between the kth and $(k+1)$th largest eigenvalues of the normalized adjacency matrix is large.

Finally, there is the issue of nodes entering or leaving the network over time. We deal with these nodes in the following manner. Nodes that leave the network between times $t-1$ and t can simply be removed from $\bar{W}^{t-1}$ in (1). Nodes that enter the network at time t have no corresponding rows and columns in $\bar{W}^{t-1}$. Hence, these new nodes can be naturally handled by adding rows and columns to $\bar{W}^t$ after performing the smoothing operation in (1). In this way, the new nodes have no influence on the update of the forgetting factor α^t yet contribute to the community structure through $\bar{W}^t$.

3   Experiments

3.1   Reality Mining

Data Description. The MIT Reality Mining data set [2] was collected as part of an experiment on inferring social networks by monitoring cell phone usage rather than by traditional means such as surveys. The data was collected by recording cell phone activity of 94 students and staff at MIT for over a year. Each phone recorded the Media Access Control (MAC) addresses of nearby Bluetooth devices at five-minute intervals. Using this device proximity data, we construct a sequence of adjacency matrices where the edge weight between two participants corresponds to the number of intervals where they were in close physical proximity within a time step. We divide the data into time steps of one week, resulting in 46 time steps between August 2004 and June 2005.

In this data set, we have partial ground truth to compare against. From the MIT academic calendar [5], we know the dates of important events such as the beginning and end of school terms. In addition, we know that 26 of the participants were incoming students at the university's business school, while the rest were colleagues working in the same building. Thus we would expect the detected communities to match the participant affiliations, at least during the school terms when students are taking classes.

Observations. We make several interesting observations about the community structure of this data set and its evolution over time. The importance of temporal smoothing for tracking communities can be seen in Fig. 1. On the left is the heat map of community membership over time when the proposed method is used. On the right is the same heat map when ordinary community detection at each time is used, which is equivalent to setting $\alpha^t = 0$ in (1). Notice that two clear communities appear in the heat map to the left, where the proposed method is used. The participants above the black line correspond to the colleagues working in the same building, while those below the black line correspond to the incoming business school students. On the heat map to the right, corresponding to ordinary
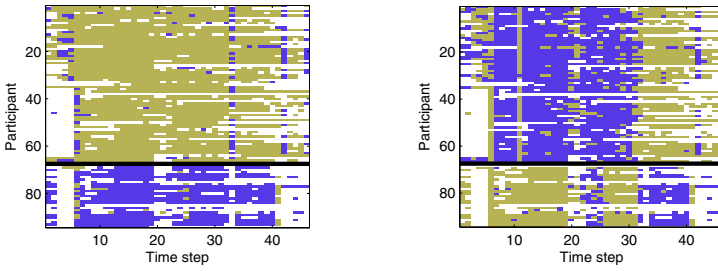
Fig. 1. Heat maps of community structure over time for the proposed method (left) and ordinary community detection (right) in the Reality Mining experiment
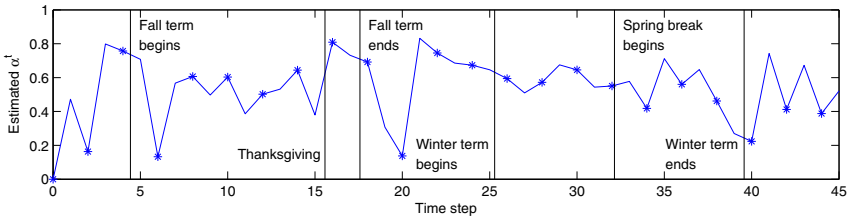
Fig. 2. Estimated forgetting factor α^t by time step in the Reality Mining experiment

community detection, the community memberships fluctuate highly over time. Thus we can see that tracking communities by the proposed method results in more stable and accurately identified communities.

The estimated forgetting factor α^t at each time step is plotted in Fig. 2. Six important dates are labeled on the plot. Notice that the estimated forgetting factor drops several times, suggesting that the structure of the proximity network changes, around these dates. This is a reasonable result because the proximity network should be different when students are not in school compared to when they are in school. Thus α^t also appears to be a good identifier of change points in the network.

3.2   Project Honey Pot

Data Description. Project Honey Pot [8] is an ongoing project targeted at identifying spammers. It consists of a distributed network of decoy web pages with trap email addresses, which are collected by automated email address harvesters. Both the decoy web pages and the email addresses are monitored, providing us with information about the harvester and email server used for each spam email received at a trap address. A previous study on the Project Honey Pot data [9] found that harvesting is typically done in a centralized manner. Thus harvesters are likely to be associated with spammers, and in this study we assume that the harvesters monitored by Project Honey Pot are indeed representative of spammers. This allows us to associate each spam email with a spammer so that we can track communities of spammers.
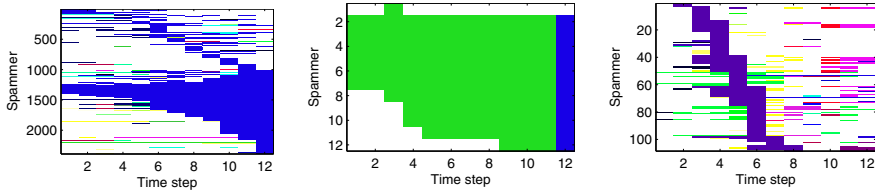
Fig. 3. Temporal evolution of the giant community (left), a persistent community (middle), and a "staircase" community (right) in the Project Honey Pot experiment

Unlike in the previous experiment, we cannot observe direct interactions between spammers. The interactions must be inferred through indirect observations. We take the edge weight between two spammers i and j to be the total number of emails sent by i and j through shared email servers, normalized by the product of the number of email addresses collected by i and by j. Since servers act as resources for spammers to distribute emails, the edge weight is a measure of the amount of resources shared between two spammers. We divide the data set into time steps of one month and consider the period from January 2006 to December 2006. The number of trap email addresses monitored by Project Honey Pot grows over time, so there is a large influx of new spammers being monitored at each time step. Some spammers also leave the network over time.

Observations. In this data set, we do not have ground truth for validation so the experiment is of an exploratory nature. At each time step, there are over 100 active communities, so rather than attempting to visualize all of the communities, as in Fig. 1, we instead try to visualize the evolution of individual communities over time. We discover several interesting evolution patterns, shown in Fig. 3. On the left, there is a giant community that continually grows over time as more and more spammers enter the network. The appearance of this giant community is quite common in large networks, where a core-periphery structure is typically observed [4]. In the middle, we illustrate a community that is persistent over time. Notice that no spammers change community until time step 12, when they all get absorbed into the giant community.

Perhaps the most interesting type of community we observe is pictured on the right. We call this a "staircase" community due to the shape of the heat map. Notice that at each time step, many new spammers join the community while some of the existing spammers become inactive or leave the community. This suggests that either the members of the community are continually changing or that members assume multiple identities and are using different identities at different times. Since spamming is an illegal activity in many countries, the latter explanation is perhaps more likely because it makes spammers more difficult to track due to the multiple identities. Using the proposed method, it appears that we can indeed track these types of spammers despite their attempts to hide their identities.

4 Related Work

There have been several other recent works on the problem of tracking communities in dynamic social networks. [11] proposed to identify communities by graph coloring; however, their framework assumes that the observed network at each time step is a disjoint union of cliques, whereas we target the more general case where the observed network can be an arbitrary graph. [3] proposed a method for tracking the evolution of communities that applies to the general case of arbitrary graphs. The method involves first performing ordinary community detection on time snapshots of the network by maximizing modularity. A graph of communities detected at each time step is then created, and meta-communities of communities are detected in this graph to match communities over time. The main drawback of this approach is that no temporal smoothing is incorporated, so the detected communities are likely to be unstable.

Other algorithms for evolutionary clustering have also been proposed. Relevant algorithms for community tracking include [6] and [1], which extend modularity maximization and spectral clustering, respectively, to dynamic data. [10] proposed an evolutionary spectral clustering algorithm for dynamic multi-mode networks, which have different classes of nodes and interactions. Such an algorithm is particularly interesting for data where both direct and indirect interations can be observed. However, one shortcoming in these algorithms is that they require the user to determine to choose the values for parameters that control how smoothly the communities evolve over time. There are generally no guidelines on how these parameters can be chosen in an optimal manner.

5 Conclusion

In this paper, we introduced a method for tracking communities in dynamic social networks by adaptive evolutionary clustering. The method incorporated temporal smoothing to stabilize the variation of communities over time. We applied the method to two real data sets and found good agreement between our results and ground truth, when it was available. We also obtained a statistic that can be used for identifying change points. Finally, we were able to track communities where the members were continually changing or perhaps assuming multiple identities, which suggests that the proposed method may be a valuable tool for tracking communities in networks of illegal activity.

The experiments highlighted several challenges that temporal tracking of communities presents in addition to the challenges present in static community detection. One major challenge is in the validation of communities, both with and without ground truth information. Another major challenge is the selection of the number of communities at each time step. A poor choice for the number of communities may create the appearance of communities merging or splitting when there is no actual change occurring. This remains an open problem even in the case of static networks. The availability of multiple network snapshots may actually simplify this problem since one would expect that the number of communities, much like the community memberships, should evolve smoothly

over time. Hence, the development of methods for selecting the number of communities in dynamic networks is an interesting area of future research.

References

1. Chi, Y., Song, X., Zhou, D., Hino, K., Tseng, B.L.: Evolutionary spectral clustering by incorporating temporal smoothness. In: Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2007)
2. Eagle, N., Pentland, A., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 106(36), 15274–15278 (2009)
3. Falkowski, T., Bartelheimer, J., Spiliopoulou, M.: Mining and visualizing the evolution of subgroups in social networks. In: Proc. IEEE/WIC/ACM International Conference on Web Intelligence (2006)
4. Leskovec, J., Lang, K.J., Dasgupta, A., Mahoney, M.W.: Statistical properties of community structure in large social and information networks. In: Proc. 17th International Conference on the World Wide Web (2008)
5. MIT academic calendar 2004-2005,
   http://web.mit.edu/registrar/www/calendar0405.html
6. Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.P.: Community structure in time-dependent, multiscale, and multiplex networks. Science 328(5980), 876–878 (2010)
7. Newman, M.E.J.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103(23), 8577–8582 (2006)
8. Project Honey Pot, http://www.projecthoneypot.org
9. Prince, M., Dahl, B., Holloway, L., Keller, A., Langheinrich, E.: Understanding how spammers steal your e-mail address: An analysis of the first six months of data from Project Honey Pot. In: Proc. 2nd Conference on Email and Anti-Spam (2005)
10. Tang, L., Liu, H., Zhang, J., Nazeri, Z.: Community evolution in dynamic multimode networks. In: Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008)
11. Tantipathananandh, C., Berger-Wolf, T., Kempe, D.: A framework for community identification in dynamic social networks. In: Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2007)
12. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17(4), 395–416 (2007)
13. Xu, K.S., Kliger, M., Hero III, A.O.: Evolutionary spectral clustering with adaptive forgetting factor. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (2010)
14. Yu, S.X., Shi, J.: Multiclass spectral clustering. In: Proc. 9th IEEE International Conference on Computer Vision (2003)

Bayesian Networks for Social Modeling

Paul Whitney, Amanda White, Stephen Walsh,
Angela Dalton, and Alan Brothers

Pacific Northwest National Laboratory, Richland, WA USA 99354
{paul.whitney,amanda.white,stephen.walsh,
angela.dalton,alan.brothers}@pnl.gov

Abstract. This paper describes a body of work developed over the past five years. The work addresses the use of Bayesian network (BN) models for representing and predicting social/organizational behaviors. The topics covered include model construction, validation, and use. These topics show the bulk of the lifetime of such model, beginning with construction, moving to validation and other aspects of model "critiquing", and finally demonstrating how the modeling approach might be used to inform policy analysis. The primary benefits of using a well-developed computational, mathematical, and statistical modeling structure, such as BN, are 1) there are significant computational, theoretical and capability bases on which to build 2) the ability to empirically critique the model, and potentially evaluate competing models for a social/behavioral phenomenon.

Keywords: Social modeling, calibration, validation, diagnostics, expert elicitation, evidence assessment.

1 Introduction

Computational modeling of human behavior (CMHB) is a rapidly growing enterprise. Drivers for this enterprise are both practical (prediction, understanding, and policy) and theoretical. We describe a collection of work in CMHB that is based on using Bayesian networks (BN) models [1] as the underlying computational engine.

Computational approaches for CMHB include systems dynamics (SD), agent-based modeling (ABM), discrete event simulations (DES), and statistical methods. These modeling methods are used in dynamic settings. BN models are arguably best suited for semi-static phenomena or assessing over some time interval. Accordingly, we view BN and the dynamic approaches as complimentary.

Regression analyses and related statistical methods are frequently used approaches for developing models. Relative to statistical approaches, the primary benefit of using BN-based modeling approaches is that models can be constructed in advance of data collection, and for phenomena for which data are or will not be available.

The remainder of the paper reflects the life cycle of a BN model. The next section describes the construction of a BN model. Then we present technologies for calibrating the parameters and validating the structure of a BN model *given empirical data*. Finally, we illustrate how evidence can be attached to a BN through a notional example.

2   Constructing Bayesian Network Models

This section focuses on two fundamental aspects of model construction. The first is the methodology used to construct a model with information obtained from technical literature or direct discussions with domain experts. The second aspect is a methodology used to infer the model parameters based on domain experts' feedback elicited from pair-wise scenario comparisons.

2.1   Model Construction

We have executed the following approach to arriving at a BN model structure for a human or organization's behavior. The approach takes as raw input materials from a social science domain. We execute the following steps:

- Key concepts influencing the phenomena of interest are identified
- Relationships among these concepts, as suggested by the literature, are identified
- A BN is constructed, based on the concepts identified, that is consistent with the relationships among the concepts.

While the three-step outline is simple, the execution can be challenging to various degrees, depending on the source literature and complexity of the concept relationships within that literature.

A source literature example that leads rapidly to BN models is political radicalization, as described in [2]. This paper provides a list of mechanisms that can lead to radicalization at various scales (individual, organization, and "mass"). Each of these in turn has influencing factors. One of the mechanisms involves the relationships – both connectedness and radicalization levels – between an individual and a group.
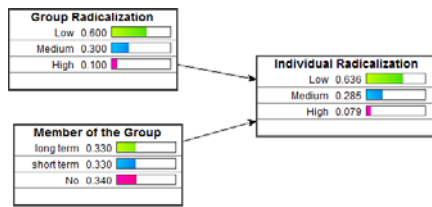
Fig. 1. Bayesian network model fragment representing one aspect of radicalization

Some of the *concepts* and their relationships within this source document include:

- If the individual is a *member of a group*, then that *individual's radicalization* levels will be similar to the *group's*
- If the individual is a *recent* member of the group, then their *radicalization* characteristics will be somewhere between the general population and that *group*
- If the individual is not a *member of the group*, their individual *radicalization* characteristics will be presumed to be similar to the *larger population*

A model fragment consistent with these relationships is shown in the Figure 1.

Both model structure and parameter values are needed for a BN. In this case, the parameters were set so that the above three relationships would hold. However, there are multiple BN structures and parameter values that could satisfy the above three relationships. The next subsection shows a method for mapping experts' assessments of scenarios to parameter values. The subsequent section provides empirical methods for examining both the structure and model parameter values.

2.2 Experts' Assessments Calibration

Experts' assessments play two prominent roles in BN model development: specifying the network structure, and providing parameters values for which pertinent data are lacking. The latter task presents considerable challenges in Bayesian network (BN) modeling due to heuristics and biases in decision making under uncertainty [3, 4]. The available data to calibrate a BN often provide limited coverage, or could be of poor quality or completely missing. Domain experts are sometimes the only viable information source. Developing a valid and efficient method to better capture domain experts' insight is crucial for BN modeling.

Providing assessments directly in probabilistic terms can be especially difficult for domain experts who are unfamiliar with probabilities. Elicitation approaches that demand less statistical esotericism are potentially useful for BN calibration. A popular method in marketing research, conjoint analysis [5] offers a novel applicability to elicitations for BN models. We developed a web-based interactive conjoint analysis expert elicitation user interface that offers pair-wise scenario comparisons for experts' feedback as shown in Figure 2. The scenarios are combinations of nodes and node states from the underlying BN model. The expert indicates his/her assessment by moving the anchor along the scale, then submits the assessments, and immediately engages in another comparison. This process is run iteratively until sufficient information is collected for model calibration. Our initial test with conjoint analysis-based probability elicitation suggests that a fairly large amount of assessments can be obtained within a rather short time frame and that the elicitation approach has great potential for model calibration [6].
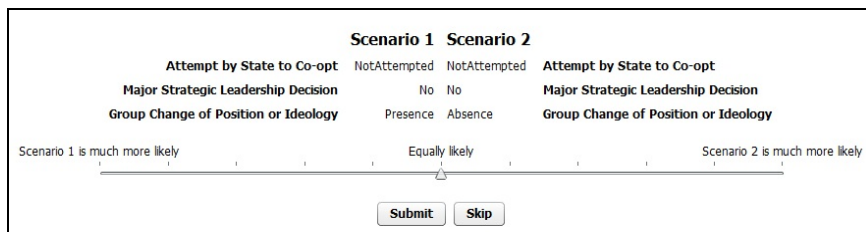
Fig. 2. Bayesian network model fragment representing one aspect of radicalization

3   Empirical Model Validation

In the first sub-section, we describe how empirical observations are used in model building and validation. We define model calibration as estimating the model

parameters from empirical observations. In the second sub-section, we present a graphical diagnostic for the structural fit of the model with data. The major benefits of such a calibration include:

1. A calibrated model that can, with some measurable level of precision, forecast observations within the same data space
2. The ability to measure the quality of the model fit to the observations
3. The ability to inspect the influence of more extreme observations on the model fit
4. The ability to compare the relative fits of different hypothesized model structures
5. The ability to measure and inspect the degree of consistency of the model structure.

The above capabilities are integral to testing domain theory against the state of nature and to arriving at a "good," trustable model that may be utilized in making prediction or other inference.

3.1   Empirical Model Calibration

The Bayesian approach to calibration requires the specification of a prior distribution on the model parameters, $\Pr(\theta)$, and may be combined with the data through the likelihood $L(\theta|X)$ to arrive at an updated posterior distribution $\Pr(\theta|X)$ [7]. This is achieved through the general Bayesian estimation equation:

$$\Pr(\theta \mid X) \propto L(\theta \mid X) \times \Pr(\theta).$$

Point estimates and uncertainties are readily derived from the posterior. The arguments to adopt a Bayesian approach to parameter estimation for BNs are compelling [8, 9]. The Bayesian approach enables incorporation of expert opinion through the prior distribution and produces a posterior which aggregates the data and opinions. Further, the Bayesian approach enables calibration of some model variables with data even if other variables are not observed. We implemented a Bayesian approach to model calibration [10]. The calibration results were shown for a model describing Group Radicalization.

3.2   Data Driven Model Diagnostics

Diagnostic checking is the act of inspecting the consistency of the modeling assumptions with the data. Inferences drawn from a calibrated model are only as trustable as the modeling assumptions are valid [11]. The major assumptions in specifying the structure of the BN are the conditional independencies among the variables. Thus, a methodology to inspect the validity of the conditional independencies in data is desirable. Statistical diagnostics are often approached through graphical inspection of model artifacts (such as residuals) for their consistency with the modeling assumptions. We accomplish this for BNs by plotting the hypothesized conditional independencies by using back-to-back bar charts. If the assumption (that is, the structure) is supported by the data, we should find reasonable symmetry in the back-to-back distributions.

The Group Radicalization model [10] assumes that the variables *Isolated and Threatened* and *Competing for Support* are conditionally independent, given the level of *Radicalization*. Figure 3 represents a visual inspection of this assumption in the

MAROB data [12]: if the assumption were true, the back-to-back distributions of *Isolated and Threatened* given each level of *Competing for Support* and *Radicalization* should be symmetric. The black dots represent a data driven quantification of this assumed symmetry. Their expectation is 0 under the hypothesized independence. The bands around the black dots given an uncertainty measure on this quantity are coded as green if the symmetry is reasonable, and red, otherwise. The graphic provides some evidence that, for the level of *Radicalization* in the "High" state, *Isolated/Threatened* is not independent of *Competing for Support*. Further, the graphic indicates that groups that are competing for supporters are less likely to be isolated and threatened than groups that are not competing. This result indicates that we may wish to alter the model structure by adding a direct dependence between *Isolated and Threatened* and *Competing for Support* in the BN.
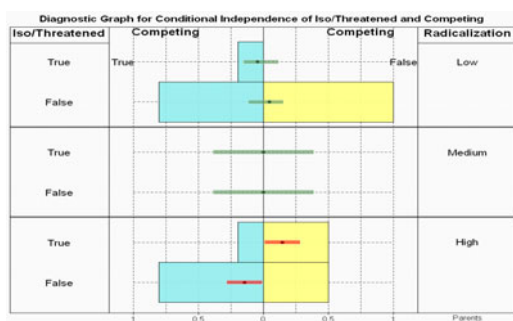
Fig. 3. Bayesian network model fragment representing one aspect of radicalization

4   Evidence: Assessment and Linkage with BN Models

After a model has been constructed and calibrated, the next step is typically to apply the model. A common analytic task is to assess the change in likelihood of the outcome of the model in light of observed evidence such as news reports, sensor data, and first hand testimonial. This is commonly accomplished in Bayesian network analysis by "setting" the state of an indicator to 1, and allowing the BN computation engine to compute the probabilities of the other nodes' states given that new information. We have developed a more nuanced approach to evidence handling to reflect fundamental thinking in evidence assessment.

Schum [13] distinguishes between an event and the evidence of the event, and identifies three attributes of credibility that must be considered when determining the inferential force that a piece of evidence has on the model:

- Veracity – is the witness or attester of the event expected to report truthfully?
- Observational sensitivity – is it possible for the witness to have observed the event?
- Objectivity – does the witness of the event have a stake in the outcome?

Likewise, our Bayesian network models assess the effect of whether or not an event occurs on the model separately from the likelihood of an event, given the evidence at

hand. Since the indicators in our social science models are taken from academic papers, they tend to be broad categories. The events, which are suggested by the model author or added by the user, are observable and concrete items. When the analyst adds a piece of evidence to the model, they are asked to choose the event to which it is relevant, whether it supports or refutes the event, and to rate the credibility and bias of the evidence.

When evidence is attached to the Bayesian network model in the manner described above, the probabilities of the states of all indicators in the model are changed. Our system displays these changes for the user in a bar chart (Figure 4), showing the ratio of the probability with evidence over the probability before evidence. The X-axis of this plot is a quasi-log scale. The center tick means no change, the ticks to the right mean ratios of 2, 3 and 4, respectively (i.e. the probability doubled, tripled or quadrupled due to the evidence), and the ticks to the left mean 1/2, 1/3 and 1/4 (the probability was halved, etc, due to the evidence). Figure 4 shows that the probability of Proliferation Propensity being Very High was roughly halved due to the evidence that the user added to the model.

Fig. 4. Change in probabilities of proliferation propensity resulting from evidence

5 Example: Nuclear Non-proliferation

All of the modeling pieces described above – a Bayesian network model, calibration with data and experts' assessments, and the evidence gathering and attachment tool – form an integrated modeling framework, which we have applied to nuclear proliferation risk assessments. Below we discuss how this novel approach was employed to create case studies of ranking state-level proliferation propensity.

5.1 Model Development and Motivation

The proliferation risk model [14] integrates technical, social, and geopolitical factors into a comprehensive modeling framework to estimate the state actors' propensity to acquire nuclear weapons capability. This modeling approach benefits from insight offered by both technical proliferation modeling and social modeling, and is well grounded in academic literature. The Bayesian network model serves as the basic structure on which evidence is organized and relevant queries developed.

We initially quantified the model with expert judgment, and subsequently with consolidated and enhanced data sets [14]. While these data were useful in defining base rates and have allowed us to exercise the model for specific countries, the conditional probability tables are still based upon expert judgment in various forms.
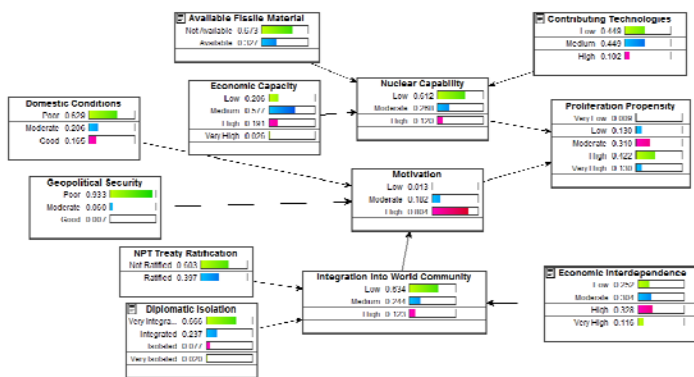
Fig. 5. Bayesian network model fragment for whether a country has a nuclear program

5.2 Data Gathering Process

The integrated data gathering approach consists of four steps: target identification, evidence gathering, evaluation and attachment, and model updating. We focused on Africa since states in this continent are markedly diverse in terms of economic development, technical capabilities, regime stability, and governance. We created country proliferation propensity rankings for the following countries: South Africa, Nigeria, Uganda, Gabon, Senegal, and Ghana. During the evidence gathering stage, researchers developed two query sets for most of the nodes in the model, and then the query sets were merged to reduce redundancy and ensure better search term representation. For instance, for the node named "Nuclear Capability," we created a broad search query such as "(Country name) has nuclear capability." In the same vein, for the node, "Democracy," which represents a state's degree of democracy, a query, "(Country name) is democratic," was created. In some cases, multiple queries were created to provide broader search coverage. The queries were run and candidate evidence retrieved. We evaluated the evidence in terms of its reliability and bias in accordance with the regular criteria.

5.3 BACH and Results

The model development and data gathering were done in a software tool we created called BACH (Bayesian Analysis of Competing Hypotheses). BACH is a Java-based tool for Bayes net model creation, evidence search and assessment, and evaluation of the effect of evidence on a user's hypotheses. BACH allows the user to create a model (as shown in Figure 5) and to document the model creation process including links and references. The tool also includes the ability to construct and save evidence queries (which at the moment search over the internet, but can be adapted to search databases or local networks) and supports evidence importation and credibility and bias assessment. When evidence is attached or the model is changed, BACH instantly updates the new hypothesis probabilities and the outcome effects (shown in Figure 4).

BACH updated the outcome probabilities for state proliferation propensity to reflect the attached evidence. The state propensity rankings are shown in Table 1 below. The likelihood of developing a nuclear weapons program from most likely to least likely is: South Africa, Nigeria, Ghana, Senegal, Gabon, and Uganda. The outcome change ratio measures the magnitude of propensity change resulting from the attached evidence for a given state and is calculated as [1- (base rate – updated calculation)/base rate]. The outcome node has five states: very low, low, moderate, high, and very high, and the results suggest that the attached evidence produced the greatest impact on the "very high" state for South Africa and Nigeria, the "very low" state for Ghana and Senegal, and the "low" state for Gabon and Uganda.

Table 1. Proliferation Propensity Rankings for Six African Countries

Propensity Ranking	Country	Propensity (p)	Max Outcome Change Ratio	State with Max Changes
1	South Africa	0.42	2.12	Very High
2	Nigeria	0.35	1.76	Very High
3	Ghana	0.18	1.34	Very Low
4	Senegal	0.16	1.31	Very Low
5	Gabon	0.12	1.13	Low
6	Uganda	0.09	1.23	Low

6 Conclusions

We have presented a "cradle to grave" perspective on how BN models can be constructed, validated, and used in socio-behavioral modeling. The modeling construction process is suitable for a wide range of human and organizational modeling activities. We have found it suitable for steady-state or a "slice of time" modeling and analysis settings. The behavioral mechanisms are proposed in the literature, once represented in the form of a BN model, can be the subject of critical empirical evaluations. Further, a well-thought approach for evidence assessment is also available in this framework. This set of features contributes to the scientific utility of this category of modeling approaches.

References

1. Jensen, F., Nielsen, T.: Bayesian Networks and Decision Graphs, 2nd edn. Springer, New York (2007)
2. McCauley, C., Moskalenko, S.: Mechanisms of Political Radicalization: Pathways Toward Terrorism. Terrorism and Political Violence 20, 416–433 (2008)
3. Kahneman, D., Tversky, A.: Judgment under Uncertainty: Heuristics and Biases. Science 4157, 1124–1131 (1974)
4. Renooij, S.: Probability Elicitation for Belief Networks: Issues to Consider. Knowledge Engineering Review 16, 255–269 (2001)

5. Green, P., Krieger, A., Wind, Y.: Thirty Years of Conjoint Analysis: Reflections and Prospects. Interfaces 73, S56–S73 (2001)
6. Walsh, S., Dalton, A., White, A., Whitney, P.: Parameterizing Bayesian Network Representations of Social-Behavioral Models by Expert Elicitation: A Novel Application of Conjoint Analysis. In: Proceedings of Workshop on Current Issues in Predictive Approaches to Intelligent and Security Analytics (PAISA), pp. 227–232. IEEE Press, Los Alamitos (2010)
7. Cox, D., Hinkley, D.: Theoretical Statistics. Chapman & Hall, London (1974)
8. Heckerman, D., Geiger, D., Chickering, D.: Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Journal of Machine Learning 20, 197–243 (1995)
9. Riggelson, C.: Learning Parameters of Bayesian: Networks from Incomplete Data via Importance Sampling. International Journal of Approximate Reasoning 42, 69–83 (2005)
10. Whitney, P., Walsh, S.: Calibrating Bayesian Network Representations of Social-Behavioral Models. In: Chai, S.-K., Salerno, J.J., Mabry, P.L. (eds.) SBP 2010. LNCS, vol. 6007, pp. 338–345. Springer, Heidelberg (2010)
11. Cleveland, W.: Visualizing Data. Hobart Press, Summit (1993)
12. Minorities at Risk Organizational Behavior Dataset,
    `http://www.cidcm.umd.edu/mar`
13. Schum, D.: The Evidential Foundations of Probabilistic Reasoning. John Wiley & Sons, New York (1994)
14. Coles, G.A., Brothers, A.J., Gastelum, Z.N., Thompson, S.E.: Utility of Social Modeling for Proliferation Assessment: Preliminary Assessment. PNNL 18438 (2009)

A Psychological Model for Aggregating Judgments of Magnitude

Edgar C. Merkle[1] and Mark Steyvers[2]

[1] Department of Psychology
Wichita State University
edgar.merkle@wichita.edu
[2] Department of Cognitive Sciences
University of California, Irvine
mark.steyvers@uci.edu

Abstract. In this paper, we develop and illustrate a psychologically-motivated model for aggregating judgments of magnitude across experts. The model assumes that experts' judgments are perturbed from the truth by both systematic biases and random error, and it provides aggregated estimates that are implicitly based on the application of nonlinear weights to individual judgments. The model is also easily extended to situations where experts report multiple quantile judgments. We apply the model to expert judgments concerning flange leaks in a chemical plant, illustrating its use and comparing it to baseline measures.

Keywords: Aggregation, magnitude judgment, expert judgment.

1   Introduction

Magnitude judgments and forecasts are often utilized to make decisions on matters such as national security, medical treatment, and the economy. Unlike probability judgments or Likert judgments, magnitude judgments are often unbounded and simply require the expert to report a number reflecting his or her "best guess." This type of judgment is elicited, for example, when an expert is asked to forecast the year-end value of a stock or to estimate the number of individuals afflicted with H1N1 within a specific geographic area.

Much work has focused on aggregating the judgments of multiple experts, which is advantageous because the aggregated estimate often outperforms the individuals [1]. A simple aggregation method, the *linear opinion pool*, involves a weighted average of individual expert judgments. If experts are equally weighted, the linear opinion pool neglects variability in: (1) the extent to which experts are knowledgeable, and (2) the extent to which experts can translate their knowledge into estimates. As a result, a variety of other methods for assigning weights and aggregating across judges have been proposed. These methods have most often been applied to probability judgments [2,3].

In the domain of probability judgments, the *supra-Bayesian* approach to aggregation [4,5] has received attention. This approach involves initial specification

of one's prior belief about the occurrence of some focal event. Experts' judged probabilities are then used to update this prior via Bayes' theorem. This method proves difficult to implement because one must also specify a distribution for the experts' judgments conditional on the true probability of the focal event's occurrence. As a result, others have focused on algorithms for weighting expert judgments. Kahn [6] derived weights for a logarithmic opinion pool (utilizing the geometric mean for aggregation), with weights being determined by prior beliefs, expert bias, and correlations between experts. This method requires the ground truth to be known for some items. Further, Cooke [7] developed a method to assign weights to experts based on their interval (i.e., quantile) judgments. Weights are determined both by the extent to which the intervals are close to the ground truth (for items where the truth is known) and by the width of the intervals. This method can only be fully utilized when the ground truth is known for some items, but it can also be extended to judgments of magnitude.

Instead of explicitly weighting experts in an opinion pool, Batchelder, Romney, and colleagues (e.g., [8,9]) have developed a series of Cultural Consensus Theory models that aggregate categorical judgments. These models account for individual differences in expert knowledge and biases, and they implicitly weight the judgments to reflect these individual differences. The aggregated judgments are not obtained by averaging over expert judgments; instead, the aggregated judgments are estimated as parameters within the model. Additionally, these models can be used (and are intended for) situations where the ground truth is unknown for all items.

In this paper, we develop a model for aggregating judgments of magnitude across experts. The model bears some relationships to Cultural Consensus Theory, but it is unique in that it is applied to magnitude judgments and can handle multiple judgments from a single expert for a given item. The model is also psychologically motivated, building off of the heuristic methods for psychological aggregation described by Yaniv [10] and others. Finally, the model can be estimated both when the ground truth is unknown for all items and when the ground truth is known for some items.

In the following pages, we first describe the model in detail. We then illustrate its ability to aggregate expert judgments concerning leaks at a chemical plant. Finally, we describe general use of the model in practice and areas for future extensions.

2   Model

For a given expert, let X_{jk} represent a magnitude prediction of quantile q_k for event j (in typical applications that we consider, $q_1 = .05; q_2 = .5; q_3 = .95$). All predictions, then, may be represented as a three-dimensional array $\boldsymbol{X}$, with the entry X_{ijk} being expert i's prediction of quantile q_k for event j. The model is intended to aggregate across the i subscript, so that we have aggregated quantile estimates for each event.

We assume an underlying, "true" uncertainty distribution for each event j, with individual experts' uncertainty distributions differing systematically from the true distribution. Formally, let V_j be the true uncertainty distribution associated with event j. Then

$$V_j \sim \mathrm{N}(\mu_j^*, \sigma_j^{2*}), \tag{1}$$

where normality is assumed for convenience but can be relaxed. For events with known outcomes, we can fix μ_j^* at the outcome. Alternatively, when no outcomes are known, we can estimate all μ_j^*.

We assume that each expert's uncertainty distribution differs systematically from the true distribution. The mean (μ_{ij}) and variance (σ_{ij}^2) of expert i's distribution for event j is given by:

$$\mu_{ij} = \alpha_i + \mu_j^*, \tag{2}$$
$$\sigma_{ij}^2 = \sigma_j^{2*}/D_i,$$

where the parameter α_i reflects systematic biases of expert i and receives the hyperdistribution:

$$\alpha_i \sim \mathrm{N}(\mu_\alpha, \phi_\alpha), \ i = 1, \ldots, N. \tag{3}$$

The parameter D_i reflects expertise: larger values reflect less uncertainty. D_i is restricted to lie in $(0, 1) \ \forall \ i$ because, by definition, expert uncertainty can never be less than the "true" uncertainty.

We assume that expert i's quantile estimates arise from his/her unique uncertainty distribution and are perturbed by response error:

$$X_{ijk} \sim \mathrm{N}(\Phi^{-1}(q_k)\sigma_{ij} + \mu_{ij}, \gamma_i^2), \tag{4}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution function. The response error reflects the fact that the expert's quantile estimates are likely to stray from the quantiles of his/her underlying normal distribution.

We are estimating a Bayesian version of the above model via MCMC, so we need prior distributions for the parameters. These are given by

$$
\begin{array}{lll}
D_i \sim \mathrm{Unif}(0,1) & \gamma_i^2 \sim \mathrm{Unif}(0,1000) & \mu_j^* \sim \mathrm{N}(0, 1.5) \\
\mu_\alpha \sim \mathrm{N}(0,4) & \phi_\alpha \sim \mathrm{Unif}(0,1000) & \sigma_j^{2*} \sim \mathrm{Gamma}(.001,.001)
\end{array}. \tag{5}
$$

The above priors are minimally informative, though this is not obvious for the priors on μ_j^* and μ_α. We further describe the rationale for these priors in the application. An additional advantage of the Bayesian approach involves the fact that we can easily gauge uncertainty in the estimated ground truth.

In summary, the model developed here is unique in that: (1) the ground truth for each item is estimated via a model parameter, and (2) specific distributional assumptions are very flexible. Psychological aspects of the model include the facts that: (1) judges are allowed to have systematic response biases; and (2) judges are allowed to vary in expertise.

3   Application: Flange Leaks

To illustrate the model, we use data on the causes of flange leaks at a chemical plant in the Netherlands [11,12]. Ten experts estimated the number of times in the past ten years that specific issues (called "items" below) caused a failure of flange connections. Issues included "changes in temperature caused by process cycles," "aging of the gasket," and "bolts improperly tightened." For each issue, the experts reported 5%, 50%, and 95% quantiles of their uncertainty distributions. There were fourteen issues assessed, yielding a total of $10 \times 14 \times 3 = 420$ observations. The ground truth (i.e., true number of times the issue led to a failure in the past ten years) was known for eight of the fourteen issues.

3.1   Implementation Details

The judgments elicited from experts varied greatly from item to item, reflecting the fact that some issues were much more frequent than others. Figure 1 illustrates this phenomenon in more detail. The figure contains 8 panels, each of which displays individual expert judgments in grey (labeled E1 through E10). The dots with horizontal lines reflect experts' 5%, 50%, and 95% quantile judgments, and the horizontal line in each panel reflects the truth. The vast differences in magnitude can be observed by, e.g., comparing judgments for item 5 with those for item 6 (keeping in mind that the y-axis changes for each item). Further, focusing on item 5, we can observe extreme intervals that are far from the truth (for item 5, close to zero).

Both the vast differences in judged magnitudes and the large outliers lead to poor model performance. To minimize these issues, we fit the model to standardized, log-transformed judgments for all items. These transformations make model estimation more stable, and the resulting model estimates can be "untransformed" to obtain aggregated estimates on the original scale. The transformed data also lead to default prior distributions on the μ_j^*: because we are dealing with standardized data, a reasonable prior would be $N(0,1)$. Because experts tend to under- or overestimate some quantities, we have found it helpful to increase the prior variance to, say, 1.5. This also leads to the prior variance on μ_α: because we are dealing with standardized data, it would be quite surprising to observe a mean bias outside of $(-2, 2)$.

We used three implementations of the model to aggregate estimates in different ways. In the first implementation (M1), we fit the model to expert judgments for all 14 items, excluding the ground truth. This implementation reflects a situation where the truth is unknown and estimates must be aggregated. In the second implementation (M2), we used the same data but included the ground truth for the first four items. This reflects a situation where the truth unfolds over time or was already known to the analyst, so that experts can be "graded" on some items but not others. The third model implementation (M3) held out the ground truth for one item at a time. Thus, the estimated interval for the
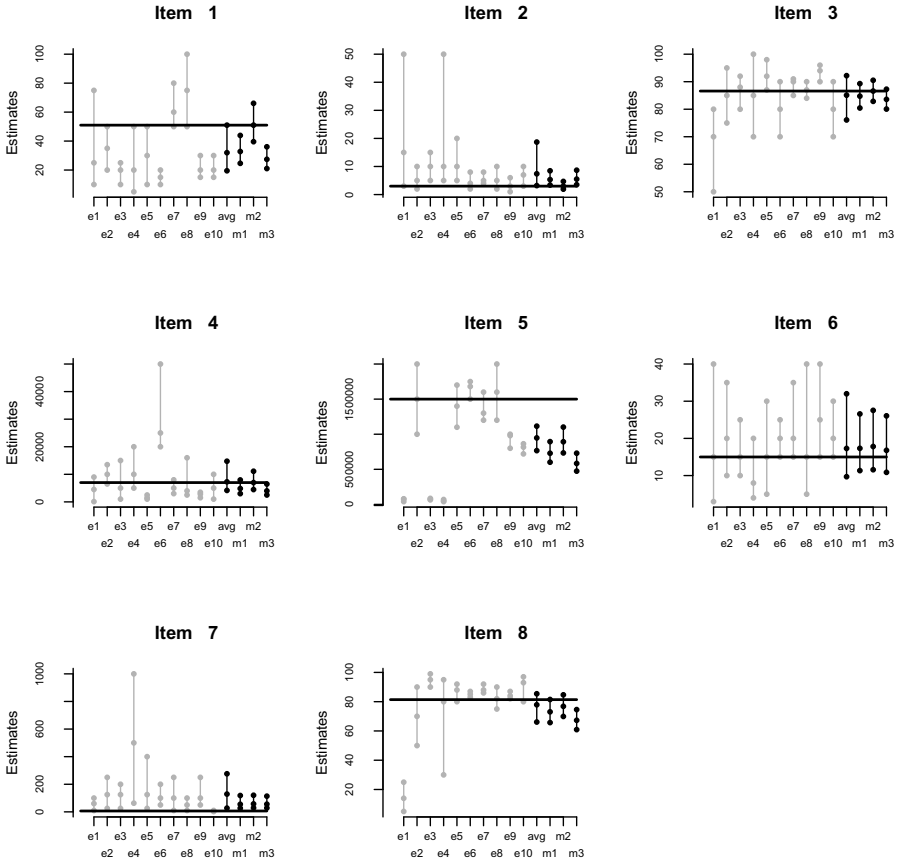
Fig. 1. Magnitude estimates of ten experts and four aggregation methods for eight items. Vertical lines with points reflects 5%, 50%, and 95% quantile judgments; "e" labels represent experts, and the "avg" label represents the unweighted average aggregation. "M1" represents model-based aggregations with no ground truth, "M2" represents model-based aggregations with ground truth for the first four items, and "M3" represents model-based aggregations for all items except the one being estimated. The horizontal line in each panel represents the ground truth.

first item utilized the ground truth for items 2–8. The estimated interval for the second item utilized the ground truth for items 1 and 3–8, and so on. Finally, for comparison, we also aggregated the estimates via an unweighted average.

3.2   Main Results

Figure 1 contains the four aggregated intervals (black lines), with `avg` being unweighted averages and `M1, M2, M3` being the model implementations. It can be seen that, as compared to the experts' reported intervals, the aggregated intervals are consistently narrower and closer to the truth. The model-based

aggregations also result in narrower intervals, as compared to the simple average across experts. Comparing the three aggregated intervals with one another, it can be seen that the (M2) intervals are centered at the true value for Items 1–4. However, for items where the ground truth was not known (items 5–8), the intervals are very similar for all three methods. This implies that the model did not benefit from "knowing" the ground truth on some items.

We now focus on statistical summaries of Figure 1. Table 1 displays four statistics for each expert and for the aggregated intervals. The first is the number of intervals (out of eight) that covered the truth. The other four are ratio measures, comparing each expert/model to the model-based aggregations with no ground truth (M1). These measures include the average ratio of interval widths, defined as:

$$\text{Width} = \frac{1}{8} \sum_{j=1}^{8} \frac{(X_{ij3} - X_{ij1})}{(\widehat{X}_{j3} - \widehat{X}_{j1})}, \tag{6}$$

where i is fixed and the $\widehat{X}$ are aggregated quantiles from M1. The next measure is a standardized bias ratio, defined as:

$$\text{Bias} = \frac{\sum_j |X_{ij2} - \mu_j^*|/\mu_j^*}{\sum_j |\widehat{X}_{j2} - \mu_j^*|/\mu_j^*}, \tag{7}$$

where μ_j^* is the ground truth. The last measure is the average likelihood of the ground truth under each expert's interval (assuming normality), relative to the sum of likelihoods under the expert's interval and under the aggregated interval from M1:

$$\text{LR} = \frac{1}{8} \sum_{j=1}^{8} \frac{\phi((\mu_j^* - X_{ij2})/\sigma_{ij})}{\phi((\mu_j^* - X_{ij2})/\sigma_{ij}) + \phi((\mu_j^* - \widehat{X}_{j2})/\widehat{\sigma}_j)}, \tag{8}$$

where $\phi()$ is the standard normal density function. This is a measure of the relative likelihood of the truth for the expert's intervals, as compared to the M1 intervals.

For the width and bias measures described above, values less than one imply that the expert (or other aggregation) did better than M1 and values greater than one imply that the expert (or aggregation) did worse than M1. The likelihood measure, on the other hand, must lie in (0,1). Values less than .5 imply that the expert did worse than M1, and values greater than .5 imply the converse.

Table 1 contains the statistics described above for each expert and for the aggregations. Statistics for M2 are not included because its estimated intervals are centered at the ground truth for the first four items (leading to inflated performance measures). Examining the individual expert judgments, the table shows that no expert beats M1 on all four measures. Expert 8 is closest: her intervals are wider than the model, but she covers more items, exhibits less bias, and exhibits greater truth likelihoods.

Examining the aggregations, M1 beats the unweighted average on both interval width and bias. The two aggregation methods are similar on coverage and

Table 1. Statistics for individual experts' judgments and model-based, aggregated judgments. "Coverage" is number of intervals that cover the truth (out of 8), "Width" is ratio of of mean interval width to that of M1, "Bias" is ratio of mean standardized, absolute bias to that resulting from M1, and "LR" is the likelihood of the truth for each interval relative to M1.

Expert	Coverage	Width	Bias	LR
e1	3	2.8	1.5	0.28
e2	6	2.1	2.1	0.53
e3	3	1.3	2.3	0.19
e4	4	4.1	8.4	0.29
e5	3	1.9	2.3	0.45
e6	2	1.7	1.9	0.34
e7	4	1.2	1.7	0.53
e8	7	1.8	0.9	0.61
e9	1	1.0	1.7	0.39
e10	4	1.1	0.4	0.36
avg	4	1.9	2.2	0.55
M1	4	–	–	–
M3	2	0.9	1.1	0.26

likelihood. Comparing M1 with M3, M3 is worse on coverage and likelihood ratio. These findings, which generally match the visual results in Figure 1, imply that having access to the ground truth for some items did not improve model estimates.

4   Conclusions

The psychological model for aggregating magnitude judgments described in this paper exhibits good statistical properties, including an ability to consistently outperform expert judgments and judgments from simple aggregation methods. Unexpectedly, the model was unable to benefit from the ground truth (though see [13] for a similar result in a different context). We view this to be related to the model assumption that experts consistently over- or underestimate the ground truth across items. This assumption is violated if, say, experts are very knowledgeable about some items but not others. If the assumption is violated, then the model cannot obtain good estimates of expert biases because the biases are changing across items. This leads to the disutility of the ground truth. A potential solution would involve the addition of "item saliency" parameters into the model, though we may quickly find that the model is overparameterized.

While these and other extensions may be considered, the model developed here exhibits reasonable performance and has not been tailored to the content area in any way. It may generally be applied to domains where magnitude judgments are utilized, and it is very flexible on the number of quantile judgments elicited, the presence of ground truth, and the extent to which all experts judge all items. The model also utilizes interpretable parameters, allowing the analyst to maintain

some control over the model's behavior. This feature may be especially useful in situations where the costs of incorrect judgments are high. More generally, the model possesses a unique set of features that give it good potential for future application.

References

1. Surowiecki, J.: The wisdom of crowds. Anchor, New York (2005)
2. Clemen, R.T., Winkler, R.L.: Combining probability distributions from experts in risk analysis. Risk Analysis 19, 187–203 (1999)
3. O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T.: Uncertain judgements: Eliciting experts' probabilities. Wiley, Hoboken (2006)
4. Morris, P.A.: Combining expert judgments: A Bayesian approach. Management Science 23, 679–693 (1977)
5. Genest, C., Zidek, J.V.: Combining probability distributions: A critique and annotated bibliography. Statistical Science 1, 114–148 (1986)
6. Kahn, J.M.: A generative Bayesian model for aggregating experts' probabilities. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 301–308 (2004)
7. Cooke, R.M.: Experts in uncertainty: Opinion and subjective probability in science. Oxford University Press, New York (1991)
8. Batchelder, W.H., Romney, A.K.: Test theory without an answer key. Psychometrika 53, 71–92 (1988)
9. Karabatsos, G., Batchelder, W.H.: Markov chain estimation for test theory without an answer key. Psychometrika 68, 373–389 (2003)
10. Yaniv, I.: Weighting and trimming: Heuristics for aggregating judgments under uncertainty. Organizational Behavior and Human Decision Processes 69, 237–249 (1997)
11. Cooke, R.M.: Experts in uncertainty: Opinion and subjective probability in science. Oxford, New York (1991)
12. Cooke, R.M., Goossens, L.L.H.J.: TU Delft expert judgment data base. Reliability engineering and system safety 93, 657–674 (2008)
13. Dani, V., Madani, O., Pennock, D., Sanghai, S., Galebach, B.: An empirical comparison of algorithms for aggregating expert predictions. In: Proceedings of the Twenty-Second Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI 2006), Arlington, Virginia, pp. 106–113. AUAI Press (2006)

Speeding Up Network Layout and Centrality Measures for Social Computing Goals

Puneet Sharma[1,2], Udayan Khurana[1,2], Ben Shneiderman[1,2],
Max Scharrenbroich[3], and John Locke[1]

[1] Computer Science Department
[2] Human-Computer Interaction Lab
[3] Department of Mathematics
University of Maryland
College Park MD 20740
{puneet,udayan,ben}@cs.umd.edu,
{john.b.locke,max.franz.s}@gmail.com

Abstract. This paper presents strategies for speeding up calculation of graph metrics and layout by exploiting the parallel architecture of modern day Graphics Processing Units (GPU), specifically *Compute Unified Device Architecture (CUDA)* by *Nvidia*. Graph centrality metrics like *Eigenvector, Betweenness, Page Rank* and layout algorithms like *Fruchterman-Rheingold* are essential components of *Social Network Analysis (SNA)*. With the growth in adoption of SNA in different domains and increasing availability of huge networked datasets for analysis, social network analysts require faster tools that are also scalable. Our results, using NodeXL, show up to 802 times speedup for a Fruchterman-Rheingold graph layout and up to 17,972 times speedup for Eigenvector centrality metric calculations on a 240 core CUDA-capable GPU.

Keywords: Social Computing, Social Network Analysis, CUDA.

1 Introduction

The enormous growth of social media (e.g. email, threaded discussions, wikis, blogs, microblogs), social networking (e.g. Facebook, LinkedIn, Orkut), and user generated content (e.g. Flickr, YouTube, Digg, ePinions) presents attractive opportunities for social, cultural, and behavioral analysts with many motivations. For the first time in history, analysts have access to rapidly changing news stories, national political movements, or emerging trends in every social niche. While temporal trend analysis answers many questions, network analysis is more helpful in discovering relationships among people, sub-communities, organizations, terms, concepts, and even emotional states.

Many tools have been developed to conduct Social Network Analysis (SNA), but our effort is tied to NodeXL[1] [3], a network data analysis and visualization [8] plug-in for Microsoft Excel 2007 that provides a powerful and simple means to graph data

---

[1] NodeXL: Network Overview, Discovery and Exploration for Excel
http://nodexl.codeplex.com/

contained in a spreadsheet. Data may be imported from an external tool and formatted to NodeXL specifications, or imported directly from the tool using one of the supported mechanisms (such as importing an email, Twitter, Flickr, YouTube, or other network). The program maps out vertices and edges using a variety of layout algorithms, and calculates important metrics on the data to identify nodes and relationships of interest.

While NodeXL is capable of visualizing a vast amount of data, it may not be feasible to run its complex computation algorithms on larger datasets (like those found in Stanford's SNAP library[2]) using hardware that is typical for a casual desktop user. And we believe that in order for data visualization tools to reach their full potential, it is important that they are accessible and fast for users using average desktop hardware.

Nvidia's CUDA[3] technology provides a means to employ the dozens of processing cores present on modern video cards to perform computations typically meant for a CPU. We believe that CUDA is appropriate to improve the algorithms in NodeXL due to the abundance of Graphical Processing Units (GPUs) on mid-range desktops, as well as the parallelizable nature of data visualization. Using this technology, we hope to allow users to visualize and analyze previously computationally infeasible datasets.

This paper has two contributions. The first contribution is to provide the strategies and directions to GPU-parallelize the existing layout algorithms and centralities measures on commodity hardware. And the second contribution to apply these ideas to widely available and accepted visualization tool, NodeXL. To our knowledge, ours is the first attempt to GPU-parallelize the existing layout algorithms and centrality measures on commodity hardware. In our modified NodeXL implementation, we targeted two computationally expensive data visualization procedures: the Fruchterman-Rheingold [1] force-directed layout algorithm, and the Eigenvector centrality node metric.

2   Computational Complexity of SNA Algorithms

Social Network Analysis consists of three major areas - Network Visualization, Centrality Metric Calculation and Cluster detection. Popular techniques for visualization use force-directed algorithms to calculate a suitable layout for nodes and edges. Computationally, these algorithms are similar to n-body simulations done in physics. The two most popular layout algorithms are Harel-Koren [2] and Fruchterman-Rheingold [1]. The computational complexity of the latter can roughly be expressed as $O(k(V^2+E))$ and the memory complexity as $O(E+V)$, where k is the number of iterations needed before convergence, V is the number of vertices and E is the number of edges in the graph. Centrality algorithms calculate the relative importance of each node with respect to rest of the graph. Eigenvector centrality [6] measures the importance of a node by the measure of its connectivity to other "important" nodes in the graph. The process of finding it is similar to belief propagation and the algorithm is iterative with the time complexity $O(k.E)$.

With the increase in the number of vertices and edges, the computation becomes super-linearly expensive and nowadays analysis of a very highly dense graph is a reasonable task. Therefore, speedup and scalability are key challenges to SNA.

2.1 Approaches for Speedup and Scalability

To achieve greater speedup, two approaches that immediately come to mind are faster algorithms with better implementation, and secondly, more hardware to compute in parallel. If we believed that the best algorithms (possible, or the best known) are already in place, the second area is worth investigating. However, it is the nature of the algorithm which determines that whether it can be executed effectively in a parallel environment or not. If the graph can be easily partitioned into separate workable sets, then the algorithm is suitable for parallel computing, otherwise not. This is because the communication between nodes in such an environment is minimal, and so all the required data must be present locally. We divide the set of algorithms used in SNA into three different categories based on feasibility of graph partitioning:

- **Easy Partitioning** - Algorithm requires only the information about locally connected nodes (neighbors) e.g. Vertex Degree Centrality, Eigenvector centrality
- **Relatively Hard Partitioning** - Algorithm requires global graph knowledge, but local knowledge is dominant and efficient approximations can be made. e.g. Fruchterman-Rheingold
- **Hard Partitioning** - Data partitioning is not an option. Global graph knowledge is required at each node. e.g. Betweenness centrality which requires calculation of All Pairs of Shortest Paths.

Algorithms that fall under the first category can be made to scale indefinitely using a distributed systems like Hadoop [7] and GPGPUs. For the second category, there is a possibility of a working approximation algorithm, whereas nothing can be said about the third category of algorithms.

3 Layout Speedup

3.1 Layout Enhancement

In NodeXL layout algorithms, the Fruchterman-Rheingold algorithm was chosen as a candidate for enhancement due to its popularity as an undirected graph layout algorithm in a wide range of visualization scenarios, as well as the parallelizable nature of its algorithm. The algorithm works by placing vertices randomly, then independently calculating the attractive and repulsive forces between nodes based on the connections between them specified in the Excel workbook. The total kinetic energy of the network is also summed up during the calculations to determine whether the algorithm has reached a threshold at which additional iterations of the algorithm are not required.

The part of the algorithm that computes the repulsive forces is $O(N^2)$, indicating that it could greatly benefit from a performance enhancement (the rest of the algorithm is $O(|V|)$ or $O(|E|)$, where v and e are the vertices and edges of the graph,

respectively). The algorithm is presented in pseudo-code and computationally intensive repulsive calculation portion of the algorithm is italicized in a longer version of this paper [9].

Due to the multiple nested loops contained in this algorithm, and the fact that the calculations of the forces on the nodes are not dependent on other nodes, we implemented the calculation of these forces in parallel using CUDA. In particular, the vector V was distributed across GPU processing cores. The resulting algorithm, **Super Fruchterman-Rheingold**, was then fit into the NodeXL framework as a selectable layout to allow users to recruit their GPU in the force calculation process.

3.2   Results for Layout Speedup

We ran our modified NodeXL on the computer with CPU (3 GHz, Intel(R) Core(TM) 2 Duo) and GPU (GeForce GTX 285, 1476 MHz, 240 cores). This machine is located in the Human Computer Interaction Lab at the University of Maryland, and is commonly used by researchers to visualize massive network graphs. The results of our layout algorithm are shown in the table below. Each of the graph instances listed may be found in the Stanford SNAP library:

Table 1. Results of layout speedup testing

Graph Instance Name	#Nodes	#Edges	Super F-R run time (ms)	Simple F-R run time (ms)	Speedup
CA-AstroPh	18,772	396,160	1,062.4	84,434.2	**79x**
cit-HepPh	34,546	421,578	1,078.0	343,643.0	**318x**
soc-Epinions1	75,879	508,837	1,890.5	1,520,924.6	**804x**
soc-Slashdot0811	77,360	905,468	2,515.5	1,578,283.1	**625x**
soc-Slashdot0902	82,168	948,464	2,671.7	1,780,947.2	**666x**

The results of parallelizing this algorithm were actually better than we had expected. The speedup was highly variable depending on the graph tested, but the algorithm appears to perform better with larger datasets, most likely due to the overhead of spawning additional threads. Interestingly, the algorithm also tends to perform better when there are many vertices to process, but a fewer number of edges.

This behavior is the result of reaping the benefits of many processing cores with a large number of vertices, while the calculation of the repulsive forces *at each node* still occurs in a sequential fashion. A graph with myriad nodes but no edges would take a trivial amount of time to execute on each processor, while a fully connected graph with few vertices would still require much sequential processing time. Regardless, the Super Fruchterman-Rheingold algorithm performs admirably even in

the worst case of our results, providing a 79x speedup in the *CA-AstroPh* set with a 1:21 vertex to edge ratio. The ability for a user to visualize their data within two seconds when it originally took 25 minutes is truly putting a wider set of data visualization into the hands of typical end users.

4   Metric Speedup

4.1   Metric Enhancement

The purpose of data visualization is to discover attributes and nodes in a given dataset that are interesting or anomalous. Such metrics to determine interesting data include *Betweenness centrality, Closeness centrality, Degree centrality,* and *Eigenvector centrality*. Each metric has its merits in identifying interesting nodes, but we chose to enhance Eigenvector centrality, as it is frequently used to determine the "important" nodes in a network. Eigenvector centrality rates vertices based on their connections to other vertices which are deemed important (connecting many nodes or those that are "gatekeepers" to many nodes). By the Perron-Frobenius theorem we know that for a real valued square matrix (like an adjacency matrix) there exists a largest unique eigenvalue that is paired with an eigenvector having all positive entries.

4.2   Power Method

An efficient method for finding the largest eigenvalue/eigenvector pair is the power iteration or power method. The power method is an iterative calculation that involves performing a matrix-vector multiply over and over until the change in the iterate (vector) falls below a user supplied threshold. We outline the power method algorithm with the pseudo-code in Section 4.2 in [9].

4.3   Power Method on a GPU

For our implementation we divided the computation between the host and the GPU. A kernel for matrix-vector multiply is launched on the GPU, followed by a device-to-host copy of the resulting vector for computation of the vector norm, then a host-to-device copy of the vector norm and finally a kernel for vector normalization is launched on the GPU. Figure 4 in [9] outlines the algorithm flow to allow execution on the GPU.

4.4   Results for Eigenvector Centrality Speedup

We tested our GPU algorithm on eleven graph instances of up to 19k nodes and 1.7m edges. As a comparison, we used the serial version of the algorithm implemented in NodeXL with C#. Also, we used the same machine used in section 3.2. Table 2 shows the speedups we achieved by running the algorithm on the GPU.

Speedups ranged between 102x and 17,972x. We investigated relationship between the problem instances and the speedups by examining how problem size, measured by the number of edges, affected the speedup.

There is an increase in the speed up as the size of the graph increases. We believe that this is due to the effects of scale on the serial CPU implementation for graphs with bigger sizes, whereas the overhead of dividing the data for a CUDA-based calculation was more of a detriment for smaller datasets.

Table 2. Tabular view of the speedups

Graph Name	#Nodes	#Edges	Time - GPU(sec)	Time - CPU (sec)	Speedup
Movie Reviews	2,625	99,999	0.187	23.0	**123x**
CalTech Facebook	769	33,312	0.015	1.6	**102x**
Oklahoma Facebook	17,425	1,785,056	0.547	9,828.0	**17972x**
Georgetown Facebook	9,414	851,278	0.187	1,320.0	**7040x**
UNC FB Network	18,163	1,533,602	1.874	13,492.0	**7196x**
Princeton FB Network	6,596	586,640	0.094	495.0	**5280x**
Saket Protein	5,492	40,332	0.109	107.0	**798x**
SPI - Human.pin	8,776	35,820	0.078	263.0	**3366x**
Wiki Vote	7,115	103,689	0.266	137.0	**516x**
Gnutella P2P	10,874	39,994	0.172	681.7	**3966x**
AstroPh Collab	18,772	396,160	0.344	2,218.7	**6454x**

5   Future Work

One facet of this work that we hope to discover more about in the future is the scalability of the CUDA Super Fruchterman-Rheingold layout algorithm. We were fortunate to have large datasets available from the SNAP library, but given the 666x speedup obtained in the largest dataset tested, which contains 82,163 nodes, we have yet to determine the limits of the algorithm. It is clear that given the restricted amount of memory on today's GPUs (about 1 GB on high-end cards) that eventually memory will become a bottleneck, and we have yet to determine where that limit occurs. There are obviously interesting networks that contain millions of nodes, and we hope to be able to provide significant speedup for those situations as well. Unfortunately, the Windows Presentation Foundation code which paints a graph on the visualization window proved to be the weakest link in the visualization of multi-million node

graphs. On large enough graphs, the WPF module will generate an out-of-memory error or simply not terminate. Overcoming this WPF shortcoming would be the first step towards finding the bounds of the Super Fruchterman-Rheingold algorithm.

Additionally, there are other algorithms within NodeXL that could benefit from GPU-based parallelization. The Harel-Koren Fast Multiscale algorithm [2] is another heavily-used layout algorithm in data visualization that is currently implemented sequentially. Finally, the speedup achieved using the eigenvector centrality metric also bring other vertex metrics to our attention. In particular, *Closeness* and *Betweenness* centrality are two measures that are used frequently in data visualization, and have yet to be implemented using CUDA.

6   Conclusions

NodeXL is a popular data visualization tool used frequently by analysts on typical desktop hardware. However, datasets can be thousands or millions of nodes in size, and a typical desktop CPU does not contain enough processing cores to parallelize the various algorithms contained in NodeXL sufficiently to execute in a reasonable timeframe. Nvidia's CUDA technology provides an intuitive computing interface for users to use the dozens of processing elements on a typical desktop GPU, and we applied the technology to two algorithms in NodeXL to show the benefit that the application can gain from these enhancements.

The results obtained from running these enhancements were impressive, as we observed a 79x - 804x speedup in our implemented layout algorithm. Further, the 102x - 17,972x speedup gained by parallelizing the Eigenvector centrality metric across GPU cores indicates that when a CUDA-capable GPU is available, it should be used to perform centrality metrics. Such improvement in these execution times brings the visualization of data previously infeasible on a standard desktop into the hands of anyone with a CUDA-capable machine, which includes low-to-middle end GPUs today. While we have by no means maxed out the parallelization possibilities in the NodeXL codebase, we hope to have provided enough preliminary results to demonstrate the advantage that CUDA technology can bring to data visualization.

References

1. Fruchterman, T., Reingold, E.: Graph Drawing by Force-directed Placement. Software – Practice And Experience 21(11), 1129–1164 (1991)
2. Harel, D., Koren, Y.: A Fast Multi-Scale Algorithm for Drawing Large Graphs. Journal of Graph Algorithms and Applications 6(3), 179–202 (2002)

3. Hansen, D., Shneiderman, B., Smith, M.: Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, San Francisco
4. Perer, A., Shneiderman, B.: Integrating Statistics and Visualization: Case Studies of Gaining Clarity During Exploratory Data Analysis. In: ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 265–274 (2008)
5. Harish, P., Narayanan, P.J.: Accelerating large graph algorithms on the GPU using CUDA. In: Proc. 14th International Conference on High Performance Computing, pp. 197–208 (2007)
6. Centrality – Wikipedia, `http://en.wikipedia.org/wiki/Centrality`
7. Hadoop, `http://wiki.apache.org/hadoop/`
8. Social Network Analysis, `http://en.wikipedia.org/wiki/Social_network#Social_network_analysis`
9. Sharma, P., Khurana, U., Shneiderman, B., Scharrenbroich, M., Locke, J.: Speeding up Network Layout and Centrality Measures with NodeXL and the Nvidia CUDA Technology. Technical report, Human Computer Interaction Lab, University of Maryland College Park (2010)

An Agent-Based Simulation for Investigating the Impact of Stereotypes on Task-Oriented Group Formation

Mahsa Maghami and Gita Sukthankar

Department of EECS
University of Central Florida
Orlando, FL
mmaghami@cs.ucf.edu,gitars@eecs.ucf.edu

Abstract. In this paper, we introduce an agent-based simulation for investigating the impact of social factors on the formation and evolution of task-oriented groups. Task-oriented groups are created explicitly to perform a task, and all members derive benefits from task completion. However, even in cases when all group members act in a way that is locally optimal for task completion, social forces that have mild effects on choice of associates can have a measurable impact on task completion performance. In this paper, we show how our simulation can be used to model the impact of stereotypes on group formation. In our simulation, stereotypes are based on observable features, learned from prior experience, and only affect an agent's link formation preferences. Even without assuming stereotypes affect the agents' willingness or ability to complete tasks, the long-term modifications that stereotypes have on the agents' social network impair the agents' ability to form groups with sufficient diversity of skills, as compared to agents who form links randomly. An interesting finding is that this effect holds even in cases where stereotype preference and skill existence are completely uncorrelated.

Keywords: Group formation, Multi-agent social simulations, Stereotypes.

1 Introduction

Group membership influences many aspects of our lives, including our self-identities, activities, and associates; it affects not only what we do and who we do it with, but what we think of ourselves and the people around us. It can also give rise to stereotypic thinking in which group differences are magnified and importance of individual variations are discounted. Thinking categorically about people simplifies and streamlines the person perception process [12], facilitating information processing by allowing the perceiver to rely on previously stored knowledge in place of incoming information [11]. Whereas a variety of stereotypes are based on real group differences (e.g., cultural stereotypes about food preferences), stereotypes based on relatively enduring characteristics, such as race, religion, and gender, have an enormous potential for error [11] and can give rise to performance impairments [13]. We hypothesize that when stereotypes are formed independent of real group differences, it can result in negative effects for the collective system. However, studying the long-term effects of stereotypes can be difficult, especially to quantify the effects over a population rather than an individual. In this paper,

we describe an agent-based simulation for evaluating the impact of stereotypes on the performance of task-oriented groups. Understanding the role that stereotypes play in group formation can refine existing theory while providing insight into the efficacy of methods for combating the adverse effects of stereotypes on behavior and decision-making.

We base our simulation on a model of multi-agent team formation [6] since task-oriented groups share many characteristics with teams, although lacking in shared training experience. In our simulation, the population of agents is connected by a social network that is locally updated over time by unoccupied agents depending on their preferences. Stereotypes are represented as an acquired preference model based on prior experience and observable agent features. In multi-agent environments, stereotypes have been used to provide faster modeling of other agents [4,5] and to bootstrap the trust evaluation of unknown agents [2]. In contrast, we examine the case of non-informative stereotypes; stereotypes affect the agents' preferences for forming social attachments but do not affect the agents' willingness or ability to cooperate with other agents.

2 Problem Statement

To explore the impact of stereotype on group formation and network evolution, we have selected a simple multi-agent system model first introduced by Gaston and des-Jardins [6] and used in [7,8] to describe team formation. Since task-oriented groups are similar to teams, this is a reasonable method for modeling the task performance of group behavior on shared utility tasks in absence of stereotypes. In this model, there is a population of N agents represented by the set $A = \{a_1, \ldots, a_N\}$. Each agent can be considered as a unique node in the social network and the connection between the agents is modeled by an adjacency matrix E, where $e_{ij} = 1$ indicates an undirected edge between agent a_i and a_j and the degree of agent a_i is defined as $d_i = \sum_{a_j \subseteq A} e_{ij}$. Each agent is assigned a single skill given by $\sigma_i \in [1, \sigma]$ where σ is the number of available skills. Accomplishing each task requires a coalition of agents with the appropriate skills. Tasks are globally advertised for γ time steps at fixed intervals μ. Each task, T_k, has a size requirement, $|T_k|$, and a $|T_k|$-dimensional vector of required skills, R_{T_k}, which are selected uniformly from $[1, \sigma]$. When a coalition has formed with the full complement of skills required for the task, it takes α time steps for the group to complete the task.

2.1 Group Formation

In the no stereotype case, we simply follow the group formation algorithm used to allocate agents to teams in [7].

During the team formation process, each agent, a_i, can be in one of three states, s_i, UNCOMMITTED, COMMITTED, or ACTIVE. An agent in the UNCOMMITTED state has not been assigned to any task. An agent in the COMMITTED state has been assigned to a task but is still waiting for the enough agents with the right skills to join the group. Finally, an ACTIVE agent is currently working on a task with a complete group possessing the right complement of skills.

On each iteration, agents are updated in random order. UNCOMMITTED agents have the opportunity to adapt their local connectivity (with probability of P_i) or can attempt to join a group. If a task currently has no other agents committed to it, an agent can initiate a group with a probability that is proportional to the number of its immediate UNCOMMITTED neighbors defined as follows:

$$IP_i = \frac{\sum_{a_j \subseteq A} e_{ij} I(s_i, \text{UNCOMMITTED})}{\sum_{a_j \subseteq A} e_{ij}}, \tag{1}$$

where $I(x, y) = 1$ when $x = y$ and 0 otherwise.

Agents are only eligible to join task-oriented groups in their local neighborhood, where this is at least one link between the agent and the group members. The algorithm used by an agent to initiate or join a group is presented in Algorithm 1.

Algorithm 1. Group formation algorithm

for all $T_K \subseteq T$ **do**
    if $|M_k| = 0$ and s_i = UNCOMMITTED **then**
        $r \leftarrow UniformRandom([0, 1])$
        if $r \leq P_i$ **then**
            if $\exists r \in R_{T_k} : r = \sigma_i$ **then**
                $M_k \leftarrow M_k \cup \{a_i\}$
                $s_i \leftarrow$ COMMITTED
            end if
        end if
    else if $\exists a_j : e_{ij} = 1, a_j \in M_k$ and s_i = UNCOMMITTED **then**
        if $\exists r \in R_{T_k} : r = \sigma_i$ and r is unfilled **then**
            $M_k \leftarrow M_k \cup \{a_i\}$
            $s_i \leftarrow$ COMMITTED
        end if
    end if
end for

2.2   Network Adaptation

To adapt the network structure, the agents modify their local connectivity based on the notion of preferential attachment [1]. Therefore, the probability of connecting to a given node is proportional to that node's degree. As mentioned before, at each iteration the agent can opt to adapt its connectivity, with probability P_i. Modifying its local connectivity does not increase the degree of the initiating agent since the agent severs one of its existing connections at random and forms a new connection.

To form a new connection, an agent considers the set of its neighbors' neighbors designated as $N_i^2 = \{a_m : e_{ij} = 1, e_{jm} = 1, e_{im} = 0, m \neq i\}$. The adapting agent, a_i, selects a target agent, $a_j \subseteq N_i^2$, to link to based on the following probability distribution:

$$P(a_i \longrightarrow a_j) = \frac{d_j}{\sum_{a_l \subseteq N_i^2} d_l}, \tag{2}$$

where d is the degree of agents.

The results in [7] and [8] show that this simple algorithm can be used to adapt a wide variety of random network topologies to produce networks that are efficient at information propagation and result in scale-free networks similar to those observed in human societies. For the non-stereotype group formation condition, we did not observe any differences between random attachment and preferential attachment.

3   Learning the Stereotype Model

As noted in a review of the stereotype literature [11], stereotypes are beliefs about the members of a group according to their attributes and features. It has been shown that the stereotypes operate as a source of expectancies about what a group as a whole is like as well as what attributes individual group members are likely to possess [9]. Stereotype influences can be viewed as a judgment about the members of a specific group based on relatively enduring characteristics rather than their real characteristics.

Here, we do not attempt to capture the rich cognitive notions of stereotype formation as it applies to humans, but rather view a stereotype as a function $\mathcal{F} : \overrightarrow{V} \longrightarrow S$, mapping a feature vector of agents, $\overrightarrow{V}$, to a stereotypical impression of agents in forming friendships, S, which we will designate as the stereotype value judgment. This value represents the agents' judgments on other groups and is based on observable features rather than skills or prior task performance.

In most contexts, humans possess two types of information about others: 1) information about the individual's attributes and 2) the person's long-term membership in stereotyped groups [9]. Therefore, to learn the stereotype model, the simulation offers these two sources of information, $\overrightarrow{V}$ and its corresponding S which are related to the agents' group membership, for a specific period of time. In our simulation, this initial period lasts for I time steps and helps the collaborating agents gain experience about the attributes of different groups of agents. Note that membership in these groups is permanent and not related to the agent's history of participation in short-term task-oriented groups.

In our work, we propose that each agent, a_i, can use linear regression to build its own function, $\mathcal{F}_i$, and to estimate the stereotype value of another agent, a_j, according to the observable features $\overrightarrow{V_j}$. After the initial period, I time steps, the estimated stereotype value of agent a_j by agent a_i will be calculated as $\hat{S_{ij}} = \mathcal{F}_i(\overrightarrow{V_j})$.

In our model, this stereotype value judgment affects the connection of agents during the network adaptation phase, as we will describe in the following section.

3.1   Network Adaptation with Stereotype Value Judgments

In the stereotype case, the group formation algorithm is the same as described in Algorithm 1 but the network adaptation is based on the learned stereotype. If an agent decides to adapt its local network, again with probability P_i , it will do so based on its own stereotype model. To adapt the local connectivity network, each agent uses its learned model to make stereotype value judgment on other neighboring agents. This network adaptation process consists of selecting a link to sever and forming a new link.

Specifically, the agent a_i first searches for its immediate neighbor that has the lowest stereotype value judgment, a_j, and severs that link. The agent then searches for a new agent as a target for link formation. To form this link, it searches its immediate neighbors and the neighbors of neighbors. First the agent selects the neighbor with the highest stereotype value judgment, a_m, for a referral as this agent is likely to be a popular agent in its neighborhood. Then the adapting agent, a_i, will establish a new connection with a_n, one of the most popular neighbors of a_m, assuming that it is not already connected.

$$a_n = \arg_{a_k \in A, e_{ik}=0} \max \hat{S_{ik}}.$$

Note that all of these selections are the result of the stereotype value judgment model that agent a_i has about the other agents in its neighborhood.

4   Evaluation

4.1   Experimental Setup

We conducted a set of simulation experiments to evaluate the effects of stereotype value judgments on group formation in a simulated society of agents. The parameters of the group formation model for all the runs are summarized in Table 1. In task generation, each task is created with a random number of components less than or equal to ten and a vector of uniformly-distributed skill requirements. To generate the agent society, each agent is assigned a specific skill, a feature vector, and a class label. The agents' skills are randomly generated from available skills. Inspired by [2], four different long-lasting groups with different feature vector distributions are used as the basis for stereotype value judgments. Agents are assigned a six-dimensional feature vector, with each dimension representing an observable attribute, and a hidden stereotype value judgment drawn from Gaussian distribution assigned to the group. Table 1(b) shows the mean and standard deviations of the Gaussian distributions and the observable feature vector assigned to each group. The observable feature vectors are slightly noisy. To indicate the existence of an attribute, a random number is selected from distribution $N(0.9, 0.05)$ to be close to 1 and to indicate the lack of an attribute this number is selected from distribution $N(0.1, 0.05)$ to be close to zero. During the initial training period, agents are allowed to observe the hidden stereotype value judgment to learn the classifier. During the rest of the simulation these values are estimated by the agent's learned model.

In these experiments all the runs start with a random geometric graph (RGG) as the initial network topology among the agents. A RGG is generated by randomly locating all the agents in a unit square and connecting two agents if their distance is less than or equal to a specified threshold, d [3]. The random network we generated is a modified version of the RGG, proposed by [7]. In this version the d is selected as a minimal distance among the agents to guarantee that all the agents have at least one link to other agents.

When the initial network is generated, the group formation will run for an initial period with no adaptation. During these initial steps, the agents can form groups and participate in task completion to gain some experiences about other agents. Results are based on the average of 10 different runs with a different initial network for each run.

(a) Experimental parameters		
Parameter	Value	Descriptions
N	100	Total number of agents
σ	10	Total number of skills
γ	10	Task advertising time steps
α	4	Agents active time
μ	2	Task interval
$|T|$	max 10	Number of task required skills
$N_{Iterations}$	8000	Number of iterations
$N_{Initial}$	2000	Number of initial iterations

(b) Stereotype groups and feature vectors								
Group	Mean Value	StDev	f_1	f_2	f_3	f_4	f_5	f_6
G1	0.9	0.05	X					X
G2	0.6	0.15		X		X		
G3	0.4	0.15			X	X		
G4	0.3	0.1			X	X		X

4.2 Results

Performance. The performance of the system, like [7], is calculated as follows:

$$Performance = \frac{T_{SuccessfullyDone}}{T_{total}}, \tag{3}$$

which is the proportion of successfully accomplished tasks over the total number of introduced tasks in the environment. Figure 1 shows the performance of system in the system with stereotypes and without stereotypes by iteration.
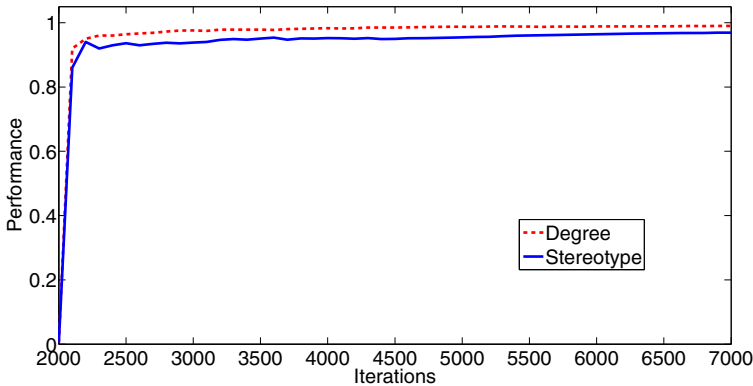
Fig. 1. The performance of task-oriented groups (with and without stereotypes) vs. the iterations

The main effect of the stereotype is to move network toward a sparse network structure with a dramatic increase in isolate nodes. This drop in performance is even more pronounced with fewer total agents.

Network Structure. Here, we examine the network structure to determine the evolution of the agent society. Figure 2 shows the Fiedler-embedding [10] of networks in the final connectivity network of $N = 200$ agents with and without stereotype value judgments. The degree-based strategy moves the structure toward being similar to a scale-free network whereas with stereotype value judgments the network becomes progressively more star-shaped.
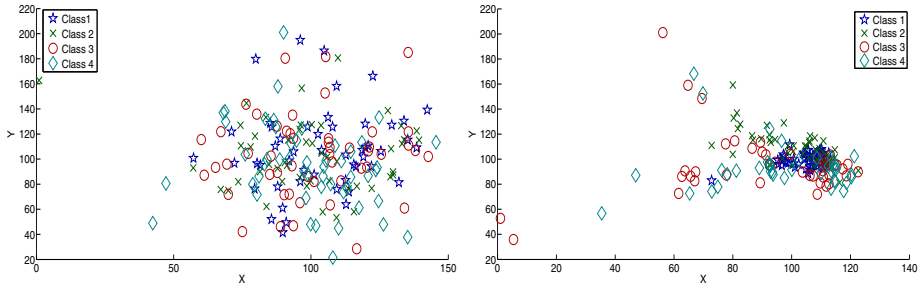
Fig. 2. Fiedler embedding of the final network structures in non-stereotype (left) and stereotype (right) based network evolution ($N = 200$)

Effects of Rapid Attachment Modification. Here we examine the effects of modifying the parameter P, the probability of updating the network, on the performance of the system, both with and without stereotypes. We varied this parameter from 0.1 to 0.9 with a step size 0.2. Figure 3 shows the performance modification during 5000 iterations in both strategies. As shown in the figure, the performance does not change significantly with P values before a certain threshold. After that threshold, the performance drops dramatically, as the agents spend more time updating the network than accomplishing tasks. This threshold is dependent on the total agents and number of skills required in the environment. In both strategies, the performance has dropped by $P = 0.7$ but in stereotype strategy the system drops at an earlier iteration since the information transmission efficiency of the network has been sabotaged by the stereotype-value judgments.
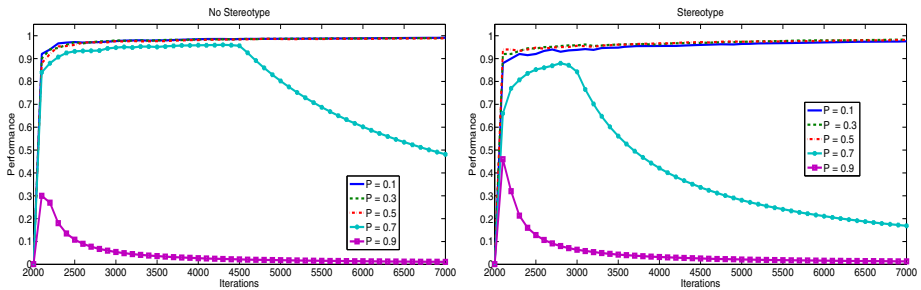
Fig. 3. The effect of the network adaptation probability, P_i

5   Conclusion and Future Work

In this paper we introduced an agent-based simulation for examining the effects of stereotypes on task-oriented group formation and network evolution. We demonstrate that stereotype value judgments can have a negative impact on task performance, even in the mild case when the agents' willingness and ability to cooperate is not impaired. By modifying the social network from which groups are formed in a systematically suboptimal way, the stereotype-driven agents eliminate the skill diversity required for

successful groups by driving the network toward specific topological configurations that are ill-suited for the task. The results show that making connections with agents solely based on group membership yields a sparser network with many isolated nodes. In future work, we plan to explore the effects of different stereotype models on network topology. Additionally, we plan to incorporate EPA (Evaluation, Potency, and Accuracy) [14] style parameters into our stereotype judgment model. Due to the technical challenges of investigating the long-term effects of stereotype across populations, we suggest our agent-based simulation method is a useful tool for investigating these research questions.

Acknowledgments

References

1. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Reviews of Modern Physics 74(1), 47–97 (2002)
2. Burnett, C., Norman, T., Sycara, K.: Bootstrapping trust evaluations through stereotypes. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, pp. 241–248 (2010)
3. Dall, J., Christensen, M.: Random geometric graphs. Physical Review E 66(1), 16121 (2002)
4. Denzinger, J., Hamdan, J.: Improving modeling of other agents using tentative stereotypes and compactification of observations. In: Proceedings of Intelligent Agent Technology, pp. 106–112 (2005)
5. Denzinger, J., Hamdan, J.: Improving observation-based modeling of other agents using tentative stereotyping and compactification through kd-tree structuring. Web Intelligence and Agent Systems 4(3), 255–270 (2006)
6. Gaston, M., Simmons, J., desJardins, M.: Adapting network structure for efficient team formation. In: Proceedings of the AAAI 2004 Fall Symposium on Artificial Multi-agent Learning (2004)
7. Gaston, M., desJardins, M.: Agent-organized networks for dynamic team formation. In: Proceedings of the International Conference on Autonomous agents and Multiagent Systems, pp. 230–237 (2005)
8. Glinton, R., Sycara, K., Scerri, P.: Agent-organized networks redux. In: Proceedings of AAAI, pp. 83–88 (2008)
9. Hamilton, D., Sherman, S., Ruvolo, C.: Stereotype-based expectancies: Effects on information processing and social behavior. Journal of Social Issues 46(2), 35–60 (1990)
10. Hendrichson, B.: Latent semantic analysis and Fiedler embeddings. In: Proceedings of the Fourth Workshop on Text Mining of the Sixth SIAM International Conference on Data Mining (2006)
11. Hilton, J., Von Hippel, W.: Stereotypes. Annual Review of Psychology 47 (1996)
12. Macrae, C., Bodenhausen, G.: Social cognition: Thinking categorically about others. Annual Review of Psychology 51(1), 93–120 (2000)
13. Marx, D.M.: On the role of group membership in stereotype-based performance effects. Journal of Social and Personality Psychology Compass 3(1), 77–93 (2009)
14. Osgood, C.: The measurement of meaning. Univ. of Illinois, Urbana (1975)

An Approach for Dynamic Optimization of Prevention Program Implementation in Stochastic Environments

Yuncheol Kang and Vittal Prabhu

Department of Industrial and Manufacturing Engineering,
Pennsylvania State University, University Park, PA 16802, USA

Abstract. The science of preventing youth problems has significantly advanced in developing evidence-based prevention program (EBP) by using randomized clinical trials. Effective EBP can reduce delinquency, aggression, violence, bullying and substance abuse among youth. Unfortunately the outcomes of EBP implemented in natural settings usually tend to be lower than in clinical trials, which has motivated the need to study EBP implementations. In this paper we propose to model EBP implementations in natural settings as stochastic dynamic processes. Specifically, we propose Markov Decision Process (MDP) for modeling and dynamic optimization of such EBP implementations. We illustrate these concepts using simple numerical examples and discuss potential challenges in using such approaches in practice.

Keywords: Markov Decision Process, Partially Observable Markov Decision Process, Evidence-based Prevention Program.

1   Introduction

The science of preventing youth problems has significantly advanced in developing evidence-based prevention program (EBP) by using randomized clinical trials [1]. Effective EBP can reduce delinquency, aggression, violence, bullying and substance abuse among youth. For instance, studies have shown the annual cost of substance abuse in the U.S. to be $510.8 billion in 1999, and effective school-based prevention programs could save $18 for every $1 spent on these programs [2]. Some of the widely used EBPs include PATHS (Promoting alternative Thinking Strategies), BBBS (Big Brothers Big Sisters), SFP (Strengthening Families Program) and LST (LifeSkills Training). EBP services are commonly delivered to youth, and sometimes their families, individually or in groups through school and community settings, i.e., the natural world settings. Usually the outcomes of EBP implemented in natural settings tend to be lower than in the clinical trials [3-5]. One of the main challenges in improving the outcomes of EBP implementations is to develop a better understanding of the complex interplay between EBP and the natural world by identifying the key factors that influence the outcomes and the manner in which these factors and the interplay evolve during an implementation [3]. Recently adaptive intervention dosage control in EBP has been proposed as a potential method to improve the outcomes while reducing the cost [6-7]. Sustaining effective EBP services under unknown and

uncertain dynamics with significant budget and capacity constraints presents an intellectual challenge and is an acute social imperative.

In this article, we present an approach for optimizing EBP implementations in stochastic dynamic environments. Our presentation is especially focused on the Markov Decision Process (MDP) approach for dynamically optimizing EBP implementations. We review MDP approach, and illustrate its potential application using numerical examples. Lastly, we discuss the challenges in applying the MDP approach for EBP implementations in practice.

2   Dynamic Optimization in Stochastic Environments

Many situations require a series of decisions to be made sequentially over time to optimize some overall objective. These problems are characterized by the need to consider the objective over the entire time horizon rather than the immediate reward at any one time step. A wide range of such problems in engineering and economics have been solved using dynamic optimization approaches since the pioneering work of Bellman in the 1950s [8]. In many situations, changes in the external and exogenous environment randomly influences the outcomes, and the decisions made at various time steps has some influence on the outcomes. Such stochastic dynamic optimization problems can be modeled as Markov Decision Processes (MDP). MDP models have been widely used in a variety of domains including management, computer systems, economics, communication, and medical treatment [9]. The "randomness" in MDP models can be caused by incomplete understanding of the process itself, or by uncertainties in interactions between the process and the environment. MDP is a variant and extended type of Markov Chain, which is a stochastic process having the property such that the next state depends on only the current state. MDP modeling approach is best suited for situations in which there are multiple actions which we can control over a Markov chain.

To construct an MDP model, we need to define the following: *state*, *action, decision epoch*, and *reward*. First, a *state* is a random variable that represents the evolution of the system at any given time. As the system evolves it transitions from one state to another. Meanwhile, an *action* indicates a possible decision in the system, which in turn influences the current state, and transition to another state. A *decision epoch* refers to the discrete time when we take an action, and we need to determine the length of the *decision horizon*. The solution to an MDP problem enables us to choose a set of actions, also known as a *policy*, which maximizes the targeted value, usually called a *reward*. In other words, a reward occurs when an action is taken at the current state, and the algorithm for solving MDP results in the policy maximizing reward over the decision horizon.

After an MDP model is constructed, we can obtain an optimal policy from the optimality equation, and corresponding structural results of the dynamics in the model. For the optimal policy, a number of algorithms have been introduced such as *value iteration*, *policy iteration* or *linear programming* [9]. Several approaches also have been developed to reduce computational burden in solving an MDP problem [10].

2.1   Modeling EBP Implementation as MDP

In this paper we propose that EBP implementations can be modeled as dynamic optimization problems in which a sequence of decisions are made over multiple time periods through which the best possible objective is obtained. More particularly, we propose that EBP implementations can be modeled as an MDP. For illustration, let us consider one of the widely used EBP, PATHS (**P**romoting **A**lternative **TH**inking **S**trategies). PATHS is a school-based program for promoting emotional and social competencies and reducing aggression and behavior problems in elementary school-aged children while simultaneously enhancing the educational process in the classroom [11]. PATHS is a "universal" program in the sense it is delivered to all the youth in the class. PATHS curriculum provides teachers with systematic, developmentally-based lessons, materials, and instructions.

There are two major stochastic processes in PATHS implementations. The first major stochastic process is various dynamic transitions in a youth's mental state. We regard this as a stochastic process from a modeling perspective because the actual learning process in a specific youth is too complex, and current science cannot deterministically predict the social and emotional changes in the youth caused by PATHS. The second major stochastic process is the dynamically changing environment around the youth. The net combined effect of PATHS and the environment is therefore modeled as causing random transitions in a youth's mental *state*, $s \in S$. A crucial aspect of the model is to quantify the probability of transitioning from one state to another. A state can be defined as a scalar or vector quantity depending on the intended purpose and richness of the model. For example, in a PATHS implementation, a youth's Social Health Profile (SHP) such as authority acceptance, cognitive concentration and social competence can be used to represent the state vector for the youth, which can be measured through periodic observations by teacher(s) and others [12]. A state could also include a youth's aggressive, hyperactive-disruptive, and prosocial behaviors measured through peer reports [12]. Depending on the purpose of the model, a state can be vector or a scalar index that quantifies the youth's social and emotional state. Instead of modeling at the individual youth level we could also consider at a higher level of aggregation such as a class room or school. The set of all the states and the transition probabilities can be used to derive statistical properties of the system's future.

Possible *actions*, $a \in A$, for controlling PATHS implementation could be improving the dosage and fidelity of the implementation, providing additional technical assistance to the teacher(s) delivering the program, or providing additional intensive intervention to high risk youth. The premise of MDP modeling here is that the actions taken will influence the dynamic state transitions even though the actions themselves cannot completely determine the outcome. The menu of actions that are available will depend on the available budget and resources for the specific implementation in practice. A series of such actions are taken at each *decision epoch*, t, over the decision horizon, N. Decision epoch in PATHS could be every month or every academic semester, and can vary depending on the specific implementation. The decision horizon can be an academic year or multiple years depending on the planning horizon of the particular implementation. There is a quantitative *reward*, $r_t(s,a)$, corresponding to each state and action. In EBP implementations reward can be a

combination of costs and benefits for the state and action. In particular, the cost can be regarded as a negative reward while the benefit can be treated as a positive reward. The core problem of MDP is to select the policy, $d_t^*(s)$, to maximize the reward over the planning horizon by solving the following *optimality equation*.

$$d_t^*(s) = arg_{a \in A} max[r_t(s, a) + \sum_{s' \in S} P(s'|s, a) u_{t+1}^*(s')].$$

$$\text{for } t = 1, ..., N - 1$$

$$(1)$$

where s' is the state at $t + 1$, $P(s'|s, a)$ is the transition probability from s to s' when the action a is taken at t and $u_{t+1}^*(s')$ is the expected future reward.

2.2   Advanced Modeling of Stochastic Processes in EBP Implementations

As prevention science unravels the "black box" model of the youth's mental processes we would be able to refine and improve the stochastic process models proposed above. A key issue that is likely to remain with us is that the youth's mental state will not be directly accessible. Any observations by teachers, parents, peers or self will at best be a probabilistic estimate of the true mental state. We can model such situations using a special case of MDP approach called Partially Observable Markov Decision Process (POMDP). In POMDP, the actual states are considered as hidden and are not directly accessible to observers. Rather, the observers have a set of observable outputs, $o \in O$, probabilistically related to the actual states and a set of belief, π, based on the observations [13]. The belief is updated using a *belief* update function, $B(\pi'|o, \pi, a)$. Thus, POMDP selects the policy, $d_t^*(s)$, to maximize the reward based upon beliefs of the states by solving the following equation.

$$d_t^*(\pi) = arg_{a \in A} max[r_t(\pi, a) + \sum_{o' \in O} P(o'|\pi, a) u_{t+1}^*(B(\pi'|o, \pi, a))].$$

$$\text{for } t = 1, ..., N - 1$$

$$(2)$$

where o' and π' is the observation and the belief at $t + 1$, $P(o'|\pi, a)$ is the probability of next observation based on the current belief and action. Since the current action influences the actual state transition, the probability of next observation is implicitly based on the state transition caused by current action chosen. $u_{t+1}^*(B(\pi'|o, \pi, a))$ is the expected future reward.

POMDP approaches have been used for modeling diseases that cannot be directly observed in patient's symptoms [14-15]. EBP implementations could be modeled as POMDP in which teachers observation report is used for probabilistic estimation of the youth's true mental state using belief functions. The belief function in POMDP would be an attractive mechanism to model the dynamics of a youth's social and emotional development trajectory for a given combination of EBP and social environment. While POMDP offers a more advanced modeling approach to deal with stochastic dynamics of EBP implementations, they are computationally much harder to solve compared to MDP.

# 3   Illustration of MDP Model for EBP Implementation

In this section we illustrate stochastic models for EBP implementations using two simple numerical examples. The first example illustrates the stochastic dynamics of EBP and the second example indicates how such dynamics can be potentially influenced by decision making using a MDP model. It should be emphasized that these examples are not based on real data.

3.1   Illustration 1: Stochastic Dynamics of EBP

Let us consider the Social Competence (SC) measure in SHP as the state for this illustration. Fig. 1 graphically represents the Markov Chain that models the stochastic dynamics using state transition probabilities without any prevention intervention. Fig. 2 represents the Markov Chain model with prevention intervention. These stochastic models can be used for a variety of analysis such as predicting the long-run social competence by computing the steady state probabilities, which is illustrated in Fig. 3. We can also use this to readily estimate long-run reward by using cost and/or benefit of each state.
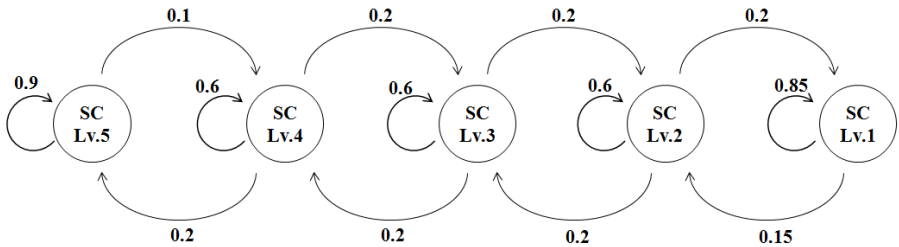
Fig. 1. Dynamics of SC without prevention intervention. Each circle represents the state indicating the level of SC of the individual participant. The level of SC in Fig. 1 and 2 is abbreviated as *Lv*. The higher level of SC represents the better outcome in terms of social competence. Arcs are the transition between the states, and the number on the arc is the probability of the transition.
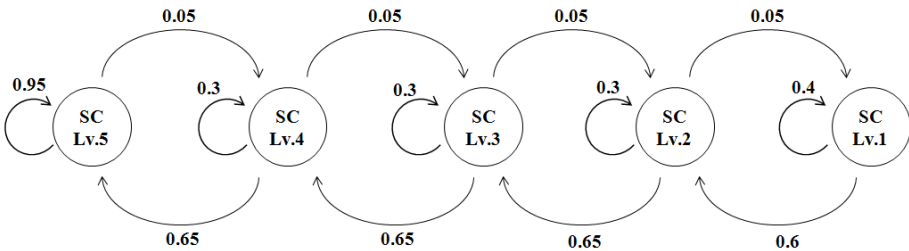
Fig. 2. Dynamics of SC with prevention intervention. Note that in this model there is higher probability of transitioning to a better SC state and compared to Fig. 1.

3.2   Illustration 2: Decision Makings in EBP Implementation Using MDP

The Markov Chain models in the previous illustration can be extended for optimizing decision making in EBP implementations using an MDP model. To illustrate this we use the transition probabilities shown in Fig. 2. Additionally, suppose the cost of the prevention intervention is 7 units for each decision epoch, and we assign 0, 13, 15, 18, 20 units as the amount of benefit from SC level 1 to level 5, respectively. We define reward for each pair of state and action, and calculate it by summing the corresponding cost and benefit. Since MDP chooses the best action for each decision epoch, we can make the best decision in terms of maximizing cumulative reward from the current decision epoch onwards.
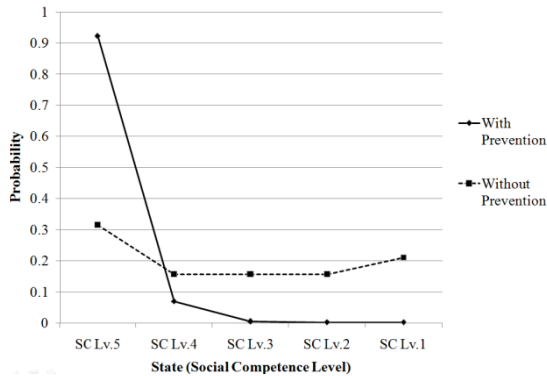
Fig. 3. Steady State Probability for scenarios with and without prevention intervention
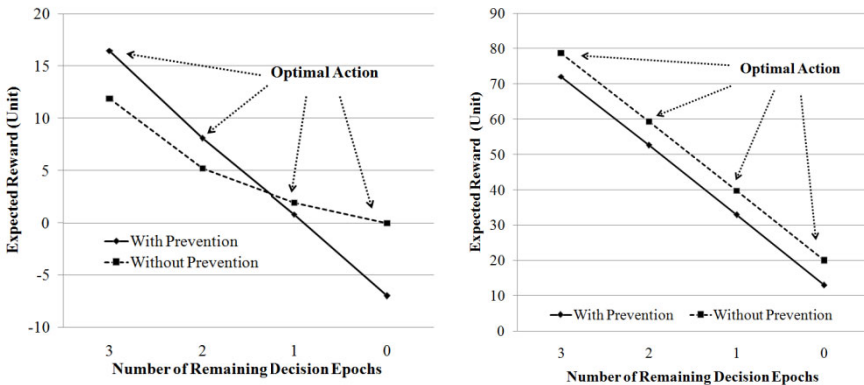

Fig. 4. Optimal Action   in SC level 1(*left panel*) and   SC level 5 (*right panel*)

Fig. 4 shows two different scenarios of decision makings for the lowest SC level, SC level 1(*left panel*) and for the highest SC level, SC level 5 (*right panel*). The left panel shows prevention intervention results in a higher reward if it is delivered for

more than two decision epochs in case of the lowest SC level. On the contrary, if the current state is in the highest SC level, then *without prevention* is always more beneficial than the other in terms of reward (*right panel* in Fig. 4). In many situations in EBP implementation, one often faces with considering a trade-off between the benefit and cost of prevention intervention. Under the circumstance, an MDP model would provide a guideline for indicating whether the cost of prevention intervention is enough to justify the benefit as this example describes.

4   Challenges in Applying MDP Approach

MDP modeling of prevention implementation could help researchers and practitioners obtain new insights into underlying dynamics and cost-effectiveness of prevention interventions. However, there are several challenges in using the MDP approach for EBP. The first challenge is to identify the structure and parameters of the MDP model such as the appropriate definition of the state and estimation of state transition probabilities. In particular, state transition probabilities and reward should reflect the characteristics of the EBP implementation with adequate accuracy. A related issue is to assign transition probabilities from sources such as prior randomized clinical trials of EBP. It should be noted that the MDP approach provides the optimal policy which maximizes the reward consisting of cost and benefit. Therefore the usefulness of the approach will be limited by accuracy of the cost and benefit estimates, especially in any monetization. Also, POMDP models can become computationally intractable even for modest sized problems and are likely to require carefully crafted solution techniques especially suitable for EBP implementation [9-10].

5   Conclusion

The need to improve the effectiveness of EBP in natural world settings has motivated the need to study EBP implementations. There is a dearth of models that mathematically characterize the dynamics of EBP processes at the individual or group levels - there is a pressing need to unravel this black box [16]. In this paper we proposed Markov Decision Process (MDP) for modeling and dynamic optimization of such EBP implementations. We illustrated these concepts using simple numerical examples and discussed potential challenges in using such approaches in practice. The modeling approaches proposed in this paper could help advance Type II translational research to create a new science of adoption and implementation of EBP in the real world and enable their ultimate goal of wide-spread public health impact [5].

Acknowledgements

References

1. Greenberg, M., Domitrovich, C., Bumbarger, B.: Preventing mental disorders in school-age children: A review of the effectiveness of prevention programs. Prevention Research Center for the Promotion of Human Development, Pennsylvania State University. Submitted to Center for Mental Health Services, Substance Abuse and Mental Health Services Administration, US Department of Health & Human Services (1999)

2. Miller, T., Hendrie, D.: Substance abuse prevention dollars and cents: A cost-benefit analysis. In: Services, US Department of Health & Human Services. (ed.), p. 59 (2009)

3. Greenberg, M.: Current and future challenges in school-based prevention: The researcher perspective. Prevention Science 5, 5–13 (2004)

4. Bumbarger, B., Perkins, D.: After randomised trials: issues related to dissemination of evidence-based interventions. Journal of Children's Services 3, 55–64 (2008)

5. Bumbarger, B.K., Perkins, D.F., Greenberg, M.T.: Taking effective prevention to scale. In: Doll, B., Pfohl, W., Yoon, J. (eds.) Handbook of Prevention Science. Routledge, New York (2009)

6. Collins, L., Murphy, S., Bierman, K.: A conceptual framework for adaptive preventive interventions. Prevention Science 5, 185–196 (2004)

7. Rivera, D., Pew, M., Collins, L.: Using engineering control principles to inform the design of adaptive interventions: A conceptual introduction. Drug and Alcohol Dependence 88, S31–S40 (2007)

8. Bellman, R., Lee, E.: History and development of dynamic programming. Control Systems Magazine 4, 24–28 (2002)

9. Puterman, M.: Markov decision processes: Discrete stochastic dynamic programming. John Wiley & Sons, Inc., New York (1994)

10. Schaefer, A., Bailey, M., Shechter, S., Roberts, M.: Modeling medical treatment using Markov decision processes. Operations Research and Health Care, 593–612 (2005)

11. Greenberg, M., Kusche, C., Cook, E., Quamma, J.: Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. Development and Psychopathology 7, 117–136 (1995)

12. Bierman, K., Coie, J., Dodge, K., Greenberg, M., Lochman, J., McMahon, R., Pinderhughes, E.: The Effects of a Multiyear Universal Social-Emotional Learning Program: The Role of Student and School Characteristics. Journal of Consulting and Clinical Psychology 78, 156–168 (2010)

13. Smallwood, R., Sondik, E.: The optimal control of partially observable Markov processes over a finite horizon. Operations Research 21, 1071–1088 (1973)

14. Hauskrecht, M., Fraser, H.: Planning treatment of ischemic heart disease with partially observable Markov decision processes. Artificial Intelligence in Medicine 18, 221–244 (2000)

15. Kirkizlar, E., Faissol, D., Griffin, P., Swann, J.: Timing of testing and treatment for asymptomatic diseases. Mathematical Biosciences, 28–37 (2010)

16. Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J.: Process evaluation in randomised controlled trials of complex interventions. Bmj 332, 413 (2006)

Ranking Information in Networks

Tina Eliassi-Rad[1] and Keith Henderson[2]

[1] Rutgers University
tina@eliassi.org
[2] Lawrence Livermore National Laboratory
keith@llnl.gov

Abstract. Given a network, we are interested in ranking sets of nodes that score highest on user-specified criteria. For instance in graphs from bibliographic data (e.g. PubMed), we would like to discover sets of authors with expertise in a wide range of disciplines. We present this ranking task as a *Top-K* problem; utilize fixed-memory heuristic search; and present performance of both the serial and distributed search algorithms on synthetic and real-world data sets.

Keywords: Ranking, social networks, heuristic search, distributed search.

1 Introduction

Given a graph, we would like to rank sets of nodes that score highest on user-specified criteria. Examples include (1) finding sets of patents which, if removed, would have the greatest effect on the patent citation network; (2) identifying small sets of IP addresses which taken together account for a significant portion of a typical day's traffic; and (3) detecting sets of countries whose vote agree with a given country (e.g., USA) on a wide range of UN resolutions. We present this ranking task as a *Top-K* problem. We assume the criteria is defined by L real-valued features on nodes. This gives us a matrix F, whose rows represent the N nodes and whose columns represent the L real-valued features. Examples of these node-centric features include degree, betweenness, PageRank, etc. We define an *L-tuple* $\langle v_1, \ldots, v_L \rangle$ as a selection of one value from each column. The score of an L-tuple is equal to the sum of the selected values (i.e., sum of the tuple's components). For a given parameter K, we require an algorithm that efficiently lists the top K L-tuples ordered from best (highest score) to worst. The L-tuple with the highest score corresponds to a set of nodes in which all features take on optimal values.

As described in the next section, we solve the Top-K problem by utilizing SMA* [4], which is a fixed-memory heuristic search algorithm. SMA* requires the specification of an additional parameter M for the maximum allotted memory-size. The choice for M has a dramatic effect on runtime. To solve this inefficiency problem, we introduce a parallelization of SMA* (on distributed-memory machines) that increases the effective memory size and produces super-linear speedups. This allows use to efficiently solve the Top-K ranking problem for large L and K (e.g., $L \in [10^2, 10^3]$ and $K \in [10^6, 10^9]$). Experiments on synthetic and real data illustrate the effectiveness of our solution.

2   A Serial Solution to the Top-K Ranking Problem

Given F, a matrix of L columns of real-valued features over N rows of vertices, and an integer K, our application of SMA* will report the top K scoring L-tuples that can be selected from F. First, we take each column of matrix F and represent it as its own vector. This produces a list X containing L vectors (of size $N \times 1$). We sort each vector in X in decreasing order. Then, we begin constructing D, another list of L vectors of real numbers. Each vector D_i has $|X_i| - 1$ entries, where $D_{ij} = X_{ij} - X_{i(j+1)}$. Note that the values in each D_i are all nonnegative real numbers; and D_i is unsorted.

The top-scoring L-tuple, R_1, can immediately be reported; it is simply the tuple generated by selecting the first element in each X_i. At this point, the problem can be divided into two subproblems whose structure is identical to the original problem (i.e. another instance of the Top-K problem). While there are several possible ways to make this division, we choose one that allows us to search efficiently. In particular, we choose the vector which would incur the least cost when its best element is discarded.[1] This can be determined by choosing the vector i with the minimum D_{i1}. Given this index i, we generate two subproblems as follows. In the first subproblem, we discard the best element in vector X_i. The resulting list of vectors will be called X^1. In the second subproblem, we keep the best element in vector X_i and remove all other elements from vector X_i. The resulting list of vectors here will be called X^0.

Let's illustrate this procedure with an example. Suppose $X = \{[10,8,5,2,1], [4,3,2,1], [30,25,24,23,22]\}$; so, $D = \{[2,3,3,1], [1,1,1], [5,1,1,1]\}$. (Recall that we assume lists in X are already sorted in decreasing order and some entries maybe missing.) The top-scoring L-tuple is $R_1 = \langle 10, 4, 30 \rangle$ with $score(R_1)$=10+4+30=44. Starting from X, the next best tuple can be generated by selecting the list i with the smallest D_{i1} and decrementing X_i in X: $X^1 = \{[10,8,5,2,1], [3,2,1], [30,25,24,23,22]\}, D^1 = \{[2,3,3,1], [1,1], [5,1,1,1]\}$, and $score(X^1)$=10+3+30=43. At this point, we can split the problem into two smaller problems. We can either "accept" the best decrement (X^1 above) or "reject" the best decrement and all future chances to decrement that list: $X^0 = \{[10,8,5,2,1], [4], [30,25,24,23,22]\}, D^0 = \{[2,3,3,1], [], [5,1,1,1]\}$, and $score(X^0)$=10+4+30=44. In this way, we can generate X^{11}, X^{10}, X^{01}, etc. This defines a binary tree structure in which each node has two children. Each series of superscripts defines a unique (L-tuple) node.

Our procedure for generating subproblems has three important properties. First, every L-tuple generated from X is either R_1, from X^1, or from X^0. Second, no L-tuple generated from X^1 or X^0 has a score greater than $score(R_1)$. Third, the top-scoring L-tuple from X^0 has the same score as R_1.

Given this formulation, a recursive solution to the Top-K problem is theoretically possible. In the base case, the input list X has L vectors of length one, and there is only one possible L-tuple to report (namely, R_1). Given arbitrary length vectors in X, we can divide the problem as above and merge the resultant top-K lists with R_1, discarding all but the top K elements of the merged list. This method, however, is impractical since each of the lists returned by the recursive calls could contain as many as K elements; hence, violating the requirement that space complexity must *not* be

---

[1] If all vectors in X contain exactly one element, no subproblems can be generated. In this case, the Top-K problem is trivial.

$O(K)$. But, a search-based approach allows us to generate the top K tuples one at a time. If we treat each Top-K instance as a node, and each subproblem generated from an instance as a child of that node, then we can treat the Top-K problem as search in a binary tree. The cost of traversing an edge is equal to the loss in score incurred by removing elements from X; thus the X^0 edge always has cost 0 and the X^1 edge has cost equal to $bestDiff(X) =_{def} min(D_{i1}|i = 1 \cdots L)$. In this context, A* search [3] clearly generates subproblems in order of their R_i scores. For the heuristic function $h(n)$, we use $bestDiff(X)$, which can be readily computed. Note that $bestDiff$ is monotone (and thus admissible) by the following argument.[2] If p is a 1-child of n (the X^1 subproblem), then $h(n) = cost(n, p)$ and $h(p) \geq 0$. Otherwise, $cost(n, p) = 0$ and $h(n) \leq h(p)$ by the definition of $bestDiff$.

Unfortunately, A* search requires storage of the $OPEN$ list in memory, and the size of $OPEN$ increases with every node expansion. This violates our memory requirements (of less than $O(K)$ storage), so we employ SMA* search [5] which stores a maximum of M nodes in memory at any given time during the execution. SMA* expands nodes in the same order as A* until it runs out of memory. At that point, the least promising node is deleted and its f-value is backed up in its parent.[3] SMA* is guaranteed to generate the same nodes as A* and in the same order. However, it may generate some intermediate nodes multiple times as it "forgets" and "remembers" portions of the tree. In certain cases, especially when the parameter M is small compared to the size of the search fringe, SMA* can experience "thrashing." This thrashing results from large parts of the search tree being generated and forgotten with very few new nodes being discovered. Our serial Top-K algorithm is essentially the same as SMA* [5] except for the use of heaps in selecting which of the X_i vectors to decrement at each node.

3   A Parallel Solution to the Top-K Ranking Problem

The aforementioned serial algorithm is very sensitive to the choice of M, the maximum amount of memory that can be allocated (see the Experiments Section). Here we present a parallel SMA* algorithm that offers dramatic improvement in runtime.

For a machine with P processing nodes, we use A* search to generate the P–1 best candidate subproblems. This covers the entire search tree. Because each subproblem is independent of the others, each processing node can perform SMA* search on its subproblem and incrementally report results to the master processing node, where the incoming tuple lists are merged and results are reported.

For small P, this algorithm works as expected and produces super-linear speedup. However as P increases, the performance boost can decrease quickly. This occurs because the runtime for this parallel algorithm is dominated by the single processing node with the longest runtime. If the initial allocation of subproblems to processing nodes is imbalanced, additional processing nodes may not improve performance at all. To ameliorate this problem, we adopt a load-balancing heuristic.

---

[2] Recall that a heuristic function $h(n)$ is monotone if $h(n) \leq cost(n, p) + h(p)$ for all nodes n and all successors p of n.

[3] The function f represents the total cost function, which equals the sum of the cost encountered so far and the estimated cost.

We run parallel SMA* on the input data with $K' \ll K$ as the threshold parameter. We then use the relative load from each subproblem as an estimate of the total work that will have to be done to solve that subproblem. We use these estimates to redistribute the initial nodes and repeat until there are no changes in the initial allocation of nodes. This distribution is then used to generate the top K L-tuples as described above. In our experiments, this heuristic correctly balanced the loads on the processing nodes. Moreover, the initial overhead to calculate the estimates was a negligible fraction of the overall runtime.

4   Experiments

4.1   Synthetic Data: Results and Discussion

To determine the runtime requirements for our Top-K algorithm, we generated a synthetic data set with $L = 100$ lists. Each list has between one and ten real values distributed uniformly in $[0, 1]$. Figure 1 shows runtime results for $M = 10^6$ nodes in memory. Note that as K increases, time complexity in K becomes near-linear. However, at approximately 20,000 tuples per second it is still too slow to be practical for large values of K.

Figures 2(a), 2(b), and 2(c) show the performance metrics for a strong scaling experiment with the parallel algorithm: *runtime*, *speedup*, *efficiency*. Runtime is the wall-clock runtime. Speedup is defined as the time taken to execute on one process (T_1) divided by the time taken to execute on P processes (T_p). Linear speedup is the ideal case where 2 processes take half the time, 3 processes take a third of the time, and so forth. Efficiency is defined as speedup divided by the number of processors. Ideal efficiency is always 1, and anything above 1 is super-linear. Values less than 1 indicate diminishing returns as more processors are added.

For the parallel Top-K experiments, we use the same 100 lists described above but with $M = 10^5$, $K = 2 \cdot 10^6$, and $K' = 10^4$. There is no I/O time in these experiments, as tuples are simply discarded once they are discovered. Note that the runtime for $P = 2$ is the same as the serial runtime plus overhead because in this case one processing node is the master and simply "merges" the results from the compute node. We see super-linear speedup for values of P up to 17 processes. However, the runtime for $P = 17$ is nearly identical to the runtime for $P = 9$. This is because in the $P = 17$ experiment, one of the large subproblems is not correctly split and redistributed. Figure 2(d) reveals the cause. When $P > 9$, the subproblems are not being correctly distributed among the nodes. Figure 3(d) shows the same results for the load-balanced version of the algorithm. In this case, it is clear that the nodes are being correctly distributed.

Figures 3(a), 3(b), and 3(c) show the parallel efficiency results for the load-balanced case. Our load-balancing heuristic drastically improves performance as the number of processors increases. These results include the time required to calculate the load-balance estimates, which explains the slightly longer runtimes at $P \le 9$.

4.2   Real Data: Results and Discussion

We tested our Top-K algorithm on four real-world data sets – namely, patent citations, DBLP bibliographic data, IP traffic, and UN voting. For brevity, we only discuss two
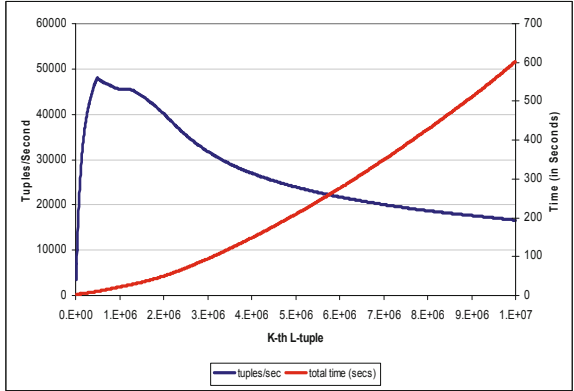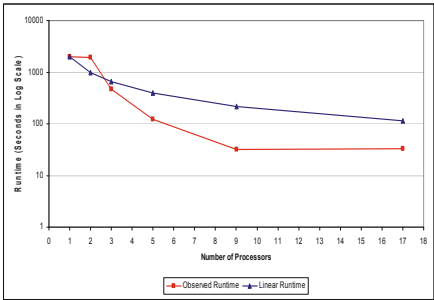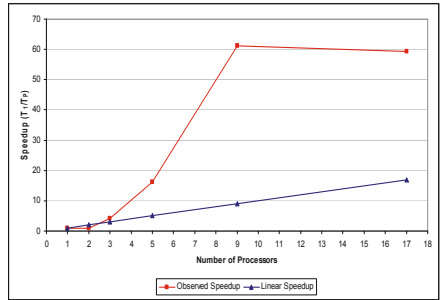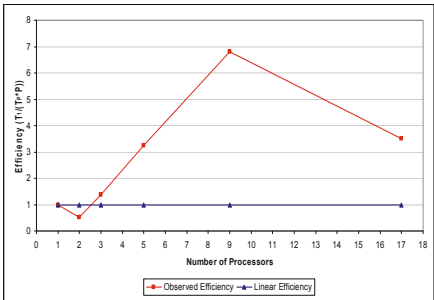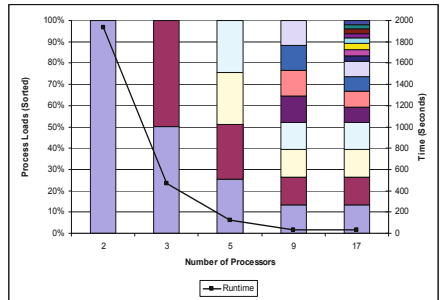
Fig. 1. Serial top-K: Runtime as a function of K

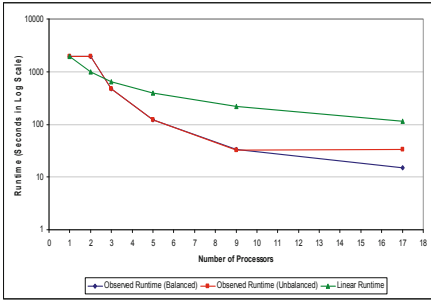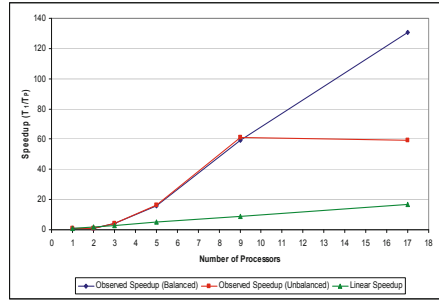(a) Runtime scaling

(b) Speedup

(c) Efficiency

(d) Load balance

Fig. 2. Unbalanced parallel top-K: $K = 2 \cdot 10^6$, $L = 100$ lists, $M = 10^5$ nodes

(a) Runtime scaling

(b) Speedup
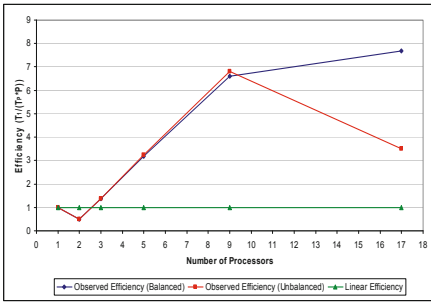
(c) Efficiency
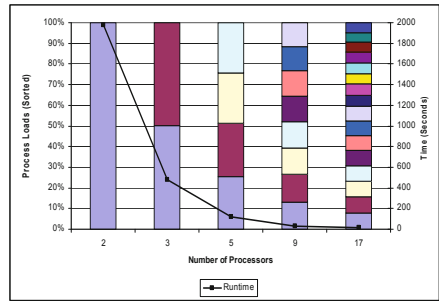
(d) Load balance

Fig. 3. Balanced parallel top-K: $K = 2 \cdot 10^6$, $L = 100$ lists, $M = 10^5$ nodes

of them here. First, our patents data set is a collection of 23,990 patents from the U.S. Patent Database (subcategory 33: biomedical drugs and medicine). The goal here is to discover sets of patents which, if removed, would have the greatest effect on the patent citation network. In particular, we are interested in patents which are frequently members of such sets; and we want to find which centrality measures correspond to these frequent nodes in the citation network. We generated a citation network with 23,990 nodes and 67,484 links; then calculated four centrality metrics on each node: degree, betweenness centrality, random walk with restart score (RWR), and PageRank. In the Top-K framework, there are 4 vectors: one for each centrality measure. The indices into the vectors are patents. The entry i in vector j is the (normalized) centrality score j for patent i. Table 1 lists the patents with the highest frequency in the top-100K tuples. As expected, the patents associated with DNA sequencing have the highest frequencies. Table 2 presents the centrality scores for the patents listed in Table 1. As it can be seen, the centrality scores alone could not have found these highly impacting patents (since no patent dominates all four centrality scores).

Our second data set is composed of 4943 proposed UN resolutions, votes from 85 countries per resolution (yes, no, abstain, etc), and 1574 categories. Each proposed resolutions is assigned to one or more categories (with an average of 1.6 categories per proposal). We select the most frequent 20 categories. Then, we score each country in each category by how often it agrees (or disagrees) with the USA vote in that category:

Table 1. Patents with the highest frequency in the top-100K tuples

Patent Number	Frequency (in Top-100K Tuples)	Patent Title
4683195	0.799022	Process for amplifying, detecting, and/or-cloning nucleic acid sequences
4683202	0.713693	Process for amplifying nucleic acid sequences
4168146	0.202558	Immunoassay with test strip having antibodies bound...
4066512	0.175828	Biologically active membrane material
5939252	0.133269	Detachable-element assay device
5874216	0.133239	Indirect label assay device for detecting small molecules ...
5939307	0.103829	Strains of Escherichia coli, methods of preparing the same and use ...
4134792	0.084579	Specific binding assay with an enzyme modulator as a labeling substance
4237224	0.074189	Process for producing biologically functional molecular chimeras
4358535	0.02674	Specific DNA probes in diagnostic microbiology

Table 2. Centrality scores for the 10 highest-frequency patents in the top-100K tuples

Patent #	Degree	Betweenness	RWR	PageRank
4683195	**1.000**	0.433	0.030	**1.000**
4683202	0.974	0.234	0.000	0.975
4168146	0.099	**1.000**	0.132	0.088
4066512	0.036	0.845	0.185	0.026
5939252	0.221	0.234	**1.000**	0.000
5874216	0.122	0.234	**1.000**	0.000
5939307	0.071	0.234	0.965	0.000
4134792	0.126	0.772	0.040	0.118
4237224	0.341	0.760	0.016	0.332
4358535	0.460	0.670	0.064	0.446

$\frac{1}{m}$ points per agreement, where m is the number of agreements per proposed resolution. The input to our top-K algorithm then is a matrix F with 85 rows (one per country) and 20 columns (one for each selected category). The entries in this matrix are the weighted agreement scores. We set K (in the top-K) to $100,000$. We repeat the experiment with scores for disagreements. Figures 4(a) and 4(b) depict the ranking results.

5   Related Work

Previous studies of the Top-K problem focus on cases where L is small (2 or 3) and each list is very long. Fredman ([2]) studies the problem of sorting $X + Y = \{x + y | x \in X, y \in Y\}$ in better than $O(n^2 \cdot log(n))$ time (where $|X| = |Y| = n$). Frederickson and Johnson ([1]) examine the slightly easier problem of discovering the Kth element of $X + Y$ in $O(n \cdot log(n))$ time. They also consider the problem of finding the Kth element of $\sum_{i=1}^{m} X_i$, proving that a polynomial algorithm in n for arbitrary $m \le n$ and K would imply that $P = NP$. Our approach uses fixed-memory heuristic search to enumerate L-tuples. It has $O(n \cdot log(n))$ runtime and at most $O(K)$ storage.
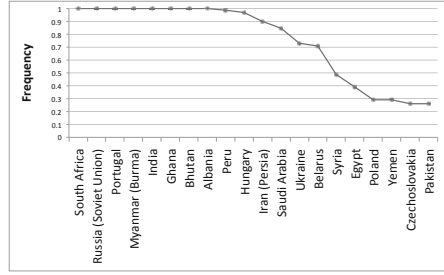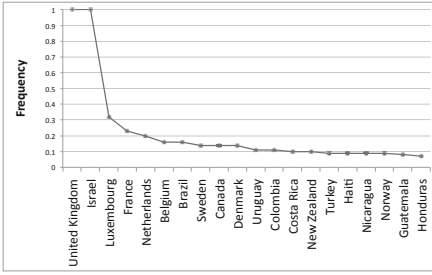
(a) Top countries agreeing with the USA in the top-100K tuples

(b) Top countries disagreeing with the USA in the top-100K tuples

Fig. 4. UN resolutions: Voting records on the 20 most frequent categories

6   Conclusions

We demonstrated that ranking sets of nodes in a network can be expressed in terms of the Top-K problem. Moreover, we showed that this problem can be solved efficiently using fixed memory by converting the problem into a binary search tree and applying SMA* search. We also parallelized SMA* for the Top-K problem in order to achieve super-linear speedup in a distributed-memory environment. Lastly we validated our approach through experiments on both synthetic and real data sets.

Acknowledgments

References

1. Frederickson, G.N., Johnson, D.B.: Generalized selection and ranking. In: Proc. of the 12th STOC, pp. 420–428 (1980)
2. Fredman, M.L.: How good is the information theory bound in sorting? Theoretic. Comput. Sci. 1, 355–361 (1976)
3. Hart, P.E., Nilsson, N.J., Raphael, B.: Correction to a formal basis for the heuristic determination of minimum cost paths. SIGART Bull. 37, 28–29 (1972)
4. Russell, S.: Efficient memory-bounded search methods. In: ECAI, pp. 1–5 (1992)
5. Russell, S.J.: Efficient memory-bounded search methods. In: Proc of the 10th ECAI, pp. 1–5 (1992)

Information Provenance in Social Media

Geoffrey Barbier and Huan Liu

Arizona State University,
Data Mining and Machine Learning Laboratory
gbarbier@asu.edu, huan.liu@asu.edu
http://dmml.asu.edu/

Abstract. Information appearing in social media provides a challenge for determining the provenance of the information. However, the same characteristics that make the social media environment challenging provide unique and untapped opportunities for solving the information provenance problem for social media. Current approaches for tracking provenance information do not scale for social media and consequently there is a gap in provenance methodologies and technologies providing exciting research opportunities for computer scientists and sociologists. This paper introduces a theoretical approach aimed guiding future efforts to realize a provenance capability for social media that is not available today. The guiding vision is *the use of social media information itself to realize a useful amount provenance data for information in social media.*

Keywords: provenance, information provenance, data provenance, social media, data mining, provenance path.

1 Introduction and Motivation

Collective behavior can be influenced by small pieces of information published in a social media setting such as a social networking site, a blog, micro-blog, or even a wiki [2,6]. Considering information as defined by Princeton's WordNet[1] as "a collection of facts from which conclusions may be drawn" and provenance as "the history of ownership of a valued object[2]", *information provenance in social media* can be considered as the origins, custody, and ownership of a piece of information published in a social media setting. Often the provenance of information published via social media is only known after a group has been influenced and motivated in a particular manner.

A lack of accurate, reliable, history or metadata about a social media information source can present problems. In March 2010, John Roberts, a United States Supreme Court Justice, was reportedly set to retire because of health issues. As it turned out, Justice Roberts had no plans to retire and a rumor that grew from a professor's teaching point, meant only for a classroom, made national headlines [2]. When Twitter was used by protestors in Iran during 2009, the source of

---

[1] http://wordnetweb.princeton.edu/perl/webwn
[2] http://www.merriam-webster.com/dictionary/provenance

some messages could not be verified and were deemed to be of no value or even antagonistic [6]. These problems might have been avoided with provenance information related to the subject, the source, or perhaps even the ideologies in play.

Some mechanisms have been designed to record provenance information for databases, the semantic web, workflows, and distributed processing [10]. However, provenance information is not routinely tracked today for social media. Although some thought has been given about the need [7] and some potential approaches [5,11], a practical approach and responsive mechanism has not be identified or implemented for today's online social media environment. In some instances, partial provenance information may suffice to inform groups in such a manner resulting in sound behaviors. Additionally, an approach for information provenance in social media needs to address the rapidly changing social media environment and must quickly respond to queries about the provenance of a piece of information published in social media.

The social media environment provides unique challenges for tracking and determining provenance. First, the social environment is *dynamic*. With more than half a billion[3] Facebook[4] users, new social media content is generated every day. Facebook is only one social media outlet. Consider there are over 40 million microblog messages posted everyday to the popular site Twitter[5]. Second, social media is *decentralized* in the sense that information can be published by almost anyone choosing one or more social media platforms and then relayed across disparate platforms to a multitude of recipients. Third, the environment provides *multiple modes* of communication such as profile updates, blog posts, microblogs, instant messages, and videos. Given this extremely challenging environment, new approaches for information provenance are needed to track where a piece of information came from in order to determine whether or not the information can be used as a basis for a decision.

Determining the provenance of a piece of information is challenging because provenance data is not explicitly maintained by most social media applications. A piece of information often moves from one social media platform to another. As a result, provenance data for a piece of information in the social media environment needs to be derived. It is proposed here that social media data itself can be used to derive the provenance or most likely provenance data about a piece of information. This departs from traditional approaches of information and data provenance which rely on a central store of provenance information.

The remainder of this paper highlights related work noting at this point in time contemporary approaches to provenance rely on a central store of provenance meta-data or an established, widely adopted, semantic framework. Following the highlights of related work, the concept of a provenance path and associated problem space applicable to today's social media environment are outlined. An example to illustrate the provenance path concept is included and future research opportunities are presented.

---

[3] http://www.facebook.com/press/info.php?statistics
[4] www.facebook.com
[5] http://blog.twitter.com/2010/02/measuring-tweets.html

2   Related Work

Provenance data is valuable in a variety of circumstances including database validation and tracing electronic workflows such as science simulations and information products produced by distributed services. Moreau published a comprehensive survey [10] that captures provenance literature related to the web. Although the survey provides over 450 references with an emphasis on data provenance, the survey does not identify a significant body of literature relating to provenance *and* social media, social computing, or online social networks.

The database and workflow communities depend on a provenance store implemented in one form or another. The proposed store is maintained centrally by one application or in a distributed scheme, each application involved in producing a piece of data maintains its own provenance store. An open provenance model (OPM) was developed to facilitate a "data exchange format" for provenance information [10]. The OPM defines a provenance graph as a directed graph representing "past computations" and the "Open Provenance Vision" requires individual systems to collect their provenance data [10].

Hasan et al. [8] define a "provenance chain." Deolalikar and Laffitte [3] investigate data mining as a basis for determining provenance given a set of documents. The World Wide Web Consortium has a provenance incubator group[6] to develop requirements for the semantic web. Golbeck [5] connects provenance with trust leveraging Resource Description Framework (RDF) supported social networks. Fox and Huang define what they call Knowledge Provenance (KP) [4,9]. Their KP construct accounts for varying levels of certainty about information found in an enterprise (i.e., the web).

In today's social media environment, applications do not collect or maintain provenance data directly. Thus, it would be practical to determine provenance dynamically without depending on a central store. Information provenance in social media must be analyzed, inferred, or mined to be realized.

3   Provenance Paths

Provenance can be characterized as a directed graph [3,5,10,12]. Within the graph, a *provenance path* can be assembled for each statement produced from the social media environment. We propose next, formal definitions for provenance paths as applied to social media.

The social media environment network will be represented by a directed graph $G = (V, E)$, $v \in V$ and $e \in E$. V is the set of nodes representing social media users publishing information using social media applications. E is the set of edges in G representing explicit transmission of social media communication between two nodes in V. An explicit transmission occurs when distinct information is communicated from one node to another or when one node directly accesses information available at another node. Publishing information alone is not

---

[6] http://www.w3.org/2005/Incubator/prov/wiki/Main_Page

considered an explicit transmission and does not create an edge in E. Given the directed graph $G = (V, E)$. We have the following definitions.

Definition: T is the set of *recipient* nodes in $G : T \subseteq V$.

Definition: A is the set of *accepted*[7] nodes in $G : A \subseteq V$ and $(T \subset A)$.

Definition: D is the set of *discarded*[8] nodes in $G : D \subset V, (D \cap A) = \emptyset$, and $(D \cap T) = \emptyset$.

Definition: $(A \cup D)$ are *identified* nodes.

Definition: M is the set of *undecided* nodes in $G : M = V - (A \cup D)$.

Definition: a *provenance path*, p, is a path in $G : p = (v_1, v_2, \ldots, v_n) : v_1 \neq v_n, v_1 \in V$, and $v_n \in T$.

Definition: P is the set of *all provenance paths* in $G : \forall p_i \in P, i = 1 \ldots m : m = \mid P \mid$ and $p_1 \neq p_2 \neq p_3 \neq \ldots p_m$.

Definition: *Accepted provenance path*, p : for all nodes, v_k, in path p, $v_k \in A$.

Definition: *Heterogeneous provenance path*, p: for all nodes, v_j, in path p, $v_j \in A, v_j \in D$, or $v_j \in M$.
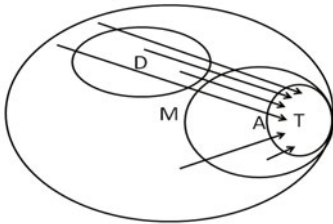
Fig. 1. Sets and abstract paths

A provenance path is a set of nodes and edges comprising a path which an element of social media information is communicated from a node in graph to a recipient or recipients. Nodes in the set T (an individual or group) are the final recipients of information along a provenance path, hereafter called recipient. The recipient makes one or more decisions based on the information received via a provenance path. Each provenance path is unique. There may be more than one provenance path for information provided to a recipient. Figure 1 illustrates the most common relationship between the subsets of V. The arrows illustrate some characteristics of possible provenance paths including accepted and heterogeneous provenance paths.

4   Provenance Path Problem Space

The social media environment provides a very challenging problem for determining information provenance. This environment, like the web, provides a theoretically bounded but practically unbounded problem space because of the large

---

[7] The criterion for accepting nodes is uniquely determined by T.

[8] The criterion for discarding nodes is uniquely determined by T.

number of users in the social media environment. Although there are a finite number of websites as part of the world-wide-web, determining the actual number of web sites is extremely challenging [1]. Similarly, there are a finite number of social media users and social media information. However, the precise actual number of users is practically intractable. This unbounded social media environment presents an unbounded problem space for provenance paths in practice although in some cases the environment will be bounded such as when considering a single social networking site or small subset of social media sites.

A provenance path can begin at an identified node ($v_1 \in A$ or $v_1 \in D$) or from a node that is undecided ($v_1 \in M$). The social media environment presents cases where a provenance path exists but all of the nodes and edges in the path are not known or only partially known to the recipient, defined as an *incomplete provenance path*. In the case of an incomplete provenance path, the complete provenance path exists in the social media environment but the complete path is not discernable to the recipient. Given an incomplete provenance path, the primary goal of solving the provenance path problem is to make all of the unknown nodes and edges known to the recipient. When all of the nodes and edges are known by the recipient, the provenance path is defined as a *complete provenance path*. When the social media environment is unbounded, it may not be possible to make a complete provenance path known to the recipient. The recipient will need to employ strategies to decide whether or not the revealed portions of the incomplete provenance path provides useful information. The problem space can be considered and approached from different perspectives depending on whether or not the path is complete or incomplete.

4.1   Complete Provenance Paths

Assessing provenance will be easiest when the recipient can access a complete provenance path with all of the node and edge relationships known to the recipient. Identified nodes are determined by a recipient as accepted or discarded. The criterion for accepting nodes can be based on one characteristic or a combination of characteristics attributed to nodes in the social media environment. However, it is intuitive that the most common problem encountered with information provenance in social media is dealing with incomplete provenance paths.

4.2   Incomplete Provenance Paths

If the actual path is not completely known to the recipient, it could be difficult to determine whether or not a discarded node contributed to or altered information presented to the recipient. Social media data could be leveraged to estimate likely paths. The nodes and links that are known to the recipient or consequently discovered can be exploited to provide a warning or calculate confidence values using probabilistic mechanisms to determine how the information might be considered.

Approaches need to be designed to infer provenance data when the path is incomplete. Decision strategies need to be developed to help the recipient authenticate information provided through social media or determine whether or

not the information itself can be corroborated via a separate provenance path including accepted social media nodes. In some cases it may enough to know whether or not the idea being presented is adversarial, complementary, or unique toward the recipient individual or group. Understanding the nuances of a publication, position, or opinion, could lend itself to a level of confidence acceptable to a recipient by using only the portion of the provenance path that is available for analysis.

4.3   Multiple Provenance Paths

Multiple provenance paths present both prospects and challenges. When multiple provenance paths are *complementary*, the paths present consistent information to the recipient. Complementary provenance paths can serve as an authentication mechanism for the information presented to the recipient. The most challenging decisions an individual or group may need to make is when the provenance paths are incomplete and multiple provenance paths provide conflicting information.

When multiple provenance paths are conflicting by presenting inconsistent or contradictory information to the recipient, the provenance paths must be reconciled. In cases where provenance paths provide conflicting information, a probabilistic approach can be applied to determine which provenance path should be accepted, if any.

5   Provenance Path Case Study

Consider the case of the Justice Roberts rumor based on a simple investigation [2]. Reference Figure 2a, a Georgetown Law School professor (node v_1) shared fictitious information in his class along edges e_1, e_2, and e_3. A student in the class, node v_3, sends a message to a blog site node v_5 along edge e_4 and the group at the blog site publishes a story based on false information. Similar provenance paths reach other blog sites and false information about a Justice in the United States Supreme Court becomes a well-circulated rumor. The information communicated along e_4 may or may not be accurate. Given the provenance path shown in Figure 2a, node v_5 should determine whether or not it should accept the information about Justice Roberts. If the recipient node v_5 analyzes the provenance path and determines that it considers each node along the provenance path as accepted, v_5 could accept the information received via the explicit communication along e_2 and e_4. However, if v_1 or v_3 are discarded nodes, the recipient will need to consider what must be done in order to authenticate the information.

In Figure 2b, an additional node, JR, is added to represent Justice Roberts. If node JR was included the provenance path, the information might be considered reliable. However, given that the node JR is not included in the path (as far as the recipient can initially discern) questions should be raised about the validity of the information. In this case, direct or indirect connections using networking
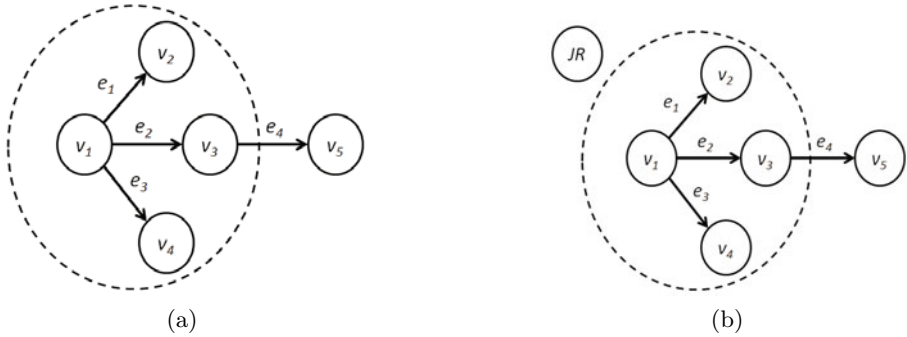
Fig. 2. A case study

information and available social media could be examined to glean additional information. As examples, comparing the "distance" from node v_1 and node JR to a common reference point in social media or analyzing the individuals' (v_1, v_3, and JR) group memberships and associated group traits.

6   Future Work

In this paper we defined the problem of information provenance in contemporary social media and presented general ideas for approaching it along with a case study. Our research is ongoing and our future work will be focussed on detailing specific methods and metrics aimed at developing solutions to this problem.

In addition to the motivating cases previously discussed, this research has exciting implications for addressing contemporary issues facing users and decision makers such as: identifying the source of an online product review to reveal fake reviews, helping to implement a practical cyber genetics[9], and determining the source when no author is evident.

Beyond the work to identify and assess provenance paths there are additional questions related to information provenance in social media such as:

- How should provenance attributes and characteristics be defined?
- How would provenance paths be valued from different recipients?
- Can provenance paths be identified and leveraged to help influence a group?
- In addition to trust, what other connections can be made between provenance and elements of social media?
- What are the implications for privacy?

---

[9] http://www.darpa.mil/sto/programs/cg/index.html

References

1. Benoit, D., Slauenwhite, D., Trudel, A.: A web census is possible. In: International Symposium on Applications and the Internet (January 2006)
2. Block, M.: Tracing rumor of John Roberts' retirement (March 2010), http://www.npr.org/templates/story/story.php?storyId=124371570
3. Deolalikar, V., Laffitte, H.: Provenance as data mining: combining file system metadata with content analysis. In: TAPP 2009: First Workshop on Theory and Practice of Provenance, pp. 1–10. USENIX Association, Berkeley (2009)
4. Fox, M.S., Huang, J.: Knowledge provenance in enterprise information. International Journal of Production Research 43(20), 4471–4492 (2005)
5. Golbeck, J.: Combining provenance with trust in social networks for semantic web content filtering. In: Moreau, L., Foster, I. (eds.) IPAW 2006. LNCS, vol. 4145, pp. 101–108. Springer, Heidelberg (2006)
6. Grossman, L.: Iran protests: Twitter, the medium of the movement. Time June 17 (2009), http://www.time.com/time/world/article/0,8599,1905125,00.html
7. Harth, A., Polleres, A., Decker, S.: Towards a social provenance model for the web. In: 2007 Workshop on Principles of Provenance (PrOPr), Edinburgh, Scotland (November 2007)
8. Hasan, R., Sion, R., Winslett, M.: Preventing history forgery with secure provenance. Trans. Storage 5(4), 1–43 (2009)
9. Huang, J., Fox, M.S.: Uncertainty in knowledge provenance. In: Bussler, C.J., Davies, J., Fensel, D., Studer, R. (eds.) ESWS 2004. LNCS, vol. 3053, pp. 372–387. Springer, Heidelberg (2004)
10. Moreau, L.: The foundations for provenance on the web. Submitted to Foundations and Trends in Web Science (2009)
11. Simmhan, Y., Gomadam, K.: Social web-scale provenance in the cloud. In: McGuinness, D.L., Michaelis, J.R., Moreau, L. (eds.) IPAW 2010. LNCS, vol. 6378, pp. 298–300. Springer, Heidelberg (2010)
12. Simmhan, Y.L., Plale, B., Gannon, D.: A survey of data provenance techniques. Tech. Rep. IUB-CS-TR618, Computer Science Department, Indiana University, Bloomington, IN 47405 (2005)

A Cultural Belief Network Simulator

Winston R. Sieck, Benjamin G. Simpkins, and Louise J. Rasmussen

Applied Research Associates, Inc.
1750 Commerce Center Blvd, N., Fairborn, OH, 45324
{wsieck,bsimpkins,lrasmussen}@ara.com

Abstract. This paper describes a tool for anticipating cultural behavior called, "CulBN." The tool uses Bayesian belief networks to represent distributions of causal knowledge spread across members of cultural groups. CulBN simulations allow one to anticipate how changes in beliefs affect decisions and actions at the cultural level of analysis. An example case from a study of Afghan decision making is used to illustrate the functionality of the tool.

Keywords: cultural network analysis, CNA, cultural model, belief network, Bayesian modeling.

1 Introduction

The purpose of this paper is to describe a prototype tool for simulating cultural behavior; the "Cultural Belief Network Simulation Tool" (CulBN). We use an example case from a study of Afghan decision making to illustrate key features of the tool and associated process for building cultural models. The tool's functionality derives from a principled, cognitive approach to culture. We first define and discuss this view of culture, prior to discussing specific details of CulBN.

The cognitive approach to culture begins by defining culture in terms of the widely shared concepts, causal beliefs, and values that comprise a shared symbolic meaning system [1, 2]. Shared meaning systems imply that members of a culture tend to perceive, interpret, and respond to events in similar ways. The cognitive approach to culture is often combined with an epidemiological metaphor from which researchers seek to explain the prevalence and spread of ideas in populations [3].

Working from this perspective, we previously developed cultural network analysis (CNA), a method for describing ideas that are shared by members of cultural groups, and relevant to decisions within a defined situation [4]. CNA discriminates between three kinds of ideas: concepts, values, and beliefs about causal relations. The cultural models resulting from CNA use belief network diagrams to show how the set of relevant ideas relate to one another. The CNA approach also includes the full set of techniques needed to build cultural model diagrams. This consists of specific methods to elicit the three kinds of ideas from native people in interviews or survey instruments, extract the ideas from interview transcripts or other texts, analyze how common the ideas are between and within cultural groups, align and assemble the common ideas into complete maps. CNA offers a unique set of theoretical constraints which distinguishes the cultural models it produces from other ways of modeling culture [5]. These aspects include an emphasis on ensuring the relevance of cultural

models to *key decisions* to provide a more direct link to actual behavior, focus on the cultural insider or *emic perspective*, modeling interrelated *networks of ideas* rather than treating ideas as independent entities, and by seeking to directly estimate the actual *prevalence of ideas* in the network rather than relying on more nebulous notions of sharedness.

This overall conception and approach to cultural analysis ideally suits simulations using Bayesian normative modeling, though with a slight twist. In typical applications of Bayesian belief networks (BNs), one elicits expert structural and probability inputs to get best possible representation of a physical system [6]. The aim in such cases is to accurately predict key physical outcomes based on expert understanding of influences within that physical system. That is, the ultimate concern is for physical reality. In contrast, the objective here is to represent human cultural knowledge as the system of interest. That is, the aim is to accurately anticipate the perceptions and decisions of groups within a specified context based on relevant culturally shared conceptions, causal reasoning, and appraised values. The key principles for application of BNs to cultural modeling are:

- BN focuses on population, rather than individual psychological level of analysis
- Nodes represent concepts held in common by some percentage of the population
- Edges represent causal beliefs, also distributed across members of the population
- Probabilities denote prevalence of ideas in the population, not strength of belief

In principle, concept nodes also contain associated value information. However, values are left implicit in the current instantiation. Next, we describe the CNA approach to building and simulating cultural models as a two step process. First, a data step is necessary to translate from raw qualitative data to structured inputs to be used in the second, modeling step. The modeling step incorporates the inputs to build executable models in CulBN. We walk through the two steps in the context of a concrete example of Afghan decision making.

2   Data Step

CulBN simulations, as a component of CNA, are grounded in the kinds of data people from a wide variety of cultures, including non-western and non-literate cultures, have been shown capable of providing in natural communications, interviews and field experiments. The first step in a CNA collection is to determine the target population and critical decision or behavior of interest. These elements are often captured in brief scenarios that provide some additional context, such as the following:

> *A unit of American soldiers in Afghanistan's Uruzgan Valley was engulfed in a ferocious fire fight with the Taliban. Only after six hours, and supporting airstrikes, could they extricate themselves from the valley. The Americans were surprised in the battle by the fact that many local farmers spontaneously joined in, rushing home to get their weapons. The farmers sided with the Taliban, fighting against the Americans. Asked later why they'd done so, the villagers insisted that they didn't support the Taliban's ideological agenda, nor were they particularly hostile toward the Americans.*

2.1   Interview Process

CNA interviews are structured around short scenarios that describe the critical decision of interest for a cultural group, such as the Afghan farmers' decision to join the fighting in the scenario above.   The information basis for the modeling effort described in this paper was a set of CNA interviews conducted with Afghans as part of a larger cultural study [7]. Each participant was interviewed individually using the same CNA interview guide. The interview guide consisted of open-ended questions to elicit participants' overall explanations of the situation, as well as their perspectives regarding the Afghan protagonists' intentions, objectives, fundamental values, and the causal links between them.

2.2   Coding and Analysis

Two independent coders read through all of the transcripts and identified segments that contained concept-causal belief-value (CBV) chains. Next, the two analysts coded each segment by identifying the antecedent, the consequent, and the direction of the relationship between the antecedent and consequent for each causal belief. For example: *"Americans bring security"* (concept) *"decreases the likelihood that"* (causal belief) *"Farmers join the Taliban side of the fight"* (concept). The coders then collaborated to resolve overlaps and redundancies between their respective lists. Percent agreement for the coding was 81% for the Farmers scenario. Prevalence information was then derived by tallying the number of times a CBV chain was mentioned across all three interviews. We converted the frequencies to proportions by using the count of the most prevalent idea as the upper bound. Finally, we created a frequency table for all ideas. The CBV chains were then integrated into a network diagram of the Afghan Farmers' decisions leading to their behavior in the scenario.

3   Modeling Step

In the modeling step, we construct the cultural belief network in CulBN, and use it as a starting point from which to perform simulations of concept and causal belief change on the decision of interest. The model is initialized with the empirical data that was collected above in the data step. The full cultural belief network for the Afghan Farmers is shown in Figure 1. Recall that these models represent how the cultural group perceives reality. They do not attempt to model physical reality. Hence, we can change various parameters of the model to simulate effects of changes in the distribution of ideas on cultural decisions and behavior. The remainder of this section describes the elements and process of the modeling step in more detail.

3.1   Diagram Semantics in CulBN

A subset of the full model is illustrated in Figure 2 to enhance readability. In these diagrams, nodes represent culturally-shared concepts, and the probability on the node
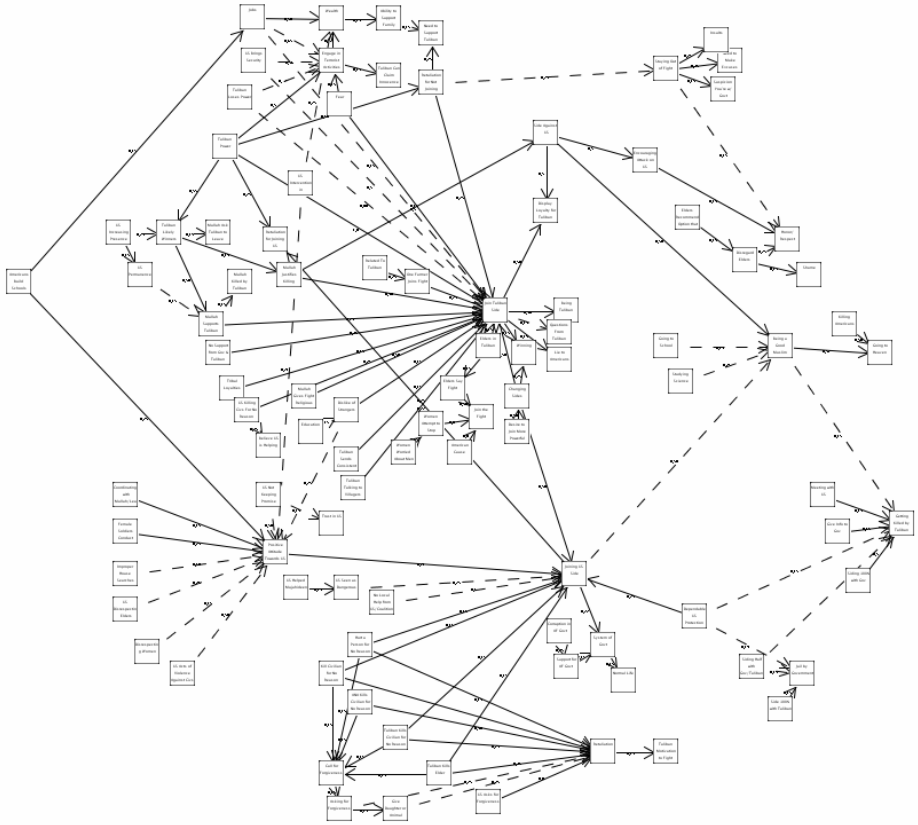
Fig. 1. Cultural belief network of Afghan Farmers' decision making in a conflict scenario as implemented in CulBN

signifies the prevalence of the concept in Afghan thinking in the scenario context. The solid arrows represent perceived positive causal relationships and dotted arrows represent negative ones. Probabilities on the lines denote prevalence of the causal beliefs. As shown in Figure 2, for example, the prevalence of the causal belief that jobs increase wealth for Afghans is 38%.

The values the model takes as input are initially determined empirically. These values represent the current state of the cultural model, as derived from the text data. To perform a simulation, the user revises the inputs with hypothetical values. The distinction is visualized in Figure 2. Here, the prevalence of the concept 'US Brings Security' is low, based on the data. However, we may wish to explore the consequences of a case where the concept in much more prevalent in the cultural group (as a result of radio announcements about security, for example). In Figure 2, we see the initial empirical value is represented by the dotted level on the node and the hypothesized values are represented with the solidfill.

After specifying input values, the user runs the simulator to acquire the probability values of the dependent nodes[1]. An output probability can be interpreted as the prevalence of the concept in the cultural group given the input values specified. For example, in Figure 2 we see that 'Join Taliban Side' is less prevalent with the increased perceptions that the 'US brings security.'
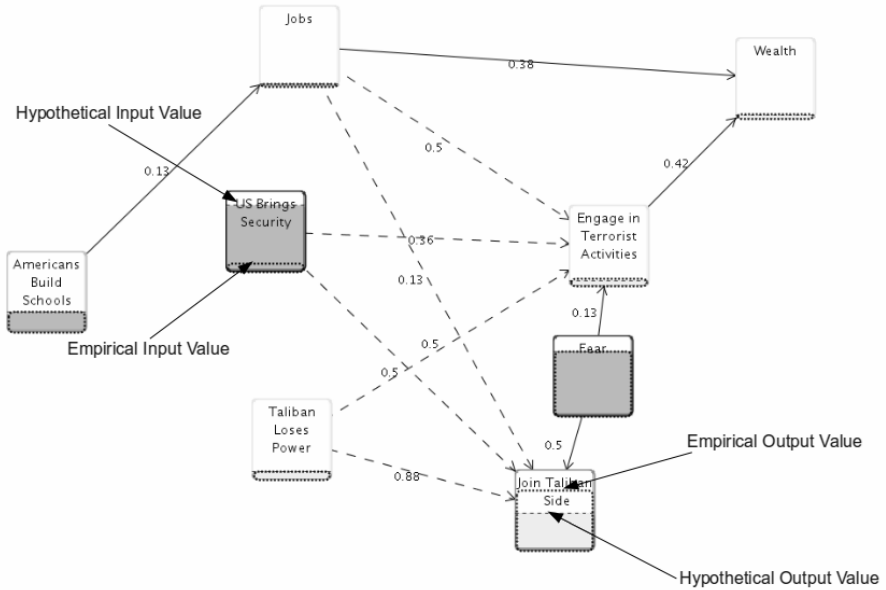
Fig. 2. Hypothetical and empirical values in CulBN

3.2   Interaction with Cultural Models in CulBN

CulBN contains a user interface that allows for creation, visualization, and manipulation of input values in a CNA-type cultural model. Combining this capability with the simulation engine allows the user to easily interact with a cultural model by supplying hypothetical data and visualizing the changes produced in the model. Figure 3 demonstrates this in the Afghan Farmer example. In the top left, we see a model initialized with only empirical data being run (the small dialog is the sample count). To the right, we see the output of the model that 75% of Afghan farmers are expected to join the Taliban in the scenario. In the bottom panels, we explore the effects of hypothesized changes to cultural concepts and causal beliefs related to perceptions of the extent that Americans are building schools locally. First, we hypothesize an increase in the prevalence of perceptions of school-building in the

---

[1] CulBN currently relies on the Java Causal Analysis Tool (JCAT) as its simulation engine 8.Lemmer, J. *Uncertainty management with JCAT*. 2007 [cited 2010; Available from: www.au.af.mil/bia/events%5Cconf-mar07%5Clemmer_jcat.pdf.. JCAT back-end uses a Monte Carlo Markov Chain (MCMC) sampler to stochastically generate the outcome likelihoods.

population (see the bottom left of Figure 3). Note that this does not result in much change in the percentage of farmers who decide to join the Taliban (notice that the solid fill on the 'Join Taliban Side' node is almost the same as the dotted fill). The implication here is that, based on the current model, influencing perceptions that schools are being built is not an effective strategy, in and of itself. However, if we incorporate an additional presumed influence on the *causal belief* that American school building leads to more jobs, then we will notice an impact on the decision to join the Taliban (see the bottom right of Figure 3). This example shows that subtle distinctions in the specific cultural ideas targeted for influence can yield sizable differences in the predicted effects on the decisions underlying cultural behavior.
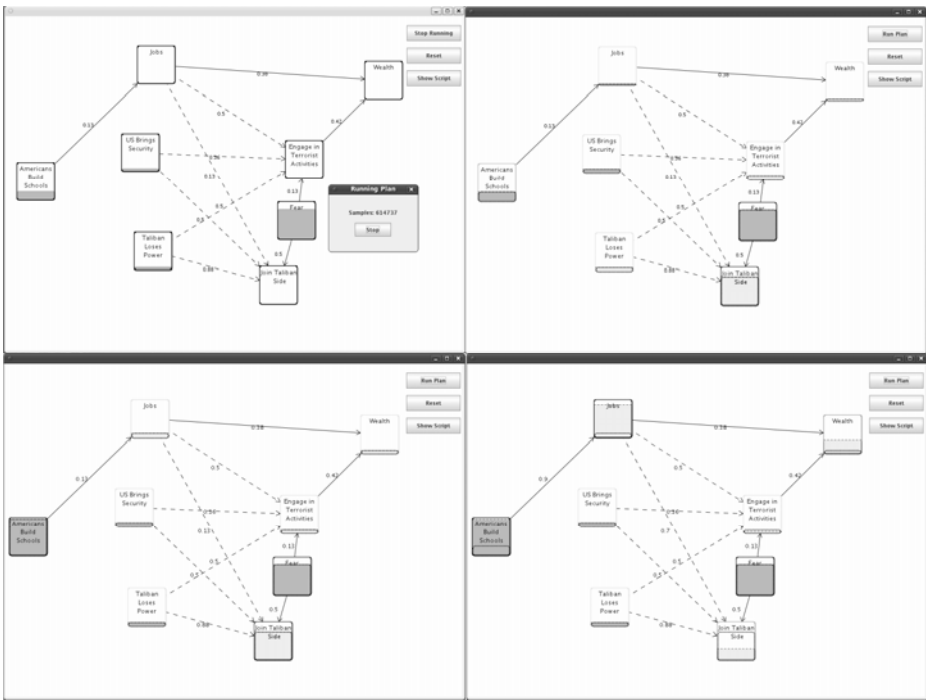
Fig. 3. Inputting hypothetical data and running the simulator in CulBN. Top left- running the empirical model, top right- result of empirical model, bottom left-result of a run with hypothetical input to a concept, bottom right- result of second run with hypothetical input added to links.

4   Discussion

4.1   Practical Implications

CulBN and its underlying principles are likely to prove useful in a wide array of applications. Here, we describe benefits of the tool for supporting analysts who need to understand the behavior of groups from other cultures. First, CulBN p*rovides a*

coherent approach for constraining and making sense of a vast array of concepts, causal beliefs, and values, as they are relevant within particular situations. CulBN, relying on the CNA process, begins with a group's key decisions, attitudes, or behaviors of interest in a given context. It then provides a systematic approach for identifying and linking the most relevant concepts, causal beliefs, and values from the group's own perspective to the key decisions. Secondly, CulBN y*ields a nuanced perspective of the thinking in diverse cultural groups. This is especially helpful to analysts who have little immersive experience within the culture.* By providing an explicit map of the group's perceived associations between concepts, the analyst has a ready reference that describes "what else members of the group would be thinking" in a given situation. Thus, it provides direct support for taking the group's perspective in that situation. Thirdly, CulBN i*ncreases the systematicity and credibility of analyst recommendations, while maintaining flexible analysis processes.* CulBN introduces a rigorous process for structuring available information regarding the cognitive content shared among members of a group, allowing analysts to freely explore the cognitive effects of hypothetical influences on the cultural group of interest. Fourthly, CulBN a*ids in promoting analysts' cultural expertise across teams and organizations.* The cultural model representations are useful knowledge products that can be shared among analysts to provide a concrete basis for discussions, including helping novice analysts get up to speed on specific topics relevant to their area of study. In addition, CulBN could be used to monitor changes in a group's, organization's, or wider society's attitudes and behaviors over time. That is, cultural models provide a basis for representing long-term cultural shifts in attitude and belief.

4.2   Theoretical Implications

The current work extends past developments in cultural network analysis by deriving quantitative cultural models from qualitative data, and incorporating them into a BN framework that enables simulated influence on cultural perceptions, decisions, and behavior. There are several advantages of the current approach to achieve cultural simulations. First, cultural agents that do not have the empirical grounding that comes by incorporating quantitative cultural models cannot be guaranteed to behave in culturally authentic ways. The performance of any model is bound by the data used to inform it. Second, a BN fits closely with the representational commitments made within CNA, allowing for the specific interpretations of elements and parameters that fit neatly within the epidemiological view of culture. Also, the quantitative cultural models incorporated into CulBN address the cultural "levels of analysis problem" that exists with many agent-models that aim to represent culture [9]. The problem is that the "cultural" level of analysis describes population characteristics, whereas cognitive agents represent individual people. By building and simulating quantitative cultural network models that estimate population-level characteristics of cultural groups, we can avoid building agents that stereotype the culture of interest. That is, CNA, combined with tools like CulBN, provides a coherent, grounded, and feasible means to model and simulate effects of changes in the distribution of mental representations in a population, as an alternative to stereotyped cultural agents that represent a canonical form of the culture.

Acknowledgments

References

1. Rohner, R.P.: Toward a conception of culture for cross-cultural psychology. Journal of Cross-Cultural Psychology 15(2), 111–138 (1984)
2. D'Andrade, R.G.: The cultural part of cognition. Cognitive Science 5, 179–195 (1981)
3. Sperber, D.: Explaining culture: A naturalistic approach. Blackwell, Malden (1996)
4. Sieck, W.R., Rasmussen, L.J., Smart, P.: Cultural network analysis: A cognitive approach to cultural modeling. In: Verma, D. (ed.) Network Science for Military Coalition Operations: Information Extraction and Interaction, pp. 237–255. IGI Global, Hershey (2010)
5. Sieck, W.R.: Cultural network analysis: Method and application. In: Schmorrow, D., Nicholson, D. (eds.) Advances in Cross-Cultural Decision Making, pp. 260–269. CRC Press / Taylor & Francis, Ltd., Boca Raton (2010)
6. Edwards, W.: Hailfinder: Tools for and experiences with Bayesian normative modeling. American Psychologist 53(4), 416–428 (1998)
7. Sieck, W.R., et al.: Honor and integrity in Afghan decision making. Manuscript in preparation (2010)
8. Lemmer, J.: Uncertainty management with JCAT (2007),
   `http://www.au.af.mil/bia/`
   `events%5Cconf-mar07%5Clemmer_jcat.pdf` (cited 2010)
9. Matsumoto, D.: Culture, context, and behavior. Journal of Personality 75(6), 1285–1320 (2007)

Following the Social Media: Aspect Evolution
of Online Discussion

Xuning Tang and Christopher C. Yang

College of Information Science and Technology
Drexel University
{xt24,chris.yang}@drexel.edu

Abstract. Due to the advance of Internet and Web 2.0 technologies, it is easy to extract thousands of threads about a topic of interest from an online forum but it is nontrivial to capture the blueprint of different aspects (i.e., subtopic, or facet) associated with the topic. To better understand and analyze a forum discussion given topic, it is important to uncover the evolution relationships (temporal dependencies) between different topic aspects (i.e. how the discussion topic is evolving). Traditional Topic Detection and Tracking (TDT) techniques usually organize topics as a flat structure but it does not present the evolution relationships between topic aspects. In addition, the properties of short and sparse messages make the content-based TDT techniques difficult to perform well in identifying evolution relationships. The contributions in this paper are two-folded. We formally define a topic **aspect evolution graph modeling** framework and propose to utilize social network information, content similarity and temporal proximity to model evolution relationships between topic aspects. The experimental results showed that, by incorporating social network information, our technique significantly outperformed content-based technique in the task of extracting evolution relationships between topic aspects.

Keywords: Social Network Analysis, Aspect Evolution Graph, Social Media.

1 Introduction

With the advance development of Web 2.0 technology, social networking websites, e.g. Facebook, Twitter, MySpace and other online communities, have become a prevailing type of media on the internet. Users consider these websites not only as networking hubs, but also as ideal sources of information and platforms of exchanging ideas. For instance, thousands of threads are developed in MySpace's news and politics forum along with the development of daily worldwide news events. However, due to the increasing number of threads and their diversity, it is extremely difficult for a user to follow these threads and fully understand the underlying blueprint. The goal of this paper is to organize threads of a given topic in an easy-to-follow fashion which reflects the evolution relationships between topic aspects, demonstrates the blueprint, and offers richer information.

A promising and effective way of organizing threads for capturing the underlying structure and easy navigation is the evolution graph, as shown in Figure 1. Given a topic of interest and a collection of threads about the topic, closely related threads can
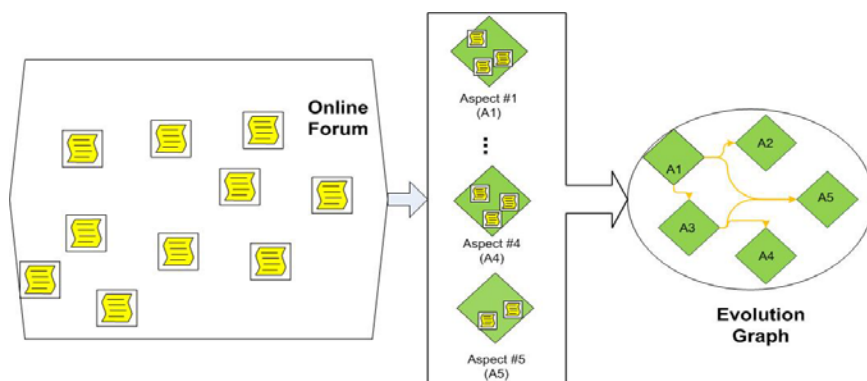
Fig. 1. Process of organizing threads from online forum into aspect evolution graph

be grouped together as topic aspects (an aspect is defined as a facet or specific subtopic of a topic here), and pairs of topic aspects with evolution relationship are linked together to form an aspect evolution graph.

Organizing threads based on topic aspects and modeling the evolution relationships between aspects are nontrivial tasks. Research in Topic Detection and Tracking (TDT) focused on the automatic techniques of identifying new events and organizing them in a flat or hierarchical structure based on their topics. However, the flat or hierarchical representation of events adopted in TDT cannot illustrate the development process of incidents [1]. To tackle this problem, Nallapati etc. [1] proposed the concept of event threading which captured the dependencies among news events. Yang etc. [2] further proposed the concept of event graph and the techniques of constructing the graph from news corpora. However, these techniques cannot be applied directly to solve the related problem in online forum, because they relied heavily on the content-based analysis techniques. Since most messages in an online forum are short and sparse, traditional TDT techniques cannot capture the topics of these threads effectively based on their content. Furthermore, a thread is very different from a news article because a news article often describes only one single story but a thread can involve multiple stories and opinions.

In this paper, our major contribution is to propose a framework of constructing aspect evolution graphs for online forum, which organizes forum threads in a clear and easy-to-follow fashion. We first cluster related threads into topic aspects, and then employ hybrid metrics to detect evolution relationship between topic aspects. Finally, we polished the evolution graph automatically by graph pruning method. To our best knowledge, there is no existing work of tracking aspect evolution and building evolution graph in the context of online forum. We conducted intensive experiments on forum data. Experimental results indicate the effectiveness of our proposed techniques.

2   Related Work

Topic Detection and Tracking (TDT) is one line of research closely related to our work, which focuses on capturing topics from a collection of documents. It also

considers how to cluster documents into different topics following either a flat-structure [3] or a hierarchical structure [4]. However, the main task of TDT does not include evolution relationship modeling as we address in this paper. Topic evolution is another line of research which focuses on how to model changes of topics over time. Some research works study topic evolution by clustering approach[5, 6]. Some other research works study topic evolution by generative models [7-9]. Although these works had their strengths on modeling dynamic topics, they focused more on topics modeling instead of the transitional flow of topic aspects. In addition, these techniques were introduced to study news article or scientific literature, their performance on forum-like data cannot be guaranteed since the forum messages are usually sparse and short.

There are a few research works concentrating on modeling event evolution on news articles. Nallapati et al. defined the concept of event threading [1]. They proposed to use a tree structure to capture the dependencies among events in news articles. The similarity between two events in their work was mainly defined by content similarity, location similarity and person name similarity. Following their work, Yang, Shi and Wei [2] introduced the event graph to describe dependencies among events in news article. They considered not only content similarity but also temporal proximity and document distributional proximity to model the relationships between events in an incident. However, none of these works consider modeling evolution relationships in the context of online forum. Forum-like data is usually short, sparse and unformatted so that it is more challenging to deal with comparing to news articles. Moreover, these existing works did not take social network information into account which can be critical in identifying the evolution relationship between topic aspects in an online forum.

3   Problem Formulation

In this section, we present the definitions of thread, aspect, topic and aspect evolution graph.

Definition 1 (Thread): A thread is a collection of user contributed messages. Its attributes include a title, timestamps, an original post which opens dialogue or makes announcement by the poster, and multiple replies from other forum members. A thread is the smallest atomic unit in our research. We represent a thread t_i with a tuple $t_i = \{W_{t_i}, N_{t_i}, opt_{t_i}\}$, where W_{t_i} is TF-IDF term vector $W_{t_i} = \{w_1^{t_i}, w_2^{t_i}, \dots, w_{|t_i|}^{t_i}\}$, N_{t_i} is the social network (defined below) associated with t_i and opt_{t_i} is the original posting time of t_i.

Definition 2 (Social Network): A social network associated with a thread t_i is a graph $N_{t_i} = <V_{t_i}, E_{t_i}>$, where V_{t_i} is a set of users, $\{p_1^{t_i}, p_2^{t_i}, \dots, p_{|V_{t_i}|}^{t_i}\}$, participated in t_i, and E_{t_i} denotes replying relationship between the users of V_{t_i}. An edge can be either undirected or directed. In this work, we only consider the directed edge. $\overrightarrow{<p_i^{t_i}, p_j^{t_i}>}$ exists *iff* $p_i^{t_i}$ replied to $p_j^{t_i}$ in thread t_i. The weight of this edge equals to the number of times $p_i^{t_i}$ replying to $p_j^{t_i}$ in t_i.

Definition 3 (Aspect): Given a collection of threads on the topic of interest, these threads can involve in different aspects (facets or subtopics) on the topic. In this paper, each aspect consists of a set of non-overlapping threads, denoted by $A_\chi = \{W_{A_\chi}, N_{A_\chi}, opt_{A_\chi}\}$ with W_{A_χ} is the centroid of the TF-IDF term vectors of the threads belonging to A_χ, $W_{A_\chi} = \frac{1}{|A_\chi|}\sum_{t_i \in A_\chi} W_{t_i}$, N_{A_χ} is the social network combining all N_{t_i} with $t_i \in A_\chi$ by taking union and $opt_{A_\chi} = \min\{opt_{t_i}|t_i \in A_\chi\}$. It's easy to extend our definition of aspect from non-overlapping threads to overlapping threads.

Definition 4 (Topic): A topic consists of a group of related aspects. It represents the common theme of these related aspects. For example, threads about ***Copenhagen Climate Conference***, ***Glacier Gate*** and ***Al Gore's Nobel Prize*** are discussing two different aspects of a common topic of global warming.

Definition 5 (Aspect Evolution Graph): An aspect evolution graph is a representation of a given topic with several aspects, where its nodes are representing the aspects and its edges are representing the evolution relationships among these aspects. An evolution relationship between two aspects in an evolution graph represents a temporal logical dependencies or relatedness between two aspects. We represent the aspect evolution graph by a tuple $AEG = \{A, R\}$, where $A = \{A_1, A_2, ..., A_{|A|}\}$ is a collection of aspects of a given topic, ***R*** is a set of evolution relationships (directed edges) among $A_i \in A$.

The major difference between this work and the previous works such as TDT and event threading is that we focus the modeling effort on evolving aspects in an online forum by making use of social network information rather than events in news articles.

Based on the definitions of these concepts, we formalize the major steps of **aspect evolution graph modeling (AEGM)** as: first of all, given a collection of threads of the same topic, $\{t_1, t_2, ..., t_m\}$, ***k*** major aspects, $\{A_1, A_2, ..., A_k\}$, are created where each aspect consists of one or more threads; secondly, given a set of aspects **A**, we compute the likelihood of evolution relationships between any pairs of aspects in **A** and then construct an aspect evolution graph; at third, graph pruning method removes those directed edges corresponding to invalid or weak aspect evolution relationships and then generates the final aspect evolution graph (**AEG**)

4 Modeling Aspect Evolution Graph with Hybrid Model

4.1 Aspect Extraction by Clustering

Every topic of discussion consists of multiple aspects. As a result, discussion threads should be clustered into aspects before modeling the evolution relationships among them. In this work, we consider the arriving threads as a finite sequence and infer the model parameters under the closed world assumption. For our online forum scenario, each thread is represented by a TF-IDF term vector. We propose to use the single-pass clustering algorithm [10] based on the TF-IDF term vectors of threads. The single-pass clustering algorithm is utilized for the following reasons: **(1)** we can apply single-pass clustering on the whole collection of threads without partitioning threads

by discretizing the time dimension into several time periods; **(2)** the single-pass clustering does not require prior knowledge, such as total number of clusters like K-Means in which the number of aspects is impossible to be pre-determined for forum threads; **(3)** the single-pass clustering is more efficient than other clustering approaches which require the computation of the similarity between all pairs of objects, such as hierarchical clustering; **(4)** it is natural and logical to cluster threads incrementally according to their posting time.

4.2   Modeling Evolution Relationship

Given a collection of aspects generated by single-pass clustering algorithm, we propose to utilize text information, temporal information, and social network information to model the evolution relationships between them.

4.2.1   Content-Based Measurement

As described in [2], a simple way to estimate the likelihood that an aspect evolution relationship exists from an aspect A_i to A_j is based on the content-based similarity and temporal information. As given in **Definition 3** of Section 3, a thread is represented as a TF-IDF term vector and an aspect is represented by the centroid of the term vectors in the aspect. Thus, the content similarity between two aspects is the cosine similarity of their centroid term vectors, defined by $\cos(W_{A_i}, W_{A_j})$. Other than the content-based similarity, the temporal order and temporal distance between two aspects are also useful to determine the evolution relationship. We assume that if one aspect i occurs later than another aspect j $(opt_{A_i} > opt_{A_j})$, it is impossible to have an evolution relationship from i to j. Moreover, if two aspects are far away from each other along the timeline, the evolution is less likely to happen between them. On the contrary, if two aspects are temporally close, it's more likely that one will evolve to another. *Temporal proximity* is used to measure relative temporal distance between two aspects, defined as the follow decaying function:

$$tp(A_i, A_j) = e^{-\alpha[diff(opt_{A_i}, \ opt_{A_j})/T]}$$

where α is a time decaying factor between 0 and 1, **diff(•)** denotes the absolute value of time difference between two aspects, **T** denotes the temporal distance between the start time of the earliest aspect timestamp and the end time of the latest aspect timestamp in the same topic. Based on the notations above, the likelihood of aspect evolution relationship from A_i to A_j is defined as:

$$score(A_i, A_j) = \begin{cases} 0 & if \ opt_{A_i} > opt_{A_j} \\ \cos\left(W_{A_i}, W_{A_j}\right) \times tp(A_i, A_j) & if \ opt_{A_i} < opt_{A_j} \end{cases}$$

4.2.2   Content-Based Measurement + Local Social Network Analysis

Although content-based similarity is an effective way to capture the evolution relationships between events in news corpora, it cannot guarantee a satisfactory performance on the forum data. Unlike news articles, forum messages are very short and sparse, which make the text similarity measurements less effective. On the other hand, if two aspects are closely related, users who are active in discussing one aspect

will tend to be active in discussing another aspect. Thus, two aspects attracting similar groups of users are more likely to be related and evolved from one to another. On the contrary, if two aspects share relatively high content-based similarity but few overlapping between their participants, it is less likely that these two aspects are related. Following this argument, we propose measurements by regularizing content-based measurement with social network analysis techniques.

To regularize the content-based measurement with the intersection of important participants involving in two aspects, we proposed the *Local Social Network Analysis* technique. It is worth emphasizing that, instead of counting the number of overlapping participants in both aspects only, we assign a significant score to each individual participant of the aspect. Two groups of participants belonging to two aspects can be considered as overlapping if most of the **important** participants overlap, even though many unimportant participants are different. A formal definition of $Overlap(N_{A_i}, N_{A_j})$ between two aspects A_i and A_j as:

$$Overlap\left(N_{A_i}, N_{A_j}\right) = \frac{\sum_{P_a \in N_{A_i} \cap N_{A_j}} \min(SS(P_a^{A_i}), SS(P_a^{A_j}))}{\sum_{P_b \in N_{A_i} \cup N_{A_j}} \max(SS(P_b^{A_i}), SS(P_b^{A_j}))}$$

where $SS\left(P_k^{A_m}\right)$ represents the significant score of the participant P_k in aspect A_m, $N_{A_i} \cap N_{A_j}$ denotes the intersection of participants of two social networks, and similarly $N_{D_i} \cup N_{D_j}$ denotes the union of participants of two social networks. From the definition above we can learn that the larger the intersection of participants involving in both aspects, the higher the value of **Overlap(•,•)**. Moreover, given a participant involving in both aspects, the smaller the difference between the significant scores of this participant in two social networks, the higher the value of **Overlap(•,•)**. In this work, we explore different ways of measuring significance of an individual participant in a social network, which includes: ***IN-DEGREE, OUT-DEGREE, IN-DEGREE PLUS OUT-DEGREE, BETWEENNESS, CLOSENESS*** and ***PAGERANK***. Based on the notations above, the likelihood of aspect evolution relationship between A_i and A_j regularizing by *Local Social Network Analysis* is defined as:

$$score(A_i, A_j)' = \begin{cases} 0 & \text{if } opt_{A_i} > opt_{A_i} \\ \left(\varepsilon \times \cos\left(W_{A_i}, W_{A_j}\right) + (1-\varepsilon) \times Overlap\left(N_{A_i}, N_{A_j}\right)\right) \times tp\left(A_i, A_j\right) & \text{if } opt_{A_i} < opt_{A_i} \end{cases}$$

where N_{A_i} and N_{A_j} are the networks associated with aspects A_i and A_j and $0 < \varepsilon < 1$.

4.3 Pruning Aspect Evolution Graph

Following the techniques proposed in Section 4.2, an aspect evolution graph with a high density of edges is constructed. A pruning technique is needed to remove those directed edges corresponding to invalid or weak aspect evolution relationships by considering the weights of edges. In this work, we simply set a static threshold λ for graph pruning. Every relationship with weight lower than the threshold will be removed.

5   Datasets

In this section, we present experiments to evaluate the aspect evolution graph modeling (AEGM) framework. We built our own crawler and retrieved threads in the topics of Haiti earthquake, Nobel Prize and Global Warming by keyword searching. We constructed three datasets for these topics, which include: half month of threads discussing Haiti earthquake (**HE** dataset), 3 months of threads about President Obama wins Nobel Prize (**NP** dataset), and 3 years of threads discussing Global Warming (**GW** dataset). HE dataset contains 17 threads and 1,188 messages. NP dataset consists of 35 threads and 2268 messages. GW dataset contains 36 threads and 5627 messages. After crawling all related threads of these three topics. We solicited human annotators to cluster threads into aspects and then create evolution relationships between these aspects. The aspect evolution graphs created by human annotators are considered as ground truth for comparison in the experiment part.

6   Experiments and Analysis

6.1   Evaluation Measures

Taking the aspect evolution graphs $AEG_{true} = \{A, R\}$ generated by human annotators as the ground truth, we employed Precision, Recall and F-Measure to evaluate the aspect evolution graph $AEG_{predict} = \{A, R'\}$, which is created automatically by our proposed techniques. Defined formally, Precision, Recall and F-Measure can be computed as follows: $sion = |R \cap R'|/|R'|$, $Recall = |R \cap R'|/|R|$.

6.2   Experiments

As introduced in section 4.2.2, we combine **Overlap** function with text similarity to measure the likelihood of evolution relationship between two aspects. We proposed different possible ways of measuring the significant of individuals in this **Overlap** function.

Table 1. Evolution relationship detection on HAITI dataset (HE), Nobel Prize dataset (NP) and Global Warming dataset (GW) pruned by static thresholding with $\lambda = 0.1$

	Best Parameter			Precision			Recall			F Measure		
	HE	NP	GW	HE	NP	GW	HE	NP	GW	HE	NP	GW
Text	$\varepsilon=1$	$\varepsilon=1$	$\varepsilon=1$	0.45	0.41	0.46	0.87	0.3	0.58	0.59	0.35	0.51
TI	$\varepsilon=0$	$\varepsilon=0.8$	$\varepsilon=0.9$	0.73	0.58	0.27	0.53	0.65	0.63	0.62	0.61	0.57
TO	$\varepsilon=0.2$	$\varepsilon=0.8$	**$\varepsilon=0.7$**	0.57	0.57	**0.61**	0.87	0.7	**0.58**	0.68	0.63	**0.59**
TIO	$\varepsilon=0.5$	$\varepsilon=0.7$	$\varepsilon=0.7$	0.56	0.53	0.58	0.93	0.74	0.58	0.7	0.62	0.58
TC	**$\varepsilon=0.1$**	**$\varepsilon=0.8$**	$\varepsilon=0.1$	**0.77**	**0.59**	0.55	**0.67**	**0.7**	0.58	**0.71**	**0.64**	0.56
TB	$\varepsilon=0.5$	$\varepsilon=0.8$	$\varepsilon=0.8$	0.67	0.5	0.63	0.67	0.35	0.53	0.67	0.41	0.57
TP	**$\varepsilon=0$**	**$\varepsilon=0.8$**	**$\varepsilon=0.8$**	**0.59**	**0.67**	**0.55**	**0.87**	**0.7**	**0.63**	**0.7**	**0.68**	**0.59**

In the experiment, we attempted to compare the effectiveness of different aspect evolution relationship scoring functions, including: (1) **TEXT:** Temporal proximity

multiplied by aspect content similarity; (2) **TEXT+INDEGREE (TI):** Temporal proximity multiplied by weighted sum of aspect content similarity and value of Overlap function using INDEGREE as significant score; (3) **TEXT+OUTDEGREE (TO):** Temporal proximity multiplied by weighted sum of aspect content similarity and value of Overlap function using OUTDEGREE as significant score; (4) **TEXT+INDEGREE+OUTDEGREE (TIO):** Temporal proximity multiplied by weighted sum of aspect content similarity and value of Overlap function using INDEGEE+OUTDEGREE as significant score; (5) **TEXT+CLOSENESS (TC):** Temporal proximity multiplied by weighted sum of aspect content similarity and value of Overlap function using CLOSENESS as significant score; (6) **TEXT+BETWEENNESS (TB):** Temporal proximity multiplied by weighted sum of aspect content similarity and value of Overlap function using BETWEENNESS as significant score; (7) **TEXT+PAGERANK (TP):** Temporal proximity multiplied by weighted sum of aspect content similarity and value of Overlap function using PAGERANK as significant score. Due to the limited space, we cannot present all of the results in parameter tuning. In Table 1, we summarized the best performance of our proposed techniques with the optimal value of the parameter ε. For each dataset, we highlighted two techniques with the highest F-Measure. Based on the results, TEXT+CLOSENESS and TEXT-PAGERANK achieved the best performance. TEXT+OUTDEGREE was slightly behind.

7 Conclusion

In this paper, we have investigated a new perspective of organizing forum threads and tracking evolving aspects in an online forum. We have presented the frame work of **aspect evolution graph modeling** for constructing aspect evolution graphs automatically from a collection of threads. By incorporating social network analysis techniques, we are able to address the weakness of content-base techniques. We proposed various approaches of combining content-based and network-based techniques of predicting evolution relationship between two aspects. The experimental results demonstrated a significant improvement of our methods. Future work can include designing more effective approach to combine content-based and network-based techniques to further improve prediction accuracy, or developing more powerful algorithm to cluster threads into aspects automatically.

References

1. Nallapati, R., Feng, A., Peng, F., Allan, J.: Event threading within news topics. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, pp. 446–453. ACM, New York (2004)
2. Yang, C., Shi, X., Wei, C.: Discovering event evolution graphs from news corpora. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 39, 850–863 (2009)
3. Allan, J., Carbonell, J., Doddington, G., Yamron, J., Yang, Y.: Topic detection and tracking pilot study: Final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, vol. 1998, pp. 194–218 (1998)

4. Allan, J., Feng, A., Bolivar, A.: Flexible intrinsic evaluation of hierarchical clustering for TDT. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management, pp. 263–270. ACM, New York (2003)
5. Morinaga, S., Yamanishi, K.: Tracking dynamics of topic trends using a finite mixture model. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 811–816. ACM, Seattle (2004)
6. Schult, R., Spiliopoulou, M.: Discovering emerging topics in unlabelled text collections. In: Manolopoulos, Y., Pokorný, J., Sellis, T.K. (eds.) ADBIS 2006. LNCS, vol. 4152, pp. 353–366. Springer, Heidelberg (2006)
7. Mei, Q., Zhai, C.: Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 198–207. ACM, New York (2005)
8. Blei, D.M., Lafferty, J.D.: Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 113–120. ACM, Pittsburgh (2006)
9. Wang, X., McCallum, A.: Topics over time: a non-markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 424–433 (2006)
10. Yang, Y., Pierce, T., Carbonell, J.: A study of retrospective and on-line event detection. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 28–36. ACM, Melbourne (1998)

Representing Trust in Cognitive Social Simulations

Shawn S. Pollock, Jonathan K. Alt, and Christian J. Darken

MOVES Institute, Naval Postgraduate School
Monterey, CA 93943
{sspolloc,jkalt,cjdarken}@nps.edu

Abstract. Trust plays a critical role in communications, strength of relationships, and information processing at the individual and group level. Cognitive social simulations show promise in providing an experimental platform for the examination of social phenomena such as trust formation. This paper describes the initial attempts at representation of trust in a cognitive social simulation using reinforcement learning algorithms centered around a cooperative Public Commodity game within a dynamic social network.

Keywords: trust, cognition, society.

1 Introduction

One of the fundamental phenomena governing human interactions is the notion of trust, without which the fabric of society quickly comes unraveled. Trust in other humans and societal institutions facilitate a market economy and a democratic form of government. The human information processing system, adept at receiving and synthesizing large amounts of sensory information from our environment, manages to identify which percepts are salient to our current task or our long term well being by the allocation of selective attention [1]. In this view the level of trust is a quality associated with each percept either directly, one sees an event as a first person observer, or indirectly, the information comes from a reliable source. The latter case, involving the interaction and development of trust between individuals, will be the topic of this paper. The representation of trust within cognitive social simulations is of fundamental importance to the exploration of macro level social phenomena.

When humans interact with one another there is a multitude of information, both verbal and nonverbal, that is exchanged between the participants. In social simulations the major issue in modeling this phenomenon is to understand more fully how information flows through a social network and how the society develops and evolves its beliefs over time [2]. Central to the modeling of information flow in social systems is the concept of trust. The main contribution of this paper is a model of trust based on reinforcement learning and demonstrated in the context of the Public Commodity Game [3].

This paper first provides a brief introduction to trust, reinforcement learning, and cognitive social simulation. This is followed by a description of the general trust model as well as the current Python implementation of the model within the context

of the Public Commodity game. Computational results of experimentation with model parameters related to the formation of norms and penalties for the violation of norms are provided as well as a summary and discussion of future work.

2   Background

This section provides an overview of the notion of trust as used in this setting, reinforcement learning, and cognitive social simulation.

2.1   Trust

In order to model trust, it is first necessary to define trust. Trust is not a simple concept (though it is easy to express its general intent) and codifying trust to the extent of incorporation into computer simulation is rather difficult. This work establishes a definition of trust suitable for use in simulation rather than a general definition. In particular the definition of trust used is chosen with an eye toward the long term goal of the research, to model communication and belief revision. Trust is viewed as an analysis performed by an individual agent to prejudice newly obtained information either based on the sender or the topic of discussion, and the history between the sender and the receiver and topic [4]. As an example, if a person who is completely trusted by another shares some information, the receiver is more likely to welcome the communication and take action on it. Additionally, if the sender is moderately trusted by the individual, but the receiver is untrusting of the topic (such as a particular political view), they may disregard the information as non-actionable. Simply making a binary decision of actionable versus non-actionable information is not robust enough to capture the nuances of human behavior requiring the incorporation of some concept of the grey area in between. The model incorporates sender and topic into the state space of a reinforcement learning algorithm to develop a two-pass notion of trust where the receiving agent first determines whether information is actionable then makes a separate decision if he should or should not revise his beliefs based on this new information.

2.2   Reinforcement Learning

An appealing approach to represent human like learning and action selection is the idea of reinforcement learning, where agents will seek select actions within their environment based on their experience. Based on the permissiveness of the environment, agents are eligible to receive percepts from the environment that inform them on the state of the environment at a given point in time. The basic elements of reinforcement learning are: a *policy* that maps states to actions; a *reward function* that maps a state of the environment to a reward; a *value function* that maps states to long term value given experience; and an optional *model* of the environment [5]. The policy provides a set of actions that are available in a given state of the environment; the agents leverage their prior knowledge of the environment, informed by the value function, to determine which action will provide the greatest reward, as defined by the modeler. Agents must strike a balance between exploration, behavior to explore the reward outcomes of state action pairs that have not been tried, and exploitation,

behavior that takes advantage of prior knowledge to maximize short term rewards, in order to avoid converging to local minima [5]. The ability to control this balance makes reinforcement learning an attractive approach for representing human behavior. The reinforcement learning technique used in this work is Q-learning in conjunction with a softmax function (the Boltzmann distribution).

Q-learning, $Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma max_{a'} Q(s',a') - Q(s,a))$, falls into a class of model free reinforcement learning methods that have the property that the learned action-value function, Q, approximates the optimal action-value function, Q*, requiring only that all state action pairs be updated as visited [5]. For each state action pair, *(s,a)*, the Q-learning function updates the current estimate based on new information received from recent actions, *r*, and discounted long term reward. In general, an action is selected from a given state, the reward outcome is observed and recorded, and the value function updated. The value associated with each action is used during each visit to a particular state to determine which action should be chosen using the Boltzmann distribution, shown below.

$$P_i = \frac{e^{\hat{Q}(\sigma,a)_i / \tau}}{\sum_j e^{\hat{Q}(\sigma,a)_j / \tau}} \quad (1.1)$$

The Boltzmann distribution uses the temperature term, τ, to control the level of exploration and exploitation. A high temperature translates into exploratory behavior, a low temperature results in greedy behavior.

2.3 Cognitive Social Simulation

Cognitive social simulations are particularly well-suited for defense applications which typically call for analysis of a military force's performance while operating as part of a complex conflict ecosystem [6]. Agent based models have been applied to the military domain previously [7], but the use of cognitive social simulation, enabled by cognitive architectures, in this application area is relatively new [8]. The relevancy of these tools is particularly highlighted by the nature and objectives of the current conflicts, where the population of the conflict area is seen as the center of gravity of the operation [9]. Gaining an understanding of potential means of transitioning the social system in these countries from an unstable to a stable state provides a challenge to leaders at all levels of the military and civilian organizations involved. The representation of trust and its impact on the effectiveness of information operations is required in order to provide greater insight into the population within the area of interest.

3 Approach

This section provides an overview of a general model of learned trust for agents in cognitive social simulations, an introduction to the test bed environment, and a description of the proof of principle implementation in Python 2.6.

3.1   General Model

This general model is a turn-based simulation in which the agents and their relationships are represented by a single network graph with the agents as the nodes and their social relationships as the edges, weighted by the value of the relationship. To be specific the network graph is a combination of two similar graphs with identical nodes, but in one the edges are bidirectional and represent a base constant value and the second one in place of each bidirectional edge contains a pair of directed edges representing the agent's individual variable contribution to the strength in the relationship. In this way, the edge weights have both a static and dynamic component. The static is based entirely on the concept of homophily (E_H), that like persons associate more frequently, utilizing a simple Euclidean distance calculation and as stated above is set as a constant [10]. The dynamic portion is completely under the control of the agents involved ($E_{A \to B}$). In the case of k total agents, each agent has k-1 choices of who to spend time with and based on that emphasis will receive some unknown reward from these relationships. For every simulation round, each agent will choose to increase, decrease or maintain their contribution to their relationships with the other agents. This contribution can be seen like a fraction of time spent with the others in that it is represented as a floating point number from 0.0 to 1.0 and such that the sum of all these components (i.e. the sum of all edge weights leaving the node representing the agent) always sums to 1.0. At the end of each turn the agent is rewarded based on the strength of their relationships. This reward takes the following form:

$$Reward = E_H \cdot \min (E_{A \to B}, E_{B \to A}) \qquad (1)$$

The equation uses the minimum of the variable contributions from each agent. In this way it more accurately can be said that the variable portion is the fraction of time that the agent would *like* to spend with the other agent and therefore using the min also provides a small penalty for those who place more emphasis on a relationship than the recipient.

The result of this basic model is the development of a simple dynamic social network. The network tends to become highly centralized around 1 or 2 agents. In particular, in runs consisting of 50 agents, the final network graph consisted of nearly every agent with a strong connection to a single central agent with no other connections present. In order to mitigate this affect it was necessary to add a second order factor in the reward calculation for the agents. Continuing along the analogy that the emphasis represents a fraction of time desired to be spent with the other agents, then it is natural to extend this and allow for groups of more than 2 agents to spend time together. In other words if two agents have a strong relationship and also each have a strong relationship to the same third agent, then all three agents should receive an additional reward for this.

The second order reward factors are based on the same reward function as used in the first order above. In this case, the reward is divided by a distribution factor and subsequently squared. For the case of agents A and B as above, but this time having a common friend in agent C, the additionally reward looks as below:

$$2nd - Reward = \min(E_H \cdot \min(E_{A \to C}, E_{C \to A}) , E_H \cdot \min (E_{A \to C}, E_{C \to A})) / D \qquad (2)$$

The closeness centrality of the network is highly sensitive to the distribution factor and is discussed in detail in section 4.

Once the second order terms are added similar network properties to what we would expect to see in real social situations emerge; namely subdivisions into clique's, pairings and the exclusion of certain individuals. It is obvious that this feature is not intended to actually model the internal and external processes that form human social networks; rather it is simply building the stage on which to build future work.

3.2 Prototype Trust Implementation

Each agent has a simple belief structure consisting of a finite set of beliefs, five in this case, represented by a single floating point number from 0.0 to 1.0. These beliefs combine in simple linear combinations to provide issue-stances, one in this case, also as a floating point from 0.0 to 1.0. During each turn of the simulation, following an initial stabilization period, set initially to 1000 rounds for a simulation of 15 agents, the agent will choose a topic based on a probabilistic Boltzmann distribution, and discuss this topic with its k nearest neighbors. Other than the k nearest there any neighbor above a specified threshold will always receive a communication, while neighbors below another threshold will never receive a communication.

Initially, the communications consist of each agent telling his closest neighbors what his value is on a selected belief. The receiving agents then use a reinforcement learning algorithm to determine whether or not belief revision is merited. In order to utilize reinforcement learning it is necessary to define some concept of a reward that the agent will receive based on their beliefs and therefore directly related to their trust and belief revision mechanisms. Our inspiration for a reward model comes from Game Theory and is called the Public Commodity Game [3]. In this game, each agent has an option to contribute to a public pot of money each round or to opt out. Following the round the money in the pot is multiplied by some amount (in this case 3.0) and then redistributed to each agent regardless of contribution. As a slight variation of this, the agents are allowed to decide an amount to contribute rather than to choose from full or no contribution.

In the current model, agents are given 1.0 possible units to play such that an agent that contributes nothing is guaranteed a reward of at least 1.0 for opting out and an unknown reward ranging from nearly 0 to 3.0 for full contribution. Game theory tells us that without cooperation the expected equilibrium for rational players would be exactly 0.0 contributions from all agents, in other words all agents take the guaranteed 1.0 and opt-out of the public commodity all together [3]. It is likely the case that some people would always contribute at least some small amount irrespective of their losses. In order to achieve this, "Faith in the Public Commodity" is the issue-stance and is used to directly control the level of their contribution to the public commodity. During each simulation round, agents communicate with one another and attempt to bring other agents closer to their beliefs. Despite the fact that a 1.0 contribution from all agents is the most mutually beneficial strategy, it is not a stable equilibrium as a single person could quickly realize that decreasing their contribution will increase their total revenue and would be easily reinforced by learning algorithms. What is seen is the agents benefit the most from a strategy of trying to make all the other agents contribute at a level consistent with their issue strength.

Now that there is a concrete idea of reward in this model, it is possible to begin applying a simple model of belief revision. Essentially, the agent will use a reinforcement learning algorithm where the state space is represented by sender-topic pairs and the action will be to dismiss the communication or to apply a revision, based on their level of trust in the information, resulting in an update of the fraction of time they desire to spend with the sender. Without any further weighting factors or limiting algorithms, the expected result of this simulation is that the beliefs of all agents will approach a stable equilibrium in which all beliefs are the same.

4   Experimentation and Analysis

This section provides a comparison of model results from several attempts to moderate belief revision within the model.

4.1   Experimentation of Second Order Factors in the Base Social Network

The first task was to evaluate the effects of the strength of the second order terms in the base social network reward functions. The distribution factor was varied and showed a fairly steep "S" curve that was centered between D = 14 to D = 24.
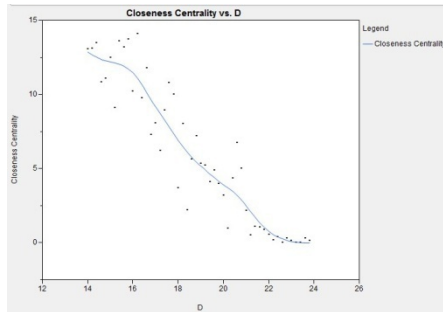
Fig. 1. Closeness centrality versus distribution factor

The distribution factor in affect allows fine tuning of closeness centrality in the base model in order to fit it to any particular purpose. There are several widely varying sources on what a real human social network should look like in terms of closeness centrality that range from 0.20 to 0.60. Therefore, for the purposes of the remainder of this initial experimentation D = 18.4 is used in order to target the fractional closeness centrality to around 0.30. The exact nature of these values is irrelevant for this initial model and only serves as a baseline for further work.

4.2   Belief Revision Experimentation

As discussed previously it is necessary to implement within the model a method for moderating how the agents can revise their beliefs. Initial methods for doing this are, at least for now, centered on a penalty for changing beliefs.

$$Net\ Dividend = (1.0 - Contribution) + 3.0\ x\ Contribution - NormPenalty$$
$$NormPenalty = e^{F*BeliefVariance}$$

Where the Belief Variance is a simple Euclidean distance measure from the agents current beliefs to what they started with at the beginning of the simulation. What is surprising is that when F is varied, there is no marked difference in the outcome of the simulation from a purely statistical point of view. What is seen however is in the Public Commodity play over time.
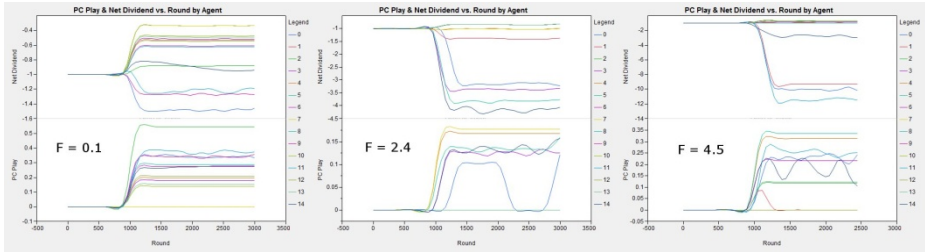
Fig. 2. Public commodity game play versus norm penalty

The higher the factor F becomes, the more unstable the Public Commodity game is. In other words, with a small norm penalty the agents will tend to find a stable equilibrium and remain there with fairly significant stability. As F is increased, the stability is decreased. The intriguing thing is that this behavior appears to be similar to actual human interaction. For example, if we look at a society there is a sense of a norm although it will really change over time it will remain fairly constant over small enough time periods. In this society there will be people or factions that challenge the social norm causing brief unstable equilibrium away from the norm that seem to return to the social norm after some time. More investigation into this phenomenon is currently underway. Once there is a satisfactory explanation for this behavior it may be possible, just as in the second order parameter, to tune the magnitude of the norm penalty in a way that is unique to the society being modeled.

5   Conclusions and Future Work

The results of the initial work are promising in that it is seen that there are factors within this simple first draft model that allow the social network to form and evolve based purely on the agents cognitive processes however the nature of the network can be tuned to match a desired case. This initial effort to model trust shows a lot of promise and merits further investigation. The next step in this process will be to utilize this algorithm within an existing social simulation and evaluating the effects.

The next generation of this model is planned to include a much more complicated belief structure that will include several metacognitive elements giving the agents some control over belief revision. There will also be included within the belief structure mental models of the other agents and the environment so that a perceived closeness in beliefs of the other agents will play directly into trust.

References

[1] Anderson, J.: Cognitive Psychology and Its Implications, 6th edn. Worth Publishers, New York (2005)

[2] Sun, R.: Cognitive social simulation incorporating cognitive architectures. IEEE Intelligent Systems, 33–39 (2007)

[3] Owen, G.: Game Theory. Academic Press, San Diego (1995)

[4] McKnight, D., Chervany, N.: The Meanings of Trust. University of Minnesota (1996)

[5] Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)

[6] Kilcullen, D.J.: Three Pillars of Counterinsurgency. In: US Government Counterinsurgency Conference

[7] Cioppa, T.M., Lucas, T.W., Sanchez, S.M.: Military applications of agent-based simulations. In: Proceedings of the 36th Conference on Winter Simulation, p. 180 (2004)

[8] National Research Council, Behavioral Modeling and Simulation: From Individuals to Societies. National Academies Press, Washington DC (2008)

[9] C. O. S. W. Department of the Army, Counterinsurgency. U.S. Army (2006)

[10] McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology 27(1), 415–444 (2001)

Temporal Visualization of Social Network Dynamics: Prototypes for Nation of Neighbors

Jae-wook Ahn[1,2], Meirav Taieb-Maimon[2,3], Awalin Sopan[1,2],
Catherine Plaisant[2], and Ben Shneiderman[1,2]

[1] Department of Computer Science
[2] Human-Computer Interaction Lab, University of Maryland,
College Park, MD, USA
{jahn,awalin,plaisant,ben}@cs.umd.edu
[3] Ben-Gurion University of the Negev, Beer-Sheva, Israel
meiravta@bgu.ac.il

Abstract. Information visualization is a powerful tool for analyzing the dynamic nature of social communities. Using Nation of Neighbors community network as a testbed, we propose five principles of implementing temporal visualizations for social networks and present two research prototypes: NodeXL and TempoVis. Three different states are defined in order to visualize the temporal changes of social networks. We designed the prototypes to show the benefits of the proposed ideas by letting users interactively explore temporal changes of social networks.

Keywords: Social Network Analysis, Information Visualization, Social Dynamics, User Interface, Temporal Evolution.

1 Introduction

Information visualization is a powerful tool for analyzing complex social networks so as to provide users with actionable insights even against large amount of data. Numerous social network analysis (SNA) methods have been coupled with visualization to uncover influential actors, find helpful bridging people, and identify destructive spammers. There are many research tools and a growing number of commercial software packages, some designed for innovative large scale data analysis, while others deal with common business intelligence needs. However, few approaches or tools sufficiently address the problem of how to analyze the social dynamics of change over time visually. Communities are not static. Like living organisms, they evolve because of cultural, environmental, economic, or political trends, external interventions, or unexpected events [4]. Technological developments have also had strong impacts on social changes, a phenomenon that has become influential with the arrival of mobile communications devices and social networking services.

Nation of Neighbors (http://www.nationofneighbors.com) (NoN) is a new web-based community network that enables neighbors to share local crime, suspicious activity, and other community concerns in real time. The NoN developers

have achieved an admirable success with "Watch Jefferson County" that empowers community members to maintain the security in their local neighborhoods. It began in Jefferson County, WV, but the NoN team is expanding their efforts across the U.S. in many communities. We are collaborating with them to provide appropriate tools that can help community managers explore and analyze the social dynamics embedded in their social networks.

This paper introduces our first efforts to visualize the temporal social dynamics on top of the NoN testbed. Due to the nature of the social dynamics, we need a method to interactively *compare* the time slices of social networks. In previous attempts, we combined statistics and the visualization [12] for interactive network analysis and to support multiple network comparisons using tabular representations [3]. The current work integrates the best of these two approaches by using the graphical representations and statistics to enable researchers and community managers to compare networks over time.

2   Related Studies

The importance of capturing the network change over time has been the origin of previous attempts of temporal network visualizations. Two common approaches were used. The first approach plots network summary statistics as line graphs over time (e.g. [1,7]). Due to its advantage in detecting the increase and decrease of certain statistics, many systems adopted this idea. In the VAST 2008 mini challenge 3:"Cell Phone Calls" [8] various teams (e.g. SocialDynamicsVis and Prajna teams) used similar approaches such as line graphs to characterize changes of the cell phone call social network over ten days. SocialAction integrated a time-based interactive stacked graph to display a network property such as degree. Each stack represented a node and the thickness of each stack on a given day represented its degree. The interactive visualization facilitated the exploration of the change in the degree of each node over time.

The second approach is to examine separate images of the network at each point in time. Powell et al. [13] used such a snapshot approach to show network changes over time. They distinguished new arrivals with triangles and incumbent with circles. They also used size encoding for degree; a shortcoming of such size and shape encoding might be that the new arrivals having small degree are very small and sometimes the triangles and circles are too small to be distinguishable. Furthermore, in snapshot view we cannot compare a network easily in two timeslots as the layout algorithms can dynamically change the node positions. Durant et al. [2] presented a method for assessing responses and identifying the evolution stages of an online discussion board. Their snapshot-based network visualizations could show different node positions between time.

Recent work offered improvements in static visualizations by applying dynamic visualization such as network "movies." Moody [11] distinguished flip-book style movies where node-positions were fixed but connectivity formation was captured and dynamic movies where node-positions changed over time. Condor (or TeCFlow) [5,6] introduced sliding time frames and animated the network

changes inside a specific time frame with/without the history before that. However, approaches which use animations might be distracting for users to track changes in the network. Specifically, it might be difficult to track new nodes when the position of the nodes and edges keep changing. Human eyes can miss minute but important changes in visual spaces. For example, users can only notice the overall growth of an entire network while missing that the inclusion of a certain node is responsible for a significant change. A possible suggestion can be found in Trier's work [14] where he used node degree as a measure of inertia so the high degree nodes move less across time and made the dynamic movie less distracting. C-Group [10] visualized the affiliation of focal pairs and dynamically defined groups over time using static, dynamic layouts, or animations. It fixed the focal pair positions and could minimize distraction in the animation mode.

In the present study, we suggest an approach that keeps the nodes from earlier time slots in fixed positions but also accentuates new arriving nodes and links using color coding while using intensity as a measure of aging. This overcomes the distraction caused by dynamic fast paced network movies and also fulfills the need of identifying the deletion and addition of nodes.

3   Social Network Dynamics Visualization – Ideas

Our main concern about the temporal dynamic visualization for social networks is how to *show* and *compare* the changes over time. We established the following principles for this purpose.

1. Static (unchanged) parts of a graph should remain fixed to avoid distraction.
2. Changes should be visually manifest for easy comparisons.
3. Comparisons across time should be easily and interactively explorable.
4. Temporal change of the node and edge attributes should be discoverable.
5. Temporal change of a sub-graph and its attributes should be discoverable.

Some approaches do not fix the locations of the static elements, but since users' main interests are about what is *changing*, fixing the locations of the unchanged nodes and edges is advantageous. For the changing components, we defined three states: (1) **addition**, (2) **removal**, and (3) **aging**.

In social networks, new relationships can be added or removed to/from the entire network. At the same time, existing relationships can become inactive after they were added to the network. Especially, there are instances where the removal of the relationships is not quite clear and older activities are just becoming inactive. For example, a reference once made in a citation network cannot be removed, even though it can be less important as time passes. We tried to distinguish these three states in our network visualization. Finally, users need to easily compare the network at any pair of time points.

4   Social Network Dynamic Visualization Systems

Based on these principles, we built two example systems: (1) by using NodeXL and (2) by adopting a time-slider for temporal exploration. NodeXL [9] is an
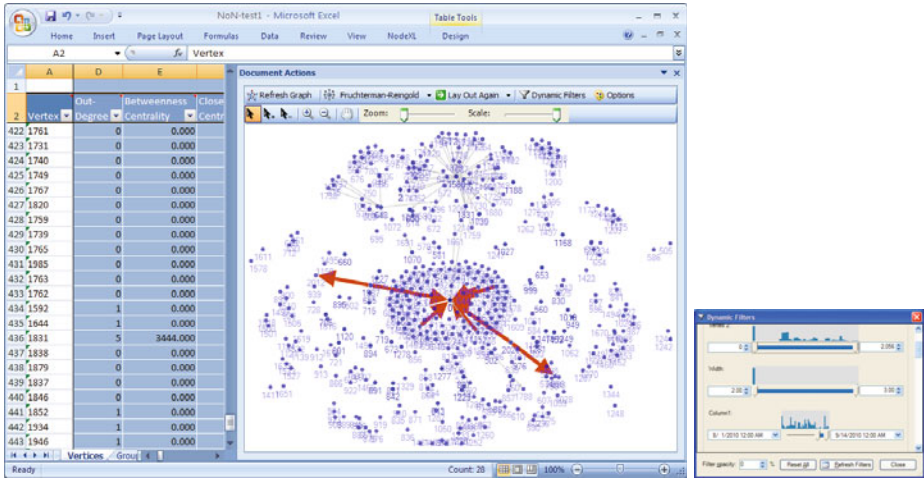
Fig. 1. Visualizing the NoN invitation network dynamics using NodeXL. Nodes and edges mean users and invitations respectively. Red edges are the ones in September 2010. Using the dynamic filter (right), it can filter out the edges before August and emphasize more recent invitations (8/1/2010 – 9/14/2010, the edges in the upper part).

extension to Microsoft Excel 2007/2010 that makes the SNA flexible and interactive. The node-edge connectivity data stored in Excel worksheets are directly converted to graph visualizations. Popular SNA statistics can be calculated and then be mapped to visual attributes. It is highly configurable by users through the common Excel operations and the NodeXL metrics, visual properties, clustering, and layout control panels. The second system called TempoVis was built as a standalone application prototype, in order to implement a time-based network exploration more interactively. We used two datasets provided by NoN: user invitation and forum conversation networks. The former contained information about who invited whom to the NoN service (July 2009 – September 2010). The latter showed which user created a thread and who replied to whom in the NoN forums (November 2005 – August 2010) with timestamps.

4.1   Spreadsheet-Based Approach Using NodeXL

Fig. 1 shows an example of temporal visualization using NodeXL. The worksheet stores the node (vertex) and edge information. The graph theory statistics such as centrality are calculated for each node too (for example, Vertex 1831's Betweenness Centrality=3444.0). In the right-hand side panel, the network is visualized using the well-known Fruchterman-Reingold force-directed layout algorithm (the layout was computed from the latest time point). It visualizes the invitation network of the NoN members accumulated for 14 months.

   The overall network layout does not change by the dynamics of the invitation network over time. However, the temporal dynamics in the individual invitation level are represented clearly. The red edges (arrowed) in Fig. 1 indicates that

they were added in a specific time point, September 2010. At the same time, the nodes and edges created earlier than a specific time (or removed) can be painted in different colors or dynamically filtered out as in Fig. 1 in order to implement the three states (Section 3). The graph elements are tightly linked to the worksheet, so that users can examine specific attributes of any nodes, edges, or subgraphs interactively while they are switching between the worksheet and the visualization. However, despite the very rich and flexible feature set that helps users to compare the data and the visual representation interactively, easy exploration across multiple time points is yet to be added to NodeXL. Even the dynamic filter requires some manual endeavors. Therefore, we built the next prototype that focused more on easier time-based exploration.

4.2 TempoVis: Adding Time-Based Interactive Exploration

TempoVis (Fig. 2) is a standalone prototype that shares the same principles but we designed it to maximize the ability to explore the time interactively. It is equipped with a time-slider with which users can navigate through different time points (months in this example). They can move the slider thumb left and right, to view the different monthly states. Users can navigate back and forth through
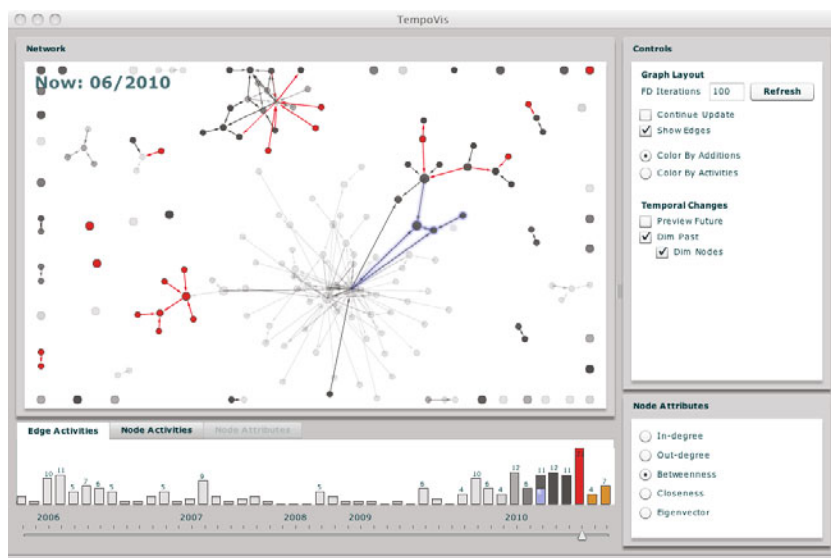
Fig. 2. TempoVis: Navigating through time using a time-slider interface (NoN conversation network). The nodes and the edges mean users and their conversations respectively. The red edges indicate the conversations of a specific month (June 2010) set by using the slider blow. The big low-intensity cluster (center below) was introduced earlier than the high intensity cluster (center top) that was added recently. Blue edges mean user selections. Node size is proportional to graph statistics dynamically calculated at each time point.

time to study the growth of the network. On top of the slider, a histogram shows the frequency of the activities. This arrangement enables users to see the overview of the activities and compare the changes over time during navigations.

TempoVis uses the force-directed layout and the red colored nodes/edges indicate new activities of the current month. The slider is set to June 2010 in Fig. 2 and the red edges and nodes represent the conversations added in that month. The gray items are the conversations that took place *before* June 2010. Fig. 3 (a) and (b) show how users can compare adjacent months using the time-slider and visualization. By switching (a) and (b) using the slider, they can easily discover how many new conversations (three) were added. Whereas (a) and (b) cannot give information about removed or aging conversations, Fig. 3 (b') distinguishes the current month's activities and the other older conversations by gradually decreasing the intensity as they age. We can see that the conversations added in the previous month (red in (a)) are painted in higher intensities while some far older nodes/edges are barely visible. This effect gives users information about which activities were more recent and which activities took place long before. Zhang et al. [15] used a similar approach to represent the age of the edges in an invitation network but they did not support the dynamic update of the edge intensities by navigation. A sigmoid function was used in order to implement the smooth drop of the color intensity by time.

Even though this network visualization provides users with information about the temporal change about the entire network, they may also be interested in changes in part of a graph. Therefore, we provided a marquee-selection tool that lets users select any part of the graph. For example, in Fig. 2, the user was interested in a small group of conversations (highlighted in blue by the user). S/he could marquee select them and then the information about the selected edges were dynamically overlaid on the histogram, showing that the conversations were done in March 2010 and the frequency was 9.
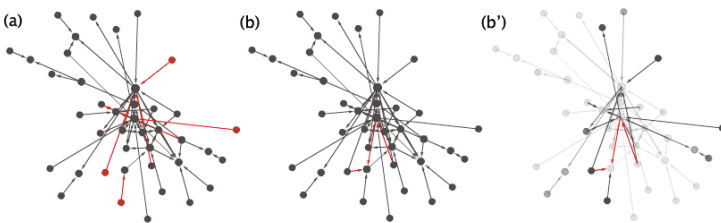
Fig. 3. Comparing February (a) and March 2007 (b and b'). Older activities can be painted in lower intensities (b') to show their age.

5   Discussions and Conclusions

This paper introduced our efforts to visualize the temporal dynamics of social networks. Five principles for implementing visualizations that support interactive exploration over time were proposed. We presented two prototypes according

to these principles and showed two datasets from the Nation of Neighbors social networking service: one utilizing the NodeXL extension to Microsoft Excel and TempoVis as a standalone application. The former approach is very flexible and easy to access because it is based on the well-known spreadsheet application and highly configurable. TempoVis was aimed to test the potential of a user interface element that could facilitate temporal navigation – a time-slider. We expect the NodeXL-based idea will be preferred by the users who are proficient with the data manipulation using spread sheets. TempoVis will be useful for the tasks where more exploratory and interactive temporal comparisons are required.

In order to evaluate our prototypes and understand their usefulness from a network manager's point of view, we consulted with Art Hanson, the NoN lead designer. He said: "The temporal visualizations would help focus limited resources on community groups who could best benefit from assistance, spot emerging trends, and identify the most influential users. The tools would also help NoN visualize the network effects resulting from changes to the NoN system and achieve the desired effect quickly with each new feature or design iteration." He could easily identify which group of NoN users showed different activities at a specific time point and use that information to manage and evaluate the community service. It was achievable by combining the time and the network information of the neighbors provided by our temporal visualization tools. We are planning more extensive user studies to validate the usefulness of our five principles and its implementations in the two visualization platforms.

The NoN team suggested to add the ability to distinguish node types (e.g. post and report) to the visualizations and to add the community group or location/distance information, so that they could link the temporal dynamics to community group activities or communications types. At the same time, we will also investigate how to locate sub-graphs and analyze their temporal dynamics. While fixing the locations of stable nodes/edges (principle 1) is useful, we are also planning to add a user option to dynamically redraw the graph layouts. The tasks such as observing large growth and change of clusters will require this feature. We are also considering to adopt multiple network views of user-selected time points in order to let users directly compare multiple time points. Enriching the visualization with additional attributes such as discussion topics will help differentiate various activities too. After more evaluation with community managers and network analysts we will embed the most valuable features into the NodeXL platform, while continuing to explore novel features in TempoVis.

Better understanding of network dynamics is necessary so that community managers can discover the triggers for growth, decay, active use, or malicious activities. Then they can develop strategies to promote growth and prevent decay and destructive interference. Social media have great potential for benefits in key domains such as community safety, disaster response, health/wellness, and energy sustainability. Better understanding of how to manage them effectively raises the potential for more successful outcomes.

References

1. Doreian, P., Kapuscinski, R., Krackhardt, D., Szczypula, J.: A brief history of balance through time. The Journal of Mathematical Sociology 21(1), 113–131 (1996)
2. Durant, K., McCray, A., Safran, C.: Modeling the temporal evolution of an online cancer forum. In: 1st ACM International Health Informatics Symposium (IHI 2010), pp. 356–365 (November 2010)
3. Freire, M., Plaisant, C., Shneiderman, B., Golbeck, J.: Manynets: an interface for multiple network analysis and visualization. In: Proceeding of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 213–222 (2010)
4. Giddens, A.: Sociology. Polity, Cambridge (2006)
5. Gloor, P., Laubacher, R., Zhao, Y., Dynes, S.: Temporal visualization and analysis of social networks. In: NAACSOS Conference, June, pp. 27–29 (2004)
6. Gloor, P., Zhao, Y.: Analyzing actors and their discussion topics by semantic social network analysis. In: Tenth International Conference on Information Visualization, IV 2006, pp. 130–135. IEEE, Los Alamitos (2006)
7. Gould, R.: The origins of status hierarchies: A formal theory and empirical test. American Journal of Sociology 107(5), 1143–1178 (2002)
8. Grinstein, G., Plaisant, C., Laskowski, S., O'connell, T., Scholtz, J., Whiting, M.: VAST 2008 Challenge: Introducing mini-challenges. In: IEEE Symposium on Visual Analytics Science and Technology, VAST 2008, pp. 195–196. IEEE, Los Alamitos (2008)
9. Hansen, D.L., Shneiderman, B., Smith, M.A.: Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, San Francisco (2010)
10. Kang, H., Getoor, L., Singh, L.: Visual analysis of dynamic group membership in temporal social networks. ACM SIGKDD Explorations Newsletter – Special Issue on Visual Analytics 9(2), 13–21 (2007)
11. Moody, J., McFarland, D., Bender-deMoll, S.: Dynamic network visualization. American Journal of Sociology 110(4), 1206–1241 (2005)
12. Perer, A., Shneiderman, B.: Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In: Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 265–274 (2008)
13. Powell, W., White, D., Koput, K., Owen-Smith, J.: Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. American Journal of Sociology 110(4), 1132–1205 (2005)
14. Trier, M.: Towards dynamic visualization for understanding evolution of digital communication networks. Information Systems Research 19(3), 335–350 (2008)
15. Zhang, J., Qu, Y., Hansen, D.: Modeling user acceptance of internal microblogging at work. In: CHI 2010 Workshop on Microblogging (2010)

Toward Predicting Popularity of Social Marketing Messages

Bei Yu[1], Miao Chen[1], and Linchi Kwok[2]

[1] School of Information Studies
[2] College of Human Ecology, Syracuse University
{byu,mchen14,lkwok}@syr.edu

Abstract. Popularity of social marketing messages indicates the effectiveness of the corresponding marketing strategies. This research aims to discover the characteristics of social marketing messages that contribute to different level of popularity. Using messages posted by a sample of restaurants on Facebook as a case study, we measured the message popularity by the number of "likes" voted by fans, and examined the relationship between the message popularity and two properties of the messages: (1) content, and (2) media type. Combining a number of text mining and statistics methods, we have discovered some interesting patterns correlated to "more popular" and "less popular" social marketing messages. This work lays foundation for building computational models to predict the popularity of social marketing messages in the future.

Keywords: text categorization, marketing, social media, prediction, media type.

1   Introduction

Social media has become a major influential factor in consumer behaviors, such as increasing awareness, sharing information, forming opinions and attitudes, purchasing, and evaluating post-purchasing experience [8]. Among those social media tools that are available for businesses, Facebook has become one of the most important media for B2C and C2C communications. Today many websites embed Facebook's "Liked" buttons, which not only allows consumers to exchange information within their Facebook network in a much faster speed but also pushes companies to pay more attention to Facebook fans' reactions to their messages sent through social media. Facebook fans' endorsement of a company's messages could be an important indicator of how effectiveness a company's social marketing strategies are.

   This research[1] aims to examine the effectiveness of different social marketing strategies. In social media such as Facebook, the long-term effectiveness of marketing strategies, such as the impact on consumer behavior, might not be directly measurable. However, Facebook provides some measures that indicate the message popularity, which is the level of attention that fans paid to the messages, such as the number of

"likes", the number of comments, and the overall tone of the comments. To some extent the popularity of a message indicates the effectiveness of the marketing strategy embedded in this message. Therefore in this research we aim to discover the characteristics of social marketing messages that affect their popularity. We focus on the number of "likes" they attract, and leave the comment analysis for future work.

Using the number of "likes" to measure the popularity of a message, we aim to answer the following research question: what characteristics of the messages contribute to their popularity? We examine the characteristics of messages from two perspectives: (1) content, and (2) media type. This article is organized as follows. Section 2 describes the data preparation process, including data sampling, downloading, and cleaning. Section 3 describes the thematic analysis of the message content. Section 4 describes an approach to predicting popularity based on the message content. Section 5 describes the correlation analysis between media type and message popularity. Section 6 concludes with limitations and future work.

2   Data Sampling, Downloading, and Cleaning

In this section we describe the data sampling, downloading, and cleaning process. We choose restaurant industry as a case study for the following reasons. Restaurant business has always been a big component of the tourism and hospitality industry, where intangible and perishable products are produced and consumed. Due to the experiential nature of these service products, consumers often seek information from online reviews and travel blogs before making purchasing decisions [10]. Restaurants understand the importance of communicating the "right" messages with their existing and potential customers over the Internet. Recent business reports also showed that restaurant entrepreneurs can substantially draw business and increase sales with Twitters and Facebook [11]. Many restaurants have been actively engaged with their customers on Facebook, which makes restaurant industry a great example for examining companies' social marketing messages.

This research adopted a criterion-based, mixed-purpose sampling procedure and selected twelve restaurants' Facebook pages in September 2010 based on the following criteria and steps: (1) we examined the top 20 restaurant chains with the highest sales volume [7] and selected all quick service restaurant chains with more than one million Facebook fans (eight in total) and the top three casual dining restaurant chains with the most Facebook fans; (2) we examined the top 20 independent restaurants and selected the top two with the most Facebook fans. We should have selected 13 restaurants in total, but eventually included 12 because KFC's Facebook page couldn't be correctly identified.

We then identified each restaurant's unique Facebook ID and retrieved all posts using Facebook FQL. We retrieved three fields: message body, message's media type ("status", "link", "video", or "photo"), and the number of people who have clicked the "likes" button on this message.

On October 5th we retrieved a total of 675 messages posted by the above 12 restaurants. See Table 1 for the number of posts collected for each restaurant. So far Facebook FQL only returns posts in recent 30 days or 50 posts, whichever is greater. We plan to continue our data collection on a regular basis for the next six months and re-examine our analysis on the larger data set in the future.

Because the number of fans for each restaurant varies significantly, messages from different restaurants with same number of "likes" could indicate significantly different level of popularity. For example, the most popular message from independent restaurant Carmines attracted only 93 likes, while the least popular message from Starbucks attracted 2,250 likes. To make the numbers of "likes" comparable among restaurants we normalized the numbers of "likes" for each restaurant using z-score. The normalized numbers of "likes" are distributed in the range (-3, 2).

Table 1. Descriptions of the 12 restaurants in this study

Restaurant Name	Facebook ID	No. of FB Fans	No. of FB posts
Quick Service (7)			
McDonald's	50245567013	3,244,673	39
Starbucks	22092443056	14,605,410	58
Subway	18708563279	2,218,422	58
Pizza Hut	6053772414	1,473,274	63
Taco Bell	18595834696	2,842,184	102
Dunkin's Donuts	6979393237	2,078,568	45
Chick-fil-A	21543405100	3,106,304	73
Casual Dining (3)			
Chilli's Grill & Bar	106027143213	470,994	51
Olive Garden	18618646804	544,278	41
Outback Steakhouse	48543634386	883,670	40
Independents (2)			
Joe's Stone Crab	404077308008	1,628	54
Carmine's	99057540626	5,190	51

3   Thematic Analysis

Identifying social marketing strategies is difficult without an established typology. However we may identify the common themes of the messages as indicators of restaurants' favorite marketing strategies. For example, if "donation" occurs frequently we can infer that charity is a commonly used marketing strategy. In order to identify the main themes we first applied Latent Dirichlet Allocation (LDA) algorithms [1] to all messages and identified ten main themes. However, the resulted themes were not as meaningful probably because LDA has proved to be more effective on larger data set. We then turned to basic keyword analysis to identify the most common content words (nouns, verbs, adjectives, and adverbs) that each restaurant used as a representation of its marketing strategies. Because of the important role of pronouns in communication [2], we also counted pronouns in this study.

During the keyword analysis process we (1) gathered all messages posted by each restaurant as an individual text collection; (2) tokenized the messages and computed each token's overall term frequency (how many times it occurs in this text collection); (3) manually picked the ten most frequent pronouns and content words that consist of the "keyword profile" of each restaurant. The resulted keyword profiles are presented in Table 2.

The keyword profiles in Table 2 depict a common communicative style signified by prevailing use of first and second person pronouns like "you", "your", "we", and "our". These profiles also uncover two common marketing strategies. The first is to promote key concepts as part of their public images. For example, McDonald's promoted "family time", Subway emphasized "fresh" food and being "fit", and Starbucks focused on the concept of "enjoy". The second common strategy is to run advertisement campaigns. Many of the keywords other than pronouns involve the major advertising campaigns, especially contests. Almost all restaurants launched some contests, for example "Super Delicious Ingredient Force" by TacoBell, "Ultimate Dream Giveaway" by Pizza Hut, "Create-A-Pepper" by Chili's, "Never Ending Pasta Bowl" by OliveGarden, "Free Outback For a Year" by Outback, "Eat Mor Chickin Cowz" by Chick-fil-A, "Fan Of The Week" by Dunkin's Donuts, etc.

Table 2. Keyword profiles for each restaurant

Restaurant	Keywords
Quick Service (7)	
McDonald's	you, your, games, all, we, time, Olympic, angus, more, howdoyoumcnugget
Starbucks	you, we, your, coffee, our, free, us, about, enjoy, more
Subway	your, you, favorite, fresh, footlong, today, chicken, fit, buffalo, breakfast
Pizza Hut	your, pizza, you, our, today, we, order, free, new, dream
Taco Bell	you, your, who, force, night, sdif, fourthmeal, super, chicken, delicious
Dunkin's Donuts	you, FOTW, our, coolatta, week, DD, next, your, bit, submit
Chick-fil-A	cowz, you, mor, your, cow, day, facebook, our, like, appreciation
Casual Dining (3)	
Chilli's Grill & Bar	we, you, fans, our, your, pepper, us, winner, St. Jude, gift
Olive Garden	you, our, your, pasta, new, we, ending, bowl, never, contest
Outback Steakhouse	bit, free, your, our, year, you, win, check, visit, chance
Independents (2)	
Joe's Stone Crab	special, week, our, dinner, crab, jazz, summer, we, you, sauce
Carmine's	your, our, we, family, memories, my, Washington DC, us, you, dinner

Did the key concepts resonate among fans? Did these contests really catch the eyes of fans? Or more broadly which strategies are more effective? By measuring the effectiveness with the popularity of relevant messages, we translated this question to the question of which messages are more popular than others, and then adopted an exploratory approach of using text categorization and feature ranking methods to discover the relationship between message content and popularity.

4   Content-Based Message Popularity Prediction

We grouped all messages into two categories: the "more popular" category for messages with positive normalized numbers of "likes" (above average), and the "less popular" category for messages with negative normalized numbers of "likes" (below average). After this process, the messages are separated into two categories: 430 messages in the "less popular" category and 207 messages in the "more popular" category. To build a sharp and balanced classifier we formed a new data set "400

messages" which consists of 200 most popular and 200 least popular messages. Each message is tokenized and converted to a word vector using bag-of-word representation. To focus on common words we excluded words that occurred in fewer than three messages and resulted in a vocabulary of 568 words. Reducing the vocabulary size also resulted in 19 empty vectors in the "400 messages" data set, which eventually has 381 (192 "more popular", and 189 "less popular") messages.

We then apply text categorization algorithms to examine the extent to which the "more popular" and "less popular" categories are separable. We employed two widely used text categorization algorithms Support Vector Machines (SVMs) and naïve Bayes (NB) to train and evaluate the content-based popularity prediction models.

SVMs are among the best methods for text categorization [6] and feature selection [4]. SVMs favor discriminative features with broad document coverage, and therefore reduce the risk of over-fitting. Studies have shown that the 10% most discriminative features identified by SVM classifiers usually work as well as, or even better than the entire original feature set [12]. In this study we used the SVM-light package [5] with default parameter setting. We compared the performance of four variations of SVM models: SVM-bool with word presence or absence (one or zero) as feature values; SVM-tf with raw word frequencies; SVM-ntf with word frequencies normalized by document length; and SVM-tfidf with word frequencies normalized by document frequencies.

Naïve Bayes (NB) is another commonly used text categorization method, simple but effective even though the text data often violate its assumption that feature values are conditionally independent given the target value [3]. We compared two variations of naïve Bayes implementations: NB-bool, the multi-variate Bernoulli model that uses word presence and absence as feature values, and NB-tf, the multinomial model that uses raw word frequencies. We implemented the two naïve Bayes variations.

We used leave-one-out cross validation to evaluate the SVMs and NB classifiers on both "all messages" and "400 messages" data sets. We used the majority vote as the baseline for comparison. The results in Table 3 show that all classifiers beat the majority vote on the "400 messages" data set by wide margins, and that SVM-ntf and SVM-bool both achieved accuracy over 75%. In contrast the cross validation accuracies on the "all messages" data set are less encouraging in that they are merely comparable to the majority vote. The good result on "400 messages" demonstrates that the "more popular" and "less popular" messages are fairly separable based on message content. The poor result on "all messages" also supports this notion in that the messages with medium level of popularity proved to be difficult to categorize.

Table 3. Leave-one-out cross validation results (in percentage)

Algorithm	All messages	400 messages
Majority vote	67.5	50.4
Svm-bool	67.1	*75.1*
Svm-tf	67.8	74.3
Svm-ntf	72.2	*75.9*
Svm-tfidf	71.9	71.7
Nb-bool	68.7	71.1
Nb-tf	68.6	71.7

Because the SVM-bool classifier is easiest to explain, we used the SVM-bool classifier to rank the word presence or absence features based on their discriminative power toward separating "more popular" and "less popular" messages on "400 messages" in the SVM-bool classifier. Table 4 lists the words that indicate the "more popular" and "less popular" messages. Surprisingly, "contest", "win", "winner", and "free" all appear as top indicators of "less popular" messages. It means that fans are not that interested in winning the contests, probably because of the low chance to win. On the other hand, after checking context of the top indicators of "more popular" messages, we found that fans are more interested in trying new dishes, for example the word "try" occurred in 15 messages, 14 of which are in the "more popular" category.

Table 4. The most indicative words ranked by the SVM-bool classifier

Category	Top indicative words
more popular	try, have, coffee, deal, come, days, october, one, then, like, order, will, who, crunchy, with, million, very, pumpkin, flavors, new, spice, fourthmeal, but, donate, join, early, medium, fall, favorites, served, support
Less popular	the, week, about, http, win, their, I, watch, from, com, last, free, starbucks, first, can, contest, by, yet, memories, tell, amp, visit, live, almost, family, on, fun, our, meal, photo, video

5   Media Type Analysis

Social marketing messages use not only words but also other information types such as links, videos, or photos. In this section we report a statistical analysis to examine the impact of the message types on their popularity. Most of the messages we collected contain a metadata entry called "Media Type" which is either "status", "link", "video", or "photo". A "status" message has only textual descriptions, usually about the business's updates, status, information, etc. A "link", "video", or "photo" message contains a URL, video, or photo respectively. They may or may not contain text. We conjecture that a restaurant chose a media type when posting a message to Facebook in that Facebook is not likely to manually or automatically assign this metadata for each message. We retrieved the media types from all messages and described their basic statistics in Table 5.

We then performed ANOVA test to examine whether there is significant difference between the normalized numbers of "likes" for the four media groups (N=670, five messages did not include media type). The homogeneity test shows that the significance is $p < .001$ in Levene's test, indicating that the equal variance hypothesis should be rejected. Therefore we used Games-Howell for post-hoc testing.

The ANOVA test result in Table 6 shows significant difference between the four media groups in terms of the normalized number of "likes" they attracted from fans, $F(3,666)=9.499$, $p < .001$. Because the data violate homogeneity assumption, we used Welch and Brown-Forsythe to further examine the significance of the result. The result is still significant ($p < .001$): $F(3,188.16)$, $p < .001$ and $F(3,561.42)$, $p < .001$ for Welch and Brown-Forsythe tests respectively. On average "status" and "photo" messages received significantly more "likes" than "link" and "video" messages. A possible reason is that many fans chose not to spend the extra effort the click the links

or the videos. "Status" messages are even more popular than "photo" messages, although there is no significant difference between them. There is no significant difference between "link" and "video" types either.

Table 5. Basic statistics of media types

Statistics	Link	Status	Video	Photo
#msg	227	190	38	215
Mean	-.22	.19	-.40	.12
SD	.76	1.15	.51	1.06

Table 6. ANOVA test of media type difference

Media type (I)	Media type (J)	Mean difference (I-J)	Sig.
Link	Status	-.41***	.000
Link	Video	.18	.279
Link	Photo	-.34**	.001
Status	Video	.59***	.000
Status	Photo	.074	.908
Video	Photo	-.51***	.000

Note: *p<.05, **p<.01, ***p<.001

6   Conclusions and Limitations

This research examined the characteristics of social marketing messages that affect their popularity. Because popularity of these messages indicates the effectiveness of the corresponding marketing strategies, this research is expected to lay the foundation in order to build computational models to predict the effectiveness of social marketing strategies based on the popularity-related factors discovered in this study.

Using messages posted by a sample of restaurants on Facebook as a case study, we investigated the relationship between the message popularity (measured by the number of "likes") and the message content and media type. By combining a number of text mining and statistics methods, we found that (1) restaurants share some common marketing strategies, for example they attempt to promote their unique public images, introduce new dishes, as well as run advertisement campaigns like contests; (2) fairly effective prediction models can be built to predict if a message is more or less popular based on the message's content; (3) some marketing strategies, like promoting new dishes, are more effective while others, such as running contests, are not, probably because of the low chance to win these contests; (4) media type also affect the message popularity: "status" and "photo" messages are more popular than the "link" and "video" ones, probably because of the extra effort to click or play.

This research has a few limitations that we will improve in the future. First, we have not yet controlled the effect of time on the message popularity. Newly-posted messages were penalized to some extent because of their short exposure time. Second, the number of "likes" captures positive feedback only because no "dislikes" button is available. However fans can leave free-text comments which should be able to

disclose both positive and negative sentiment, and thus provide more information regarding the effectiveness of social marketing strategies. We are currently analyzing the sentiment of the messages and their comments. Third, the data set is relatively small because Facebook API returns posts of the past 30 days. However, all research methods we employed in this study can be scaled to larger data set. We will continue collecting the posts and comments regularly for the next six months and re-examine if our conclusions still hold on the larger data set. We will also explore relevant approaches to enhance the performance of message popularity prediction [5, 9].

References

1. Blei, D., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. J. Machine Learning Research 3, 993–1022 (2003)
2. Chung, C.K., Pennebaker, J.W.: The Psychological Function of Function Words. In: Fiedler, K. (ed.) Social Communication: Frontiers of Social Psychology, pp. 343–359 (2007)
3. Domingos, P., Pazzani, M.: On the Optimality of the Simple Bayesian Classifier Under Zero-one Loss. Machine Learning 29, 103–130 (1997)
4. Forman, G.: An Extensive Empirical Study of Feature Selection Metrics for Text Categorization. Journal of Machine Learning Research 3, 1289–1305 (2003)
5. Hong, L., Davison, B.D.: Empirical Study of Topic Modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics, Washington D.C, July 25 (2010)
6. Joachims, T.: Text categorization with Support Vector Machines: Learning with Many Relevant Features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998)
7. Killian, K.S.: Top 400 Restaurant Chains. Restaurant & Institutions, 28–72 (2009)
8. Mangold, W.G., Faulds, D.J.: Social Media: The New Hybrid Element of the Promotion Mix. Business Horizons 52(4), 357–365 (2009)
9. Szabo, G., Huberman, B.A.: Predicting the Popularity of Online Content. Communications of the ACM 53(8), 80–88 (2010)
10. Xiang, Z., Gretzel, U.: Role of Social Media in Online Travel Information Search. Tourism Management 31(2), 179–188 (2010)
11. Young, L.: Brave New World. Foodservice and Hospitality 42(11), 24–28 (2010)
12. Yu, B.: An evaluation of text classification methods for literary study. Journal of Literary and Linguistic Computing 23(3), 327–343 (2008)

Simulating the Effect of Peacekeeping Operations 2010–2035

Håvard Hegre[1,3], Lisa Hultman[2], and Håvard Mokleiv Nygård[1,3]

[1] University of Oslo
[2] Swedish National Defence College
[3] Centre for the Study of Civil War, PRIO

Abstract. We simulate how a set of different UN policies for peace-keeping operations is likely to affect the global incidence of internal armed conflict. The simulation is based on a statistical model that estimates the efficacy of UN peacekeeping operations (PKOs) in preventing the onset, escalation, continuation, and recurrence of internal armed conflict. The model takes into account a set of important conflict predictors for which we have projections up to 2035 from the UN and the IIASA. The estimates are based on a 1970–2008 cross-sectional dataset of changes between conflict levels and new data on PKO budgets and mandates. The simulations show a strong effect of PKOs on the global incidence of major conflicts, although restricted to operations with robust mandates. Extensive use of 'transformational' PKOs can reduce the global incidence of the most lethal conflicts with 65%.

Keywords: Armed conflict, Simulations, UN Peacekeeping operations, Conflict prevention, Statistical modeling, Demographic projections.

1 Introduction

Peacekeeping has become a common tool for resolving conflicts and establishing conditions for a stable peace in war-torn countries. In this paper, we use simulations based on a statistical model to evaluate the potential of peacekeeping operations (henceforth PKOs) for reducing future conflict. The model specification builds on research on peacekeeping as well as on other trends and factors that influence the risk of conflict. The simulations evaluate different scenarios of peacekeeping in the period 2010–2035, varying budgets and mandates.

Existing research on peacekeeping shows that PKOs are important for building peace, but has been unable to assess the total substantive impact on the risk of conflict, taking into account long-term effects and regional diffusion. By using simulations, we are able to incorporate such indirect effects of PKOs. We specify eight PKO scenarios based on previous research, own statistical estimations of relevant factors, and reports by UN sources about the likely future of peacekeeping. One clear trend in recent years is a larger PKO budget and stronger mandates. The scenarios reflect different potential policies specifying PKO expenditures and mandates, in which countries PKOs are deployed, and when. The

findings imply that the more the UN is willing to spend on peacekeeping, and the stronger the mandates provided, the greater is the conflict-reducing effect.

2 Methodology

2.1 Data

The dependent variable in this study is a three-category variable denoting whether there is a minor conflict, a major conflict, or no conflict going on in a country in a given year. The conflict data are from the 2009 update of the UCDP/PRIO Armed Conflict Dataset (ACD) [3,2]. The ACD records conflicts at two levels, measured annually. Minor conflicts are those that pass the 25 battle-related deaths threshold but have less than 1000 deaths in a year. Major conflicts are those conflicts that pass the 1000 deaths threshold. We only look at internal armed conflicts, and only include the countries whose governments are included in the primary conflict dyad (i.e., we exclude other countries that intervene in the internal conflict).

We use data on PKOs from three different sources: Doyle and Sambanis' coding of two types of PKOs [1]: **Traditional missions** comprise missions that are restricted to observing truces, troop withdrawals, or buffer zones, policing e.g. buffer zones, or assisting in negotiations. Traditional missions are always deployed with the consent of the parties to the conflict. **Transformational missions** include consent-based 'multidimensional missions' where mandates are extended with activities intended to go to the roots of the conflict, such as economic reconstruction or reform of the police, and 'enforcement missions' that do not require the consent of both parties, drawing on the authority of UN Charter articles 25, 42, and 43 to apply force to protect the activities of the operation. In order to capture the size of the PKO, we have coded the yearly expenditure for each mission, based on UN General Assembly's published *appropriation* resolutions from 1946 to the present. For PKO years without expenditure data we use the average for the mission type as our best guess.[1]

We include predictor variables that are associated with the risk of conflict and for which we have good projections for the 2010–2035 period: the proportion of males aged 20–24 years with secondary or higher education of all males aged 20–24 [7,8]; Log infant mortality, log total population, the percentage of the population aged 15–24 years of all adults aged 15 years and above [9]. We also include a set of conflict variables: information on conflict status at $t-1$, the log of the number of years in each of these states up to $t-2$, and conflicts in neighboring countries. Projections for the conflict variables are produced by the simulation procedure itself.

---

[1] A small number of missions are funded directly through the UN's operating budget, and yearly expenditure data are harder to single out from other budget items. For further details on the PKO data, see [4].

2.2   Simulation Procedure

To evaluate the efficacy of peace-keeping operations we estimate the statistical relationship between the incidence of conflict and the presence of PKOs of various types and budget sizes, controlling for other factors that have been shown to affect the risk of conflict.[2] Assessing the predictive performance of variables is extremely important in order to evaluate policy prescriptions [10]. Our statistical model specification necessitates generating predictions through simulation for other reasons. The unit of analysis in our study is the country year, and the models are estimated on data for all countries for the 1970–2008 period. Since PKOs may have effects that extend beyond that year and that country, simulation for all countries over longer periods allows quantifying the effect of UN policies over a long period (2010–2035) and tracing the effect on neighboring countries. The methodology is described in detail in [5].

The general setup of the simulation procedure is shown in Fig. 1. It involves the following steps: (1) specify and estimate the underlying multinomial logistic regression model on data for 1970–2008; (2) make assumptions about the distribution of values for all exogenous predictor variables for 2009 and projected changes after that; (3) start simulation in 2009 for the actual forecasts; (4) draw a realization of the coefficients of the multinomial logit model based on the estimated coefficients and the variance-covariance matrix for the estimates; (5) calculate the 9 probabilities of transition between levels for all countries for the first year, based on the realized coefficients and the projected values for the predictor variables; (6) randomly draw whether a country experiences conflict, based on the estimated probabilities; (7) update the values for the explanatory variables, contingent upon the outcome of step 6; (8) repeat (4)–(7) for each year in the forecast period, e.g. for 2009–2035, and record the simulated outcome; (9) repeat (3)–(8) 1,000 times to even out the impact of individual realizations of the multinomial logit coefficients and individual realizations of the probability distributions.

The simulation methodology is the most accurate hitherto developed in conflict research. [5] shows that the model specification yields an AUC of 0.937 – with $p > 0.5$ as the cutoff, it correctly predicts about 63% of conflicts (minor or major) 7–9 years after the last year of data, with about 3.5% false positives.[3]

2.3   Specifying PKO Scenarios

We specify eight different scenarios to explore the effect on the global incidence of conflict of various UN policies. The following rules guide most of them: (i)PKOs are assumed to be initiated if the conflict is major (more than 1,000 deaths in

---

[2] For a review of conflict risk variables, see [6]. We carry out a systematic evaluation of the predictive performance of the control variables in [5].

[3] To assess the predictive performance, [5] estimates the relationship between predictors and risk of conflict based on data for 1970–2000, simulates up to 2009 and compares simulation results for 2007–2009 with the most recent conflict data available for the same years [3].
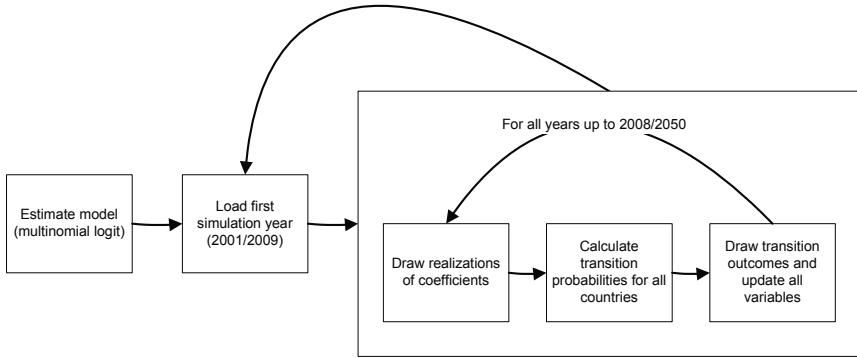
Fig. 1. Simulation Flow Chart

the previous year) and the conflict has lasted for at least two years; (ii) PKOs remain for five years after last year with conflict activity (more than 25 annual deaths); (iii) PKOs are never deployed in permanent UNSC members; (iv) except in scenarios 4, 7, and 8, PKOs are deployed only in countries that have smaller populations than 100 millions in 2009.

The first scenario (S1) is a comparison scenario where the UN terminates all PKO activity in 2010. We specify three scenarios in which the UN chooses to spend different amounts on each mission, ignoring the mandates – 100 million USD per PKO per year (S2), 800 million (S3), and 800 million plus deployment in large countries (S4). The final four scenarios vary the mandates of the PKOs, ignoring the budget of the mission: traditional mandates (S5); transformational mandates (S6): transformational mandates, also in large countries (S7); as S7, but deployment in first year of conflict (S8).

3   Results

3.1   Estimation Results

We estimated a multinomial logistic regression model with our conflict variable as the dependent variable and PKO and the other variables as predictors. Table 1 shows the results for the PKO variables from the statistical estimation.[4] All simulations shown below are based on these estimates.

3.2   Prediction Results

Figure 2 compares the simulated proportion of countries in conflict in the comparison scenario with the proportion in the other scenarios. Socio-economic development variables are important predictors of conflict, and our UN/IIASA forecasts expect positive changes for most countries over the next 25 years.

---

[4] The other estimates are reported in [4].

Table 1. Estimation Results, Determinants of Minor/Major conflict

PKO variable	Minor	Major	Minor	Major
Log expenditures	-0.00792 (-0.18)	-0.259** (-3.11)		
Traditional			-0.0757 (-0.28)	-0.462 (-1.14)
Transformational			-0.0934 (-0.29)	-2.816** (-2.68)
N	5942		5942	
log likelihood	-1518.4		-1516.2	

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimated coefficients for other predictors not shown.
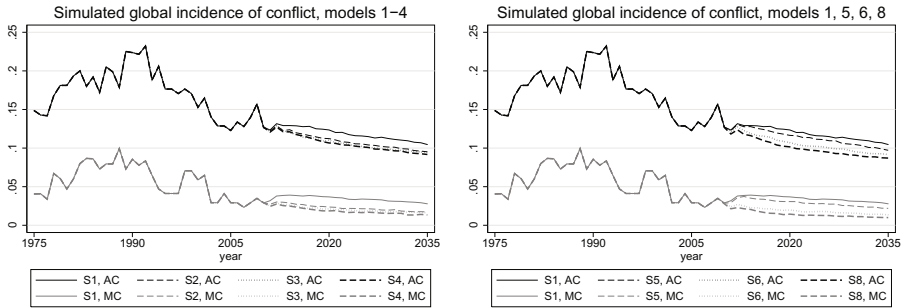
Left: Budget scenarios (S1–S4). Right: Mandate scenarios (S1, S5, S6, S8).

Fig. 2. Simulated Incidence of Conflict 2010–2035, Globally, Both Conflict Levels

Hence, we predict a moderate decline in the proportion of countries in conflict even without any PKOs, and an incidence of major conflict at the same level as in 2004–08.[5] In the left panel, we compare the baseline scenario with scenarios S2, S3, S4 varying the budget of missions (ignoring their mandates). The set of black lines represent the incidence of all armed conflicts, the gray set major conflicts only. All three scenarios imply a reduction in the incidence of conflict. The conflict reduction is substantial for major conflicts. In 2035, the predicted incidence of conflict for the most extensive scenario is less than half that of the comparison scenario. The reduction in the incidence of minor conflict, however, is negligible. There is little difference between the predictions for the various budget levels.

The right panel shows the predicted incidence of conflict varying the mandates of the PKOs. As expected from the estimates in Table 1, the effect of varying mandates is stronger than that of varying budget. The predicted reduction in the incidence of both levels of conflict is about 2%, and the reduction in the incidence of major conflict is about the same. This means that the most extensive scenario reduces the risk of major conflict in 2035 with about two thirds.

The simulated indirect effect of PKOs on minor conflict is weak. However, it is not so that PKOs merely reduce the *intensity* of conflict. If that was the

---

[5] [5] discusses the uncertainty of the simulated results.
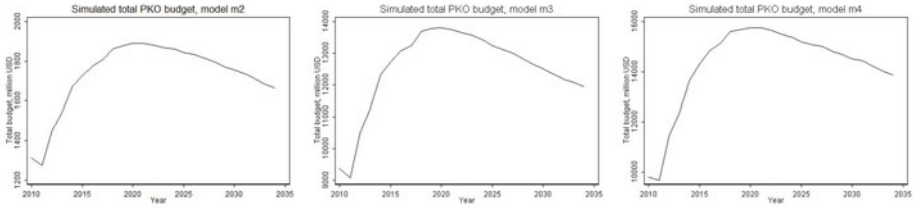
Fig. 3. Simulated Total UN PKO Budgets, 2010–2035

case, the incidence of both levels of conflict would not be reduced in any of the scenarios. Our simulations imply that for every successful transition from major to minor conflict due to the presence of a PKO, there is one transition from minor conflict to no conflict.

Our model allows capturing long-term and spatial effects of conflict. The estimates for 'log time in status' variables (not shown in Table 1) indicate that the probability of no conflict increases strongly with several consecutive years of peace, and decreases with several consecutive years of conflict. Likewise, conflicts in neighboring countries increase the risk that conflicts erupt. Given that Table 1 shows that PKOs have a clear short-term effect, we expect the difference between scenarios to increase over time. There are indications that this is the case. The simulated difference between the the no-PKO and PKO scenarios in Fig. 2 clearly widens from the first year of simulation up to 2020. After the initial 10 years, the lines are roughly parallel. The lines become parallel considerably earlier for the incidence of major conflict.

How expensive are these scenarios? Figure 3 shows the average simulated total budgets for UN PKOs under scenarios S2, S3, and S4. The low-budget scenario (S2) would imply a strong reduction in UN peace-keeping expenditures, whereas the other two scenarios represent an increase in total annual expenditures of 50–70% compared to what the UN spent in 2009. The simulations show that the budget would only increase for approximately ten years, and then start decreasing as a consequence of the reduction in the incidence of conflict.

3.3   Regional Effects

We define 8 regions. In Fig. 4, we show the simulated incidence of conflict in six of these regions for scenarios S1, S5, S6, and S8. We also plot the observed proportion of countries in conflict for the 1995–2008 period for each region. Since there are only a handful of conflicts within each region, individual conflicts are discernible in the plots. In the 1995–2004 period in West Africa, for instance, the total number of conflicts fluctuated between 2 and 4. In the 2010–25 projections for this region, 10% of the countries are in conflict. This corresponds to about 2 conflicts every year. The expected number of major conflicts (more than 1,000 battle-related deaths) is less than 0.5 for this region.

PKOs have the strongest effects in regions 4, 6, and 7. The other three regions have had few major conflicts in the post-Cold war period, and the model predicts
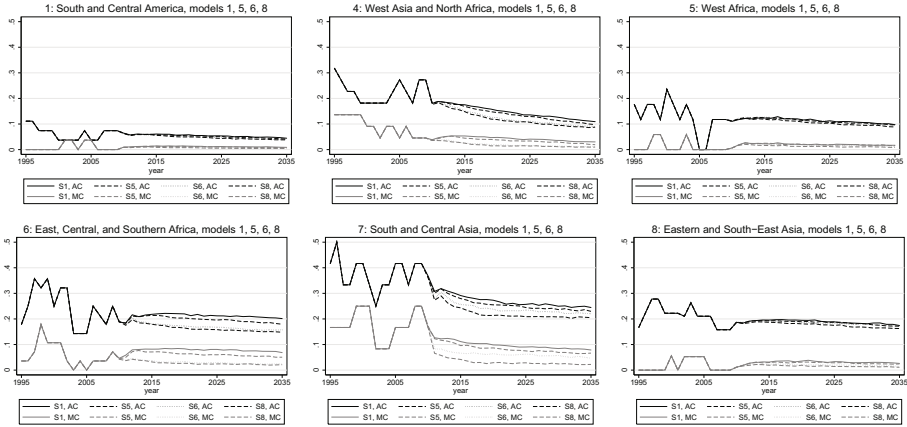
Fig. 4. Simulation 2010–2035, Both Conflict Levels, Six Regions

a continued low incidence of these conflicts. Since PKOs in our scenarios are initiated only in major conflicts, we predict fewer deployments in these regions, and they therefore only marginally affect the regional incidence of conflict.

In the other three regions, however, PKOs substantially reduce the number of conflicts. In 'West Asia and North Africa', we see a strong decline in the incidence of conflict because of the relatively high levels of socio-economic development in the region. Particularly in the first 15 years of the simulation, PKOs with strong mandates reinforce this declining trend. Since there are few large countries in the region, there is little difference between S6 and S8.

The 'East, Central, and Southern Africa' and 'South and Central Asia' regions are the ones with the highest incidence of conflict after the Cold War (and in our simulations). In region 6, scenario S6 (with enforcement mandates for all conflicts in smaller countries) reduces the predicted incidence of major conflict from about 8% of the countries to about 3% – from more than two conflicts every year to less than one. In region 7, the predicted incidence of conflict is about 11-12% in 2015 and slowly decreasing under the comparison scenario. In scenario S6 the predicted incidence decreases to about 7%. Scenario S8, which allows for PKOs in large countries (but not in the permanent members of the UNSC) and deployment in the first year of a major conflict reduces the incidence by another couple of percents.

4   Conclusion

We have evaluated the prospects of PKOs in reducing conflict in the future by simulating different scenarios for UN PKO deployment policies. The results show that PKO has a clear conflict-reducing effect. The effect of PKOs is largely limited to preventing major armed conflicts. PKOs have the strongest effect in the traditionally conflict-prone regions West Asia and North Africa, East, Central, and Southern Africa, and South and Central Asia.

In one of the most extensive scenarios, in which major armed conflicts are met with a PKO with an annual budget of 800 million USD, the global incidence of major armed conflict is reduced by half relative to a scenario without any PKOs. The total simulated UN peacekeeping budget increase by 50–70 percent relative to 2009 levels. Moreover, the total PKO budget would increase for about ten years, and then start decreasing again as a result of a reduced number of conflicts in the world. In another scenario, which specifies that major conflicts get a PKO with a robust, 'transformational' mandate from the first year, the risk of conflict is reduced by two thirds in 2035 compared to a scenario without any PKO.

Previous research has shown that peacekeepers are fairly good at keeping peace but less able at ending violent conflict. Our findings suggest that PKO may also be an effective tool for ending conflict if accompanied with a strong mandate and a sufficient budget. This implies that theories focusing on the credibility of the commitment by the UN are probably the most useful for understanding how PKOs can help warring parties lay down their arms.

References

1. Doyle, M.W., Sambanis, N.: Making War & Building Peace: United Nations Peace Operations. Princeton University Press, Princeton (2006)
2. Gleditsch, N.P., Wallensteen, P., Eriksson, M., Sollenberg, M., Strand, H.: Armed conflict 1946–2001: A new dataset. Journal of Peace Research 39(5), 615–637 (2002)
3. Harbom, L., Wallensteen, P.: Armed conflicts, 1946–2009. Journal of Peace Research 47(4), 501–509 (2010)
4. Hegre, H., Hultman, L., Nygård, H.: Evaluating the conflict-reducing effect of un peace-keeping operations. Paper presented to the National Conference on Peace and Conflict Research, Uppsala, September 9-11 (2010)
5. Hegre, H., Karlsen, J., Nygård, H.M., Strand, H., Urdal, H.: Predicting armed conflict 2010–2050. Typescript, University of Oslo
6. Hegre, H., Sambanis, N.: Sensitivity analysis of empirical results on civil war onset. Journal of Conflict Resolution 50(4), 508–535 (2006)
7. Lutz, W., Goujon, A., Samir, K.C., Sanderson, W.: Reconstruction of Population by Age, Sex and Level of Educational Attainment for 120 Countries 1970–2000. IIASA, Laxenburg (2007)
8. Samir, K.C., Barakat, B., Skirbekk, V., Lutz, W.: Projection of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 2005–2050. IIASA IR-08-xx. IIASA, Laxenburg (2008)
9. United Nations. World Population Prospects. The 2006 Revision. UN, New York (2007)
10. Ward, M.D., Greenhill, B.D., Bakke, K.M.: The perils of policy by p-value: Predicting civil conflicts. Journal of Peace Research 47(4), 363–375 (2010)

Rebellion on Sugarscape: Case Studies for Greed and Grievance Theory of Civil Conflicts Using Agent-Based Models

Rong Pan

Arizona State University, Tempe, Arizona, United States
`rong.pan@asu.edu`

Abstract. Public policy making has direct and indirect impacts on social behaviors. However, using system dynamics model alone to assess these impacts fails to consider the interactions among social elements, thus may produce doubtful conclusions. In this study, we examine the political science theory of greed and grievance in modeling civil conflicts. An agent-based model is built based on an existing rebellion model in Netlogo. The modifications and improvements in our model are elaborated. Several case studies are used to demonstrate the use of our model for investigating emergent phenomena and implications of governmental policies.

Keywords: agent-based model, civil conflicts, sensitivity analysis.

1 Introduction

In this research we are interested in modeling civil conflicts using the agent-based computational approach, so as to study the behavior of the populace in a specific national environment and to investigate the impact of governmental policies. Based on the greed and grievance theory, a person rebels because 1) he has enormous grievance towards the central government and 2) he sees the opportunities (e.g., economical benefits, political advantages, etc.) of becoming a rebel. Therefore, we may say that the occurrence of civil conflict is the result of the interaction of populace within its environment, including the reactions of individuals to their life conditions and their interactions with neighbors, with the general national status (weak or strong state) and the authoritarian power of government, etc. Agent-based models (ABMs) are the natural choice of modeling such complex systems.

The NOEM model, developed by the Air Force Rome Laboratory, is a large-scale computer simulation model built on system dynamics principles for simulating nation-state operational environment. Through it we may study the causal relationship between exogenous factors and nation-state's social and economical status, as well as the impact of various policy options. We postulate that the NOEM will provide a valid nation-state environment for ABM simulation. For example, the indicators for the economic conditions of a nation, such as GDP, GINI, can be imported from the NOEM and they are used in ABMs to create a nation's economy map. The agents in ABMs will possess certain properties that link them to their living environment. They

also follow certain behavior rules and interact with the environment. The NOEM is a system dynamics (SD) model, where macroscopic difference equations are used to calculate the change of nation-state variables. The ABM technique takes a different approach. It models at the microscopic level, modeling individual agent's (it can be a person) behavior, and how it interacts with its neighbors and its environment.

2   Theory of Greed and Grievance in Civil Conflicts

In the seminal paper by Collier and Hoeffler (2002), the outbreak of civil war was explained from two possible views: severe grievance and atypical opportunity (greed) for building a rebel organization. The grievance model is in line with the motivation cause of conflict explained in the political science literature. Grievance is fueled by the polarization of a society, which is caused by poor governance. The opportunity model explains the outbreak of civil war from an economic feasibility point of view. An interesting hypothesis proposed in Collier, Hoeffler and Rohner (2008) states that where civil war is feasible it will occur without reference to motivation. This "feasibility hypothesis" is substantiated by analyzing past civil war data and finding that the variables, which are close proxies for feasibility, have powerful consequences to the risk of a civil war.

Collier and Hoeffler's theory has stirred a heated debate on the true causes of conflicts among political scientists, sociologists and economists. Although most empirical studies conducted at the World Bank support the greed model (see, Ganesan and Vines, 2004, Bodea and Elbadawi, 2007), many researchers rebutted the notion that the onset of conflict is purely determined by economic reasons or the motivation of rebellion is driven by seeking financial opportunity only. Some argued that the reductionist categories of greed and grievance not only obfuscate other social, ethnic and religious factors, but also mislead the intervention policy of international community (Kuran, 1989, Gurr and Moore, 1997, Gurr, 2000, Ballentine and Nitzschke, 2003, Regan and Norton, 2003). This is specifically demonstrated by studying conflicts in some African countries (Agbonifo, 2004, and Abdullah, 2006). It seems that this debate will continue. In our study, we do not intend to validate this theory, but simply apply it on an artificial society and to observe the interactions of greed and grievance factors and the social consequence of governmental policy.

3   Agent-Based Models

An agent-based model (ABM) is a simulation environment where agents (e.g., simulated people) are defined to possess certain properties and behaviors, and they can interact with each other and with their surrounding environment. This type of model is different from other mathematical models, such as system dynamics, in several ways. First, the ABM is built in a bottom-up fashion. It defines the microscopic details of basic elements in a society and lets the society grow (or evolve) through agent-to-agent and agent-to-environment interactions. Second, the use of ABMs is typically to provide the macroscopic picture of how a society grows and look for emergent phenomena. This is achieved by examining the statistics of

collections of agents over the simulation time. The connections between micro-level details and macro-level phenomena can be used to explain cause-and-effects for the artificial society under study.

Naturally, ABMs have been applied in many social science studies. See Epstein (1990) for supporting arguments of this computational approach to social science. Epstein and Axtell (1996) gave a full-fledged description of how to build an artificial society and examine its emergent phenomena on a sugarscape (an imagined land which provides sugars for the agents living on it). For the civil conflict study, the most famous model is the rebellion model, which implements the greed and grievance theory on modeling agent's behavior. It has been used to demonstrate certain conflict phenomena found in the real world and the effects of changing governmental policy (Epstein et al., 2001, Epstein, 2002, and Cederman and Girardin, 2007). More recently, a group in Netherlands used ABMs to simulate the war game scenario of foreign-force evasion on an island undergoing a civil war (van Lieburg et al. 2009, Borgers et al. 2009).

In our study, we built an artificial land based on the sugarscape model and let rebels grow on this land and interact with the environment; thus, we call it "rebellion on sugarscape". The sugarscape can be viewed as a resource map which is geospatially bounded. The behavior of an agent depends on the agent's living environment (whether there are enough resources to support its living and whether there are rebels/cops in its neighborhood) and the policy set imposed on this artificial land. Our intention is to use this example to demonstrate the feasibility of the marriage of civil conflict theory and a resource map, which has not been done in the original rebellion model. In addition, we studied the behavior of rebellion and the consequences of policy setting when the resource is pre-defined.

3.1   Rebellion on Sugarscape Model

Although the rebellion model has been successfully used to explain some phenomena of civil conflicts, it lacks several aspects that we are interested in. First, the grievance level of each person is randomly given. In modeling a nation, we would like to link the grievance level to the person's economic, social, and political status, or in other words, its quality of life. Second, the number of cops is fixed. In reality, a government needs to spend enormous wealth on maintaining its military, police and other civil governing means. Third, in the current model, all agents can freely, randomly move. In reality, there may be many restrictions on a person's movement for economic and political reasons.

To address these problems, we created a refined model, called "rebellion on sugarscape". The sugarscape is a layer of resource overlaid on the artificial land. Sugars (or resources) constantly grow on this land and can be collected by each agent, and by the government through taxation, to become personal wealth or the government's wealth. However, the distributions of sugars are uneven. There are areas the sugar level is highest (4) and areas there is no sugar (0). We may interpret the sugar-rich region as the resource-rich city area and the sugar-poor region as the resource-depleted rural area, or the sugar-rich region as the region with oil fields and the sugar-poor region as the region without oil. Initially, agents are randomly scattered on the land. An agent's grievance will be linked to the location it resides. If

it is in a sugar-rich region, its perceived hardship will be low, and vice versa. The general population can be defined either mobile or non-mobile. If it is mobile, a person will move to a place having at least a sugar level of 1 within its vision. This is based on the assumption that people want to seek a better life for themselves. Every person will collect wealth from the place it resides. The wealth it can collect is the same as the sugar level at its location. Part of the wealth will be taxed, and the government will use the tax revenue to either maintain the existing cops, or to create new cops, or to assist the people who are in poverty. In our model, a cop needs at least a value of 5-sugar level to survive. To create a new cop, the government needs to spend 10 sugars. The poverty line is set at 1-sugar level. If a person has less than 1 sugar, it may be assisted by the government. The assisted amount is the total sugar that the government allocated to the assistance divided by the total number of people in poverty. Our model is depicted in Figure 1.
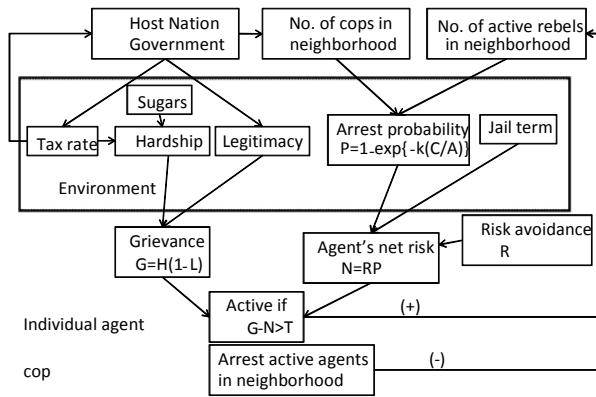

Fig. 1. The rebellion on sugarscape model

4   Case Studies

In this section, we present several case studies of the rebellion on sugarscape model. Our purposes are two-folded: 1) to validate the model against basic, intuitive scenarios; 2) to look for "surprises" that will lead to a better understanding of the model and of the consequences of policy setting.

4.1   Case Study #1: A Baseline Case

First we test a baseline case, where the number of agents (general population) is 1750 and the initial number of cops is 100. We prohibit the movement of agents (presenting a non-mobile society), and set both the wealth distributed to the poor and the wealth for creating new cops to be 0. Therefore, the government will use its wealth to maintain the initial cop size throughout the simulation. The tax rate is set at 30%.

This baseline case is a simple recast of the rebellion model using a layer that defines the resources and links the people's grievance to their wealth. Therefore, we should expect a similar behavior of rebels as in the previous rebellion simulation, except that the rebels will be most likely be in poor regions. We ran the simulation over 200 cycles. We observed that rebellions always happen in the places with less resource and the outburst of mass rebellion happens from time to time, and then it is countered by the cops. These periodic rebellion outbursts are similar to what has been described in the original rebellion model.

4.2 Case Study #2: Building Cops vs. Assisting the Poor

To suppress rebels, there are two basic strategies that the government can use – increasing the number of cops and providing financial assistance to the poor. In the following simulations, we tried both strategies and their combinations.

First, we set the Wealth distribution=0 and Wealth to create cops=0.5. This implies that the government will send half of its wealth to maintain the existing cop force and half of it to increase cops, and no assistance to the poor. As some cops will die due to lack of sugars, the number of cops will eventually be stabilized at 158. There are occasionally one or two rebels, but they will be quickly detained by the cops surrounding them. Over time, the number of jailed individuals varies around 18 to 20.

Next, we set Wealth to create cops=1.0, so all of the wealth the government collects will be spent on building new cops. Surprisingly, in the simulation, the cop size drops to 90 and a constant mass of rebels is created in the poor regions. We observe that a poor person is likely to choose cycling actions of being dormant, then actively rebel, and then be jailed. It is because they receive no help from the government and they cannot improve their lives by moving to rich regions, and in addition, there are not enough cops to suppress rebellions. The drop of cop force is surprising, given that the government wants to create more cops. But, in fact, most of the existing cops will die due to lack of sugar supply from the government, the new cops built by the government cannot keep up the pace of the dying cops. Increasing the number of new cops is more expensive, and the new cops are located randomly (less experience), less effective countering the rebels in the poor regions. Also, with more people put into jail, less tax revenue will be collected by the government, thus less funds for the cops. Moreover, from this simulation we observe that the periodic rebellion outbursts have been replaced by constant rebellions, which indicates the overall lack of a policing force.

However, when we set Wealth distribution=0.5 (half of the wealth will be redistributed to the people under the poverty line) and Wealth to create cops=0, we observe that there is still enough wealth to support 100 cops. In fact, there are very few incidences of rebellion, and even if there is an active rebel, these will be quickly detained by the cops. Over the course of the simulation, the number of jailed individuals varies from 0 to 5. This is a peaceful society. The poor will be assisted by the government, so their grievance levels cannot be too high. Also, the government will maintain a sizable force of cops.

4.3 Other Cases

Table 1. Summary of the cases being studied and their emerging phenomena

Case	Parameter	Emerging phenomenon
Baseline case	As in Section 4.1	Periodic outburst of rebellion as in the rebellion model
Increasing the number of cops	wealth-distribution=0, wealth-to-create-cop=0.5	Number of cops increases, but still have rebels from time to time
	wealth-distribution=0, wealth-to-create-cop=1	Number of cops decreases, have a large number of rebels
Assisting the poor	wealth-distribution=0.5, wealth-to-create-cop=0	Very few rebels, the number of cops is maintained at the initial level
Freedom to move	movement?=on, wealth-distribution=0, wealth-to-create-cop=0	People move to rich regions, cops move to poor regions, rebellion happens at the boundaries of rich and poor regions
	movement?=on, wealth-distribution=0.5, wealth-to-create-cop=0.2	No rebel
Less population	number of agents=250, initial number of cops=15, movement?=on, wealth-distribution=0	The percentage of population who rebels increases, the frequency of rebellion increases
	number of agents=250, initial number of cops=15, movement?=on, wealth-distribution=0.5	No rebel
	number of agents=250, initial number of cops=15, movement?=on, wealth-distribution=0, wealth-to-create-cop=0.5	The cop number increases to 22, but still have rebels in less wealthy regions
High tax	tax=0.5	Massive rebels in less wealthy regions even when the government redistributes some of its wealth to the poor
Less education	vision=3 patches	Some people will be isolated in poor regions, rebels happen at these poor regions, government assistance will help reduce rebels
Less government legitimacy	government-legitimacy=0.35	More rebels than quiet population, cop number is reduced, the government collapses. The legitimacy level of 0.75 is the boundary value for maintaining an effective government.

5   Some Issues for Future Research

In this simple artificial society we simulated, an agent's grievance is determined by its perceived hardship and its perceived government legitimacy only. Although we are able to link the agent's perceived hardship to its wealth (sugar level), the model is far away from a real world where people have various motivations for becoming rebels. Regarding to the motivations of violent non-state actors (VNSA), Bartolomei et al. (2004) postulated a framework of summarizing grievance from the perspective of human needs. On the national level, we may identify a group of variables that are important to the health of the nation, but to use ABMs we have to associate these macroscopic measurements with microscopic perceptions of individual agents. Little research has been done on this front so far, but this is a critical link between SD model and ABM.

Finally the details of how to utilize the NOEM as a nation-state environment for ABMs still need to be researched. The overall architecture that we envisioned is that from the NOEM we can create multiple layers of maps, corresponding to the nationals economic, social and political variables, and use them to define the properties and behavior rules of agents living on the artificial land. However, to materialize this concept, in-depth investigations of the NOEM's capabilities and its potential use for ABMs are required.

Acknowledgement

References

Abdullah, I.: Africans do not live by bread alone: against greed, not grievance. Africa Review of Books, 12–13 (March 2006)

Agbonifo, J.: Beyond greed and grievance: negotiating political settlement and peace in africa. Peace, Conflict and Development (4) (2004) ISSN: 1742-0601

Bartolomei, J., Casebeer, W., Thomas, T.: Modeling violent non-state actors: a summary of concepts and methods. In: Institute for Information Technology Applications (IITA) Research Publication 4, Information Series, United State Air Force Academy, Colorado (2004)

Ballentine, K., Nitzschke, H.: Beyond greed and grievance: policy lessons from studies in the political economy of armed conflict. IPA Policy Report. Program on Economic Agendas in Civil Wars (EACW), International Peace Academy (2003)

Bodea, C., Elbadawi, I.A.: Riots, coups and civil war: revisiting the greed and grievance debate. Policy Research Working Paper, 4397. The World Bank, Development Research Group, Macroeconomics and Growth Team (2007)

Borgers, E., Kwint, M., Petiet, P., Spaans, M.: Training effect based reasoning; developing a next generation of training simulators. In: Interservice/Industry Training, Simulation, and Education Conference, I/ITSEC (2009)

Cederman, L.-E., Girardin, L.: A roadmap to realistic computational models of civil wars. In: Takahashi, S., Sallach, D., Rouchier, J. (eds.) Advancing Social Simulation: The First World Congress. Springer, Japan (2007)

Collier, P., Hoeffler, A.: Greed and grievance in civil war. World Bank (2000)

Collier, P., Hoeffler, A., Rohner, D.: Beyond greed and grievance: feasibility and civil war. Technical Report, the World Bank (2008)

Epstein, J.M.: Agent-based computational models and generative social science. Complexity 4(5), 41–60 (1999)

Epstein, J.M.: Modeling civil violence: An agent-based computational approach. Proceedings of National Academy of Science 99 (suppl. 3), 7243–7250 (2002)

Epstein, J.M., Axtell, R.L.: Growing Artificial Societies. The MIT Press, Cambridge (1996)

Kuran, T.: Sparks and prairie fires: a theory of unanticipated political revolution. Public Choice 61, 41–74 (1989)

Ganesan, A., Vines, A.: Engine of war: resources, greed, and the predatory State. Technical Report, the World Bank (2004)

Gurr, T.R., Moore, W.H.: Ethnopolitical rebellion: a cross-sectional analysis of the 1980s with risk assessments for the 1990s. American Journal of Political Science 41(4), 1079–1103 (1997)

Gurr, T.R.: People Versus States: Minorities at Risk in the New Century. Unitied States Institute of Peace Press, Washington, DC (2000)

van Lieburg, A., Petiet, P.J., Le Grand, N.P.: Improving wargames using complex system practices. Technical Paper. TNO Defense and Security, The Hague, Netherlands (2007)

von Neumann, J.: Theory of Self-Reproducing Automata. In: Burks, A.W. (ed.), University of Illinois Press, Urbana (1966)

Regan, P.M., Norton, D.A.: Greed, grievance, and mobilization: the onset of protest, rebellion, and civil war. Technical Paper. Department of Political Science, Binghamton University (2003)

Wilensky, U.: Netlogo rebellion model. In: Center for Connected Learning and Computer-Based Modelling, Northwestern University, Evanston (2004)

A Culture-Sensitive Agent in Kirman's Ant Model

Shu-Heng Chen[1], Wen-Ching Liou[2], and Ting-Yu Chen[2]

[1] Department of Economics, National Chengchi University
Taipei, Taiwan 116
chchen@mail2.nccu.tw
[2] Department of Management Information System
National Chengchi University Taipei, Taiwan 116
{wenqing.liou,timvickie2003}@gmail.com

Abstract. The global financial crisis brought a serious collapse involving a "systemic" meltdown. Internet technology and globalization have increased the chances for interaction between countries and people. The global economy has become more complex than ever before. Mark Buchanan [12] indicated that agent-based computer models will prevent another financial crisis and has been particularly influential in contributing insights. There are two reasons why culture-sensitive agent on the financial market has become so important. Therefore, the aim of this article is to establish a culture-sensitive agent and forecast the process of change regarding herding behavior in the financial market. We based our study on the Kirman's Ant Model[4,5] and Hofstede's Natational Culture[11] to establish our culture-sensitive agent based model. Kirman's Ant Model is quite famous and describes financial market herding behavior from the expectations of the future of financial investors. Hofstede's cultural consequence used the staff of IBM in 72 different countries to understand the cultural difference. As a result, this paper focuses on one of the five dimensions of culture from Hofstede: individualism versus collectivism and creates a culture-sensitive agent and predicts the process of change regarding herding behavior in the financial market. To conclude, this study will be of importance in explaining the herding behavior with cultural factors, as well as in providing researchers with a clearer understanding of how herding beliefs of people about different cultures relate to their finance market strategies.

Keywords: Culture, national culture, simulation, agent-based model, herding, conformity.

1   Introduction

The global financial crisis started in 2007 and occurred according the "Butterfly Effect" bringing serious collapse similar to a "systemic" meltdown. The crisis affected all financial markets and the entire global economy. We cannot imagine the full scope or degree of the impact. Mark Buchanan [12] suggested that agent-based computer models will prevent another financial crisis and he has been particularly influential in contributing insights. Conversely, as internet technology and globalization increase, so

does the interaction among countries. Additionally people and culture bring more variation and dissimilarity. In the past, numerous studies have attempted to explain culture, but studies of embedded agent-based models are still few. For those reason affecting the new global economy, designing a culture-sensitive agent-based model has become a central issue for financial markets. Consequently, we expect to establish a culture-sensitive agent and observe or forecast the process of change regarding herding behavior in financial markets.

The word "culture" most commonly consists of the knowledge, belief, attitudes, values, and behavior that are related to the ability for representative thinking and social learning. Culture influences the way people make decisions and performs actions. Cultural dissemination or acculturation is influenced by both external and internal factors from people, society, business, or country. Examples of external factors are new societies and technology, while internal factors are the powers of encouraging and resisting social structures and natural events. When one culture faces new ideas or events, it decides whether to accept or reject it. The phenomenon of culture change will be preceded. Therefore, culture has a significant influence on individuals, business, society, and country.

The global financial crisis of 2007 began from a liquidity deficit in the United States banking system that had a tremendous effect on society. It should be noted that herding behavior (also called conformity) also expanded the extent of the hazards caused by the financial crisis. Herding behavior describes different human behavior in various countries. Different countries have various norms regarding their cultures. The phenomenon of herding behavior in different cultures is affected by various norms. In research on culture and herding behavior, Asch [1,2] investigated the relationship between herding behavior and culture based on individualism-collectivism. The herding behavior is determined by goals of individual and is a significantly affected by cultural. Based on obedience experiments, Stanley Milgram [3] proved that Norway engages in more herding behavior than France does, confirming the existence of culturally different factors affecting herding behavior.

There has been a significant growth in research regarding agent based simulation in recent decades. In herding behavior of agents, Kirman's model[4,5] established herding behavior (also called conformity) in aggregate expectations stimulating agent interactions. The phenomenon of model is more psychologically driven than economically driven. This herding mechanism was inspired by an observation in entomology. "Ants, faced with two identical food sources, were observed to concentrate more on one of these, but after a period they would turn their attention to the other." Inspired by observing the behavior of ants, Kirman characterized the swinging potential of each individual by two parameters, namely, a probability of self-conversion and a probability of being converted. The self-conversion probability raises the probability that an individual will change to other types of agents without external influence, while the probability of being converted increases the probability that the individual will be influenced by other individuals. This changing process is discrete and can be deliberated in a general theoretical framework of a Polya urn process [13] (N-path dependence).

Culture-sensitive agent models have focused on the interactions of individuals from different cultures. The Axelrod [6] model presents a viewpoint of formal cultural dissemination. Axelrod used cellular automata techniques to simulate individual communication and interplay regarding each other's belief in local networks when they have similar traits. The traits of all the individuals could directly converge to new a

culture. Individuals were affected by society. Therefore, Axelrod's model is understood to be a formalized form of culture. A formal culture can be defined apart from any diffusion and the changing rules involved. This is conducted by designating a set of mathematical symbols and a set of formation rules that determine which mathematical symbols are well-formed formulas. When transformation rules (also called rules of inference) are fired, and certain traits are accepted as axioms (together called a deductive system or a deductive apparatus), a logical system is formed. A new culture is created and influenced by these firing rules. Therefore, the rules of logic depend upon their traits.

According to the Axelrod definition, behavior of individuals is followed by a certain pattern, and if individuals detach from this model they are subject to a penalty. The individual's behavior pattern is affected by the culture of society.

Furthermore, Geert Hofstede's five dimensions of culture have been used to describe cultural differences. Recently, Geert Hofstede, Gert Jan Hofstede, Catholijn Jonker and Tim Verwaart had incorporated four of these five dimensions into an agent-based model of trade negotiation. The four dimensions are power distance [7], individualism [8], uncertainty avoidance [9], and long-term orientation [10].

To what extent does culture have influence on herding behavior? How do different countries with diverse cultural norms cope with herding behavior? In this paper, we present herding behavior regarding culture in four scenarioes: Eastern culture, Western culture, confliction, and fusion. We combined Kirman's Ant Model of herding behavior and one of the five dimensions in Hofstede's model of national culture, individualism, and collectivism. We formulated rules of culture and herding behavior to discuss how and why herding behavior is embedded in different cultures.

2   The Model

The culture-sensitive Kirman's Ant Model consists of two parts: culture-sensitive and herding behavior. In the culture-sensitive part, we focus on individualism and collectivism aspects of Hofstede's cultural consequences [11]. National culture has some distinctions between norms in collectivist and individualist societies, as shown in Table 1. Individualist societies are loose and most individuals look out for himself or

Table 1. Distinctions between norms in collectivist and individualist societies

Collectivist	Individualist
Maintain harmony, avoid confrontation	Speak your mind
High-context, implicit communication	Low-context, explicit communication
Use the word "we"	Use the word "I"
Show favor to in-group customers	Treat all customers equally
No business without a personal relation	Task is more important than a good relation
A relation brings rights and obligations	Mutual advantage is the basis of relations
Relations are given	Build and maintain relations actively
Save face for in-group	Keep self-respect
Responsible for group interests	Responsible for personal interests

Source : Hofstede [11].

herself. Whereas the collectivist society integrates into cohesive groups and protection is exchanged for loyalty. Asch [1,2] proved that agents increase herding behaviors from collectivist societies and decrease from individualist societies.

However, Hofstede [11] examined numerous different cultures on individualism and collectivism dimension (IDV). China, Hong Kong, Singapore, and Taiwan have lower IDV value; their cultures are called Eastern culture. Australia, Canada, United Kingdom, and United States have higher IDV value, are called Western culture, as shown in Table 2.

Table 2. Individualism and Collectivism Dimension Value

Eastern Country	IDV	Western Country	IDV
China *	20	Australia	90
Japan	46	Canada	80
South Korea	18	United Kingdom	89
Hong Kong	25	United States	91
Singapore	20	Germany	67
Taiwan	17	Sweden	71

Source : Hofstede[11].

Conversely, Kirman applied ant behavior, also known as the urn process path [13] (N-path dependence), to explain the phenomenon of herding and epidemics, similar to the stock market. The model was an entomologist's experiments. The ants of model faced with two food sources: black and white state of the system. The observation concentrated one of the states, after a period they would change their mind to the others[4].

This system had two dynamic parameters (δ, ε), converting rate, and self-converting rate. The Converting Rate $(1-\delta)$ determines the probability of the first ant and the second ant meeting. The Self-Converting Rate (ε) decides the probability of the first ant changing its color independently before meeting the other ant.

This paper combines the cultural parameter (I) of individualism and collectivism and Kirman's ant model as a culture-sensitive agent. We report the observation of how culture affects herding behavior. The switching rate of the modified Kirman's Ant Model is as follows.

(1) k to k+1 with probability

$$p_1 = P(k, k-1) = \left(1 - \frac{k}{N}\right) \left(\varepsilon + (1 - \delta - I) \frac{k}{N-1}\right) \tag{1}$$

(2) k to k-1 with probability

$$p_2 = P(k, k-1) = \left(\frac{k}{N}\right) \left(\varepsilon + (1 - \delta - I) \frac{N-k}{N-1}\right) \tag{2}$$

The algorithm is shown in Figure 1.

Step 1. Randomly initialize ant's position on black or white.
Step 2. Is the algorithm enough stop criteria? If it is enough then stop run.
Step 3. Self-Converting Rate Rule
      Select first ant and use self-converting rate (ε) to determine if the ant can covert by itself.
Step 4. Converting Rate Rule
      Select first ant and second ant, then use converting rate: $(1-\delta-I)$ to determine if the first ant can convert.
Step 5. Calculate black ants of next state
      applying a probabilistic state transition rule to calculate equilibrium distribution, K/N, and frequency.
Step 6. The model stops at its criteria, if not it will go back to Step 3.

Fig. 1. Algorithm of culture-sensitive Kirman's ant model

3 Experiment Result

Our simulation was based on the uniform distribution of Kirman's Ant Model with converting rate $(1-\delta, \delta=0.02)$ and self-converting rate $(\varepsilon, \varepsilon=0.01)$.The cultural ingredient affected on delta value(δ) and converting rate$(1-\delta)$ according to previous investigation. We then fixed the number of agents at N=100 and ran 100,000 times in six scenarios with different initial conditions.

Case 0: No Cultural Factor
      Initial condition: 100 individuals put on black and white source randomly.
Case 1: Eastern Culture
      Initial condition: 100 Eastern individuals put on black and white source randomly.
Case 2: Western Culture
      Initial condition: 100 Western individuals put on black and white source randomly.
Case 3: Cultural Conflict
      (1) Initial condition: 50 Eastern individuals put on black source and 50 Western individuals put on white source.
      (2) Initial condition: 50 Western individuals put on black source and 50 Eastern individuals put on white source.
Case 4: Cultural Fusion
      (1) Initial condition: 50 Eastern individuals put on black source and 50 Western individuals put on white source.
      (2) Initial condition: 50 Western individuals put on black source and 50 Eastern individuals put on white source.
      The first ant change cultural ingredient (δ) when meet the second ant. The new cultural ingredient (δ) of the first ant will be equal to the average of the first cultural ingredient (δ_1) and the second cultural ingredient (δ_2).
Case 5: Forecasting Culture
      Initial condition: 100 individuals with random cultural factor put on black and white source randomly.

Case 0	Case 1	Case 2
Case 3-(1)	Case 3-(2)	Case 4-(1)
Case 4-(2)	Case 5-(1)	Case 5-(2)
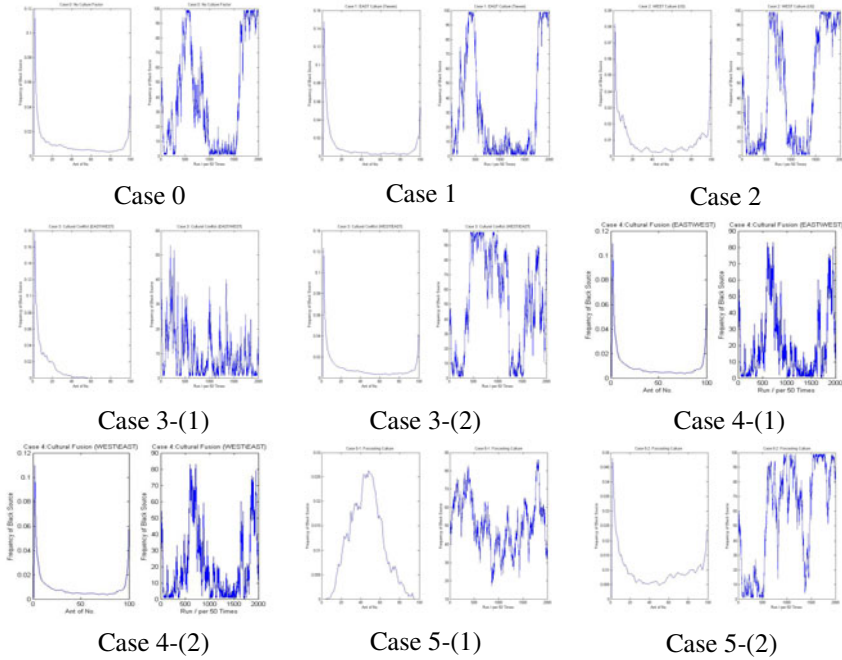
Fig. 2. The simulation results

The simulation results are illustrated in Figure 2. In Figure 2, Case 0 is a non-culture-sensitive factor of herding behavior and is an original Kirman's Ant Model. In Case 1 and Case 2, the ants have only one pure culture (Eastern or Western). We compared Case 0, Case 1, and Case 2 and determined the culture of individuals increasing the economic situation of optimism and pessimism.

Case 1 refers to the Eastern society and its collectivist cultural context. The individuals are affected by groups in herding behavior. This society has consciousness that is more consistent and sticks on a black or white source for a long time. If the economic environment is not favorable, the individuals are more pessimistic. If they want to become optimistic, they must spend more time to influence each other. Because they have a strong sense of collective consciousness and do not easily break, and vice versa. Hence, government financial policies probably apply a long-term and powerful policy to manipulate investor views. Case 2 refers to the Western society and its individualist cultural context. The individuals exhibit strong minds and diverse beliefs. This society fluctuates and swings on a black or white source quickly. If the economic environment is terrible, the investors are pessimistic at the start. After a period, they will have dissimilar opinions on the economic situation. The individuals contain the differences from cognitive and on the consciousness wandering between optimism and pessimism. They have a strong sense of individual and easily break, and vice versa. Therefore, the financial policies of government may use a fast and dynamic policy to persuade investor views. In Case 3, we considered the experimental case of cultural conflict. A society has two different cultures, and do not completely accept each other.

The difference between the experiments in Cases 3-(1) and 3-(2) involves the initial setup condition in which culture ants are in black source, Eastern or Western? That means our concern which culture.

We determined that the herding of Eastern society in Case 3-(1) sticks more than Case 1. The herding of Western society in Case 3-(2) fluctuates more than Case 2. The experimental results show that they are impacted by group distance and are influenced increasingly more by the original cultural consciousness. The individualist is more toward individualism. The collectivist is more toward collectivism.

In Case 4, we considered the experimental case of cultural integration or fusion. A society has two different cultures that accept each other and the respective cultures one hundred percent. The difference between experiments in Case 4-(1) and 4-(2) involves the initial setup condition in which culture ants are in black source, Eastern or Western. Regardless of which initial condition the experiments start, they have the consistent result and the same simulating process of herding behavior. The experimental results produce a new culture similar to a higher culture's group to integrate two different cultures. We can realize the interaction of two cultures in a society and accommodate each other. They will create a new culture; the extent of their new culture and characteristics of the collective consciousness will be moderated between two cultures.

Cases 5-(1) and 5-(2) predicted herding behavior of which country culture. The results observe individualism value (IDV) of Case 5-(1) is 0.80 closing to Canada culture behavior and Case 5-(2) is 0.30 closing to Hong Kong culture behavior.

4 Conclusion

The paper presents the effect of financial market herding behavior on national culture of individualism versus collectivism. We take a step in the direction of inserting cultural factors between simulating agents and herding behavior in financial markets. Furthermore, we create a model to simulate and forecast herding behavior with different cultural styles. Such findings underscore the importance of recognizing cultural styles with herding behavior. Finally, three of these findings are worth summarizing:

(1) We establish a culture-sensitive agent based model and observe the swing process with herding behavior in financial markets.
(2) We can reproduce and recognize the important herding behavior of the stock market through a mechanism of simulation with cultural factors.
(3) We consider the differences between individual and collective consciousness through national culture. Additionally, we examine how the investors within the stock market face the same issue and determine how to evaluate integration or divergence of different cultures.

Future research is obviously required, but this is an exciting first step. We will hopefully clarify the impact from other dimensions, such as power distance, uncertainty avoidance, long-term orientation, masculinity of Hofstede's national culture, determining the herding behavior pattern (transiting frequency and type), and research more complex culture-sensitive herding behavior regarding social networks in the financial market.

References

1. Asch, S.E.: Socialpsychology. Prentice-Hall, Englewood Cliffs (1952b)
2. Asch, S.E.: Studies of independence and conformity. A minority of one against a unanimous majority. Psychological Monographs 70(9), whole no. 416 (1956)
3. Milgram, S.: Obedience to Authority; An Experimental View. Harpercollins, New York (1974)
4. Axelrod, R.: The dissemination of culture. Journal of Conflict Resolution 41(2), 203–226 (1997)
5. Kirman, A.: Epidemics of opinion and speculative bubbles in financial markets. In: Taylor, M.P. (ed.) Money and Financial Markets, pp. 354–368. Blackwell, Cambridge (1991)
6. Kirman, A.: Ants, rationality and recruitment. Quarterly Journal of Economics 108, 137–156 (1993)
7. Hofstede, G.J., Pedersen, P.B., Hofstede, G.: Exploring Culture: Exercise, Stories and Synthetic Cultures. Intercultural Press, Yarmouth (2002)
8. Hofstede, G.J., Jonker, C.M., Verwaart, T.: Modeling Power Distance in Trade. In: David, N., Sichman, J.S. (eds.) MAPS 2008. LNCS, vol. 5269, pp. 1–16. Springer, Heidelberg (2009)
9. Hofstede, G.J., Jonker, C.M., Verwaart, T.: Modeling Culture in Trade: Uncertainty Avoidance. In: 2008 Agent-Directed Simulation Symposium (ADSS 2008), Spring Simulation Multiconference 2008, pp. 143–150. SCS, San Diego (2008)
10. Hofstede, G.J., Jonker, C.M., Verwaart, T.: Individualism and Collectivism in Trade Agents. In: Nguyen, N.T., Borzemski, L., Grzech, A., Ali, M. (eds.) IEA/AIE 2008. LNCS, vol. 5027, pp. 492–501. Springer, Heidelberg (2008)
11. Hofstede, G.: Culture's Consequences, 2nd edn. Sage Publications, Thousand Oaks (2001)
12. Buchanan, M.: Meltdown modeling, Could agent-based computer models prevent another financial crisis. Nature 460(6), 680–682 (2009)
13. Brian Arthur, W.: Increasing Returns and Path Dependence in the Economy, pp. 6–8, 17, 112–113. University of Michigan Press, Ann Arbor (1994)

Modeling Link Formation Behaviors in Dynamic Social Networks

Viet-An Nguyen[1], Cane Wing-Ki Leung[2], and Ee-Peng Lim[2]

[1] Department of Computer Science, University of Maryland,
College Park, MD, USA
[2] School of Information Systems,
Singapore Management University, Singapore

Abstract. Online social networks are dynamic in nature. While links between users are seemingly formed and removed randomly, there exists some interested link formation behaviors demonstrated by users performing link creation and removal activities. Uncovering these behaviors not only allows us to gain deep insights of the users, but also pave the way to decipher how social links are formed. In this paper, we propose a general framework to define user link formation behaviors using well studied local link structures (i.e., triads and dyads) in a dynamic social network where links are formed at different timestamps. Depending on the role a user plays in a link structure, we derive different types of *link formation behaviors*. We develop models for these behaviors and measure them for a set of users in an Epinions dataset.

Keywords: Link formation behaviors, link formation rules.

1 Introduction

Social links represent a rich set of structural knowledge about the linked users beyond their individual attributes (e.g., age, gender, etc.). At the network level, network properties such as density, diameter, user degree distribution, etc., have been well studied by both social and computer science researchers[6,3]. At the micro or node level, the social links reveal a user's preference of friends, and her preferred way to form social links with others. Links can be further utilized in a number of commercially interesting applications including item recommendation, information diffusion and community discovery.

In this paper, we examine link formation behaviors of each user and derive some models for measuring the behaviors. The link formation behaviors here are motivated by a set of dyadic and triadic pattern rules discovered from dynamic social network data. We define a *dynamic directed social network* to be a graph $G = (V, E, T, t)$. V is a set of vertices/nodes representing individuals in the network. E is a set of directed edges representing social links, such as friendship and trust links, between individuals. An element $(v_i, v_j) \in E$, where $v_i, v_j \in V$, is an edge from v_i to v_j. $t : E \to T$ is a mapping between edges and their timestamps. Without loss of generality, we represent timestamps as $T = \{t_h | h \geq 0\}$, such that $\forall t_{h_1}, t_{h_2} \in T, t_{h_1} < t_{h_2}$ *iff* $h_1 < h_2$. A graph may evolve with new

nodes and edges joining at different time points. G can be viewed as a snapshot of a social network taken at a certain time point t_h, such that G contains nodes and edges that were formed at or before t_h. For simplicity, we assume that nodes and edges are not removed after creation.

We first define a representation for *link formation (LF) rules* (see Section 3). Unlike most earlier link patterns studied, we require each link formation rule to have pre-conditional link structure formed before the consequence link. For link formation rule, we then derive link formation behaviors based on the way a user forms links with other users in Section 4. In particular, we introduce several link formation behaviors including *rule usage* and *rule confidence* associated with different link formation rules. Such behaviors can be defined at both node (user) and instance levels. Here, an instance refers to a subgraph of users and their links following the rule structure. We assign a score function to each behavior so as to measure it quantitatively. We finally apply our proposed behaviors on the Epinions data (see Section 5).

2   Related Work

There are very little work on link formation behavior modeling and analysis for dynamic social networks. Most research in the past studied known dyadic and triadic structures in static social networks[10]. Recently, these local structures are extended with time order when analyzing dynamic social networks[2,8,7]. There are also new research on mining network specific local structures[4].

Leskovec, et al. studied the "edge destination selection process" based on some simple triangle closing models[2]. The networks they have considered are temporal but undirected networks, and the models introduced do not consider individual's link formation behaviors. Romero and Kleinberg, in a recent paper [8], showed that directed closure (also known as transitivity) is used by Twitter users in forming links with one another. A measure known as *closure ratio* was introduced to measure how likely the incoming links of a user node exhibits directed closure. Instead of examining links formed by transitivity only, this paper covers behaviors related to a variety of link formation rules. In our earlier paper [7], we introduced trust reciprocity related behaviors that are shown to help predicting reciprocal trust links. This paper enlarges the behavioral study to include other link formation rules.

3   Link Formation Rules

Link Formation rule (LF-rule) is designed to describe an *observable social effect* facilitating or affecting the formation of a link from a certain node v_i to another node v_j, where v_i and v_j are called the *start* node and the *end* node of a LF-rule respectively. A LF-rule captures two important constraints. The first constraint is a certain *pre-condition* link structure related to the start node and/or the end node. The second constraint is the *temporal constraint* that the link from the start node to the end node must be formed after the pre-condition is formed.

Our LF-rules consist of *dyadic and triadic structures* which have long been recognized as interesting structures for understanding and predicting the dynamics of large, complex networks [5,1,9,8]. Figure 1 depicts five LF-rules (labeled r_1 to r_5) constructed from basic dyadic and triadic structures, which are the building blocks of local structures. In each LF-rule, the nodes labeled as s and e respectively correspond to its start node and end node. Recall that the (s, e) link (shown in blue in Figure 1) in a LF-rule is required to be formed later than other links in the same rule.
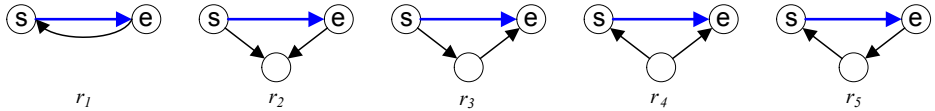
Fig. 1. LF-rules based on basic dyads and triads, including reciprocity (r_1), common out-neighbor (r_2), transitivity (r_3), common in-neighbor (r_4) and cycle (r_5)

Hence, the LF-rule set $\mathcal{R}$ in this study consists of the above five rules, i.e., $\mathcal{R} = \{r_1, \cdots, r_5\}$. As we are interested in link formation behaviors caused by local structural effect instead of some non-local (e.g., preferential attachment) and out-of-network (e.g., users already know each other before joining the network) effects, we focus on links formed with users within two-hop distance. Hence, users within two-hop distance apart must satisfy the pre-condition(s) of at least one of the given rules.

4   Link Formation Behaviors

Given the LF-rule set $\mathcal{R}$, we characterize individual nodes by their Link Formation behaviors (LF-behaviors), which describe the extent to which the nodes follow specific LF-rules in forming links. Consider that v_i takes on the role of start node, we can derive two LF-behaviors that corresponds to the usage and confidence of a rule r_l. We therefore define the following two kinds of LF-behaviors: *rule usage* and *rule confidence*.

To define the rule usage and confidence as measurable behaviors, we first introduce some important notations. Given an input dynamic social network $G = (V, E, T, t)$, an *instance* of the LF-rule r_l is a subgraph in G that: (a) is isomorphic to the graph of r_l; and (b) has the consequence edge formed after the pre-condition link structure. The set of instances of r_l with v_i and v_j taking the start node and end node roles respectively is known as the *instance set* of (v_i, v_j) w.r.t. r_l, and is denoted by $\mathbf{R}_{ij}^l$. The instance set of start node v_i w.r.t. r_l, $\mathbf{R}_{i*}^l$, is defined as $\cup_j \mathbf{R}_{ij}^l$. The node set of v_i w.r.t. r_l, $\mathbf{UR}_{i*}^l$, is defined as $\{v_j | \mathbf{R}_{ij}^l \neq \phi\}$. We define the instance of the pre-condition of r_l as a subgraph in G that is isomorphic to the pre-condition link structure of r_l. The set of instances of pre-condition of r_l with v_i and v_j taking the start node and end node roles respectively is known as the *pre-instance set* of (v_i, v_j) w.r.t. r_l, and is denoted

by $\mathbf{P}^l_{ij}$. The pre-instance set of v_i w.r.t. r_l $\mathbf{P}^l_{i*}$ is therefore defined as $\cup_j \mathbf{P}^l_{ij}$. The pre-node set of v_i w.r.t. r_l, $\mathbf{UP}^l_{i*}$, is defined as $\{v_j | \mathbf{P}^l_{ij} \neq \phi\}$.

We define the *rule usage at node level, rule usage at instance level, rule confidence at node level, rule confidence at instance level* behaviors of a user v_i w.r.t. rule r_l in Equations 1, 2, 3, and 4 respectively.

$$\mathbf{NUsage}^l_{i*} = |\mathbf{UR}^l_{i*}| / \sum_{r_k \in \mathcal{R}} |\mathbf{UR}^k_{i*}| \quad (1) \qquad \mathbf{Usage}^l_{i*} = |\mathbf{R}^l_{i*}| / \sum_{r_k \in \mathcal{R}} |\mathbf{R}^k_{i*}| \qquad (2)$$

$$\mathbf{NConf}^l_{i*} = |\mathbf{UR}^l_{i*}| / |\mathbf{UP}^l_{i*}| \quad (3) \qquad \mathbf{Conf}^l_{i*} = \sum_j |\mathbf{R}^l_{ij}| / \sum_j |\mathbf{P}^l_{ij}| \qquad (4)$$

The rule usage at node level in Eq 1 reveals the proportion of users v_i has links to are based on rule r_l. Note that it is possible that v_i may link to a user v_j using different link formation rules, i.e., $\mathbf{UR}^l_{ij} \cap \mathbf{UR}^k_{ij} \neq \phi$ for $r_l \neq r_k$. The rule usage at instance level in Eq 2 measures the proportion of v_i's rule instances are based on rule r_l. Multiple rule instances based on a rule can be associated with a link from v_i to another user v_j, as there can be multiple instances of the rule's pre-condition occurring before the (v_i, v_j) link.

The rule confidence at node level in Eq 3 measures the proportion of users v_i has connections to based on rule r_l are subsequently linked from v_i directly. The rule confidence at instance level in Eq 4 measures the proportion of v_i's pre-condition instances based on rule r_l are subsequently linked from v_i directly. $\mathbf{NConf}^l_{i*}$ may not be equal to $\mathbf{Conf}^l_{i*}$ except for reciprocity rule which has one pre-condition user instance corresponding to one pre-condition graph instance.

5   Link Formation Behavior Analysis of Web of Trust

5.1   Dataset

We conduct an experimental study on the *Epinions* dataset available at http://www.trustlet.org/wiki/Extended_Epinions_dataset. Epinions contains a directed and time-stamped Web of Trust (WOT) with *trust* edges. About 69% of edges come with an initial timestamp of 2001/01/10 (t_0), which represents *all timestamps on t_0 or prior to t_0*. The formation date and order of all edges formed after t_0 are known. As temporal information is important in characterizing LF-behaviors, we discarded an edge (v_i, v_j) with timestamp t_0 unless both users v_i and v_j were involved in at least one edge formed after t_0. As we are interested in users with sufficient link formation history, we require each user v_i to creating at least 20 out-links and in-links. There are 1295 users meeting this selection criteria.

Figure 2 shows the distribution of outdegree of our users. Note that these out-links may be created to users more than two hops away but such links are the minority. On average, more than 90% of the out-links are directed to users within two-hop away which is shown in Figure 2(b). This also illustrates the dominance of local structural effect in forming links.

(a) Outdegree Distribution

(b) Fraction of Outlinks created to two-hop neighbors
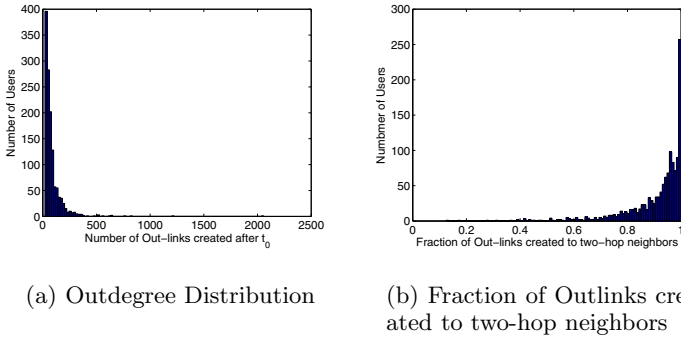
Fig. 2. Statistics of Outlinks

5.2   Link Formation Behavior Distribution

We first examine the distribution of LF-behavior scores among users. Figure 3 shows the rule usage behaviors scores of all users ordered from highest to lowest score values. The values on the x axis are the rank positions of users. For different rules, the same rank may refer to different users.

(a) At Node Level

(b) At Instance Level
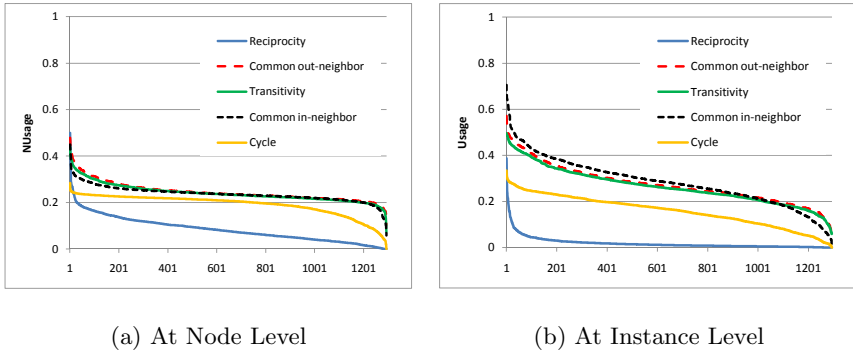
Fig. 3. Distribution of Rule Usage Behaviors

Figure 3 shows that common out-neighbor, transitivity, and common in-neighbor are the rules mostly used among the users. This is followed by cycle and reciprocity rules. Based on node level rule usage behavior values in Figure 3(a), most users have about 24% of their links formed involving each of these three rules. On average, users have only 18% and 8% of their links formed using cycle and reciprocity respectively. Similar observations can be made for the instance level rule usage behaviors (see Figure 3(b)) except that the proportions of using reciprocity and cycle are even lower. These behavior scores suggest that users may either have less distinctive opportunity or confidence to use the reciprocity and cycle rules. The exact reasons are revealed in Figure 4.

Figure 4 shows that users have relatively higher confidence for using reciprocity rule compared with other rules. On average, users have 19% confidence for this rule but not more than 5% confidence for other rules at the node and instance levels. With this, we conclude that users have low usage for reciprocity because of less distinctive opportunity instead of low confidence. While users may have more links formed with common out-neighbor, transitivity, and common in-neighbor rule, their confidence of using these rules is actually very low.

(a) At Node Level   (b) At Instance Level
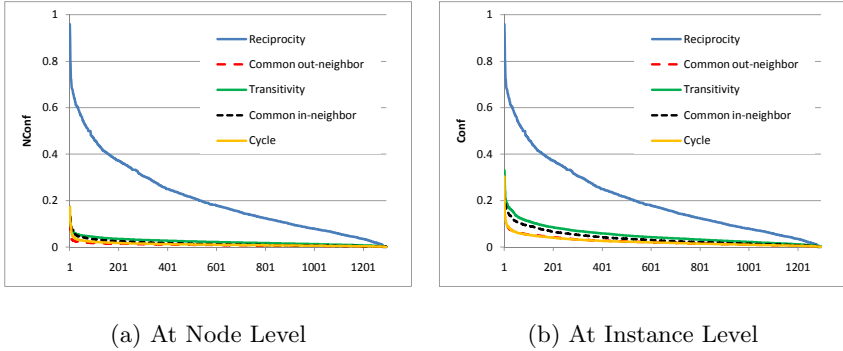
Fig. 4. Distribution of Rule Confidence Behaviors

5.3   Node Level vs. Instance Level Behaviors

Figures 5 and 6 depict the scatterplots of users' behavior scores at node and instance levels. Each figure has x and y axes representing the node and instance level behaviors respectively, and each point represents a user's node and instance level behavior scores. From the two figures, some interesting observations are: (1) The behavior scores at node and instance levels are well correlated. Hence, when a rule is used often for linked users, it is likely that the rule is also used often for the link formation structures involving the linked users; (2) Except for reciprocity, the behavior scores at node level tend to be smaller that the corresponding scores at instance level. This shows that for triadic rules, users usually form multiple pre-condition instances before actually forming the direct link.
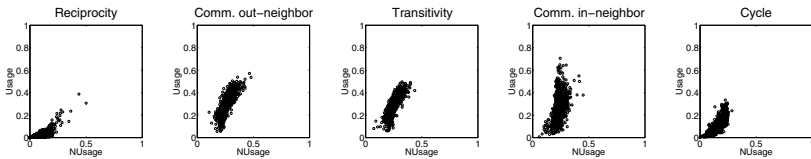
Fig. 5. Correlation of Node and Instance Level Rule Usage
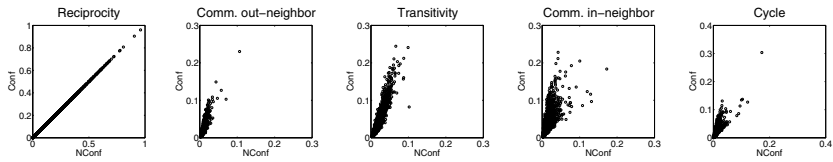
Fig. 6. Correlation of Node and Instance Level Rule Confidence

5.4 Stability of User Behaviors

Next we would like to examine the stability of user behaviors over time. The rule usage and rule confidence of each user defined in Equations 1 to 4 can be computed at different time points where $\mathbf{UR}_{i*}^{l}$, $\mathbf{UP}_{i*}^{l}$, $\mathbf{R}_{i*}^{l}$ and $\mathbf{P}_{i*}^{l}$ contain the corresponding set of users and instances up to the given time. For simplicity, we compute for each user his/her behavior scores every time the user creates an out-link (instead of for every possible time point). Thus, given a user for each type of

(a) Node level rule usage

(b) Instance level rule usage

(c) Node level rule conf.

(d) Instance level rule conf.
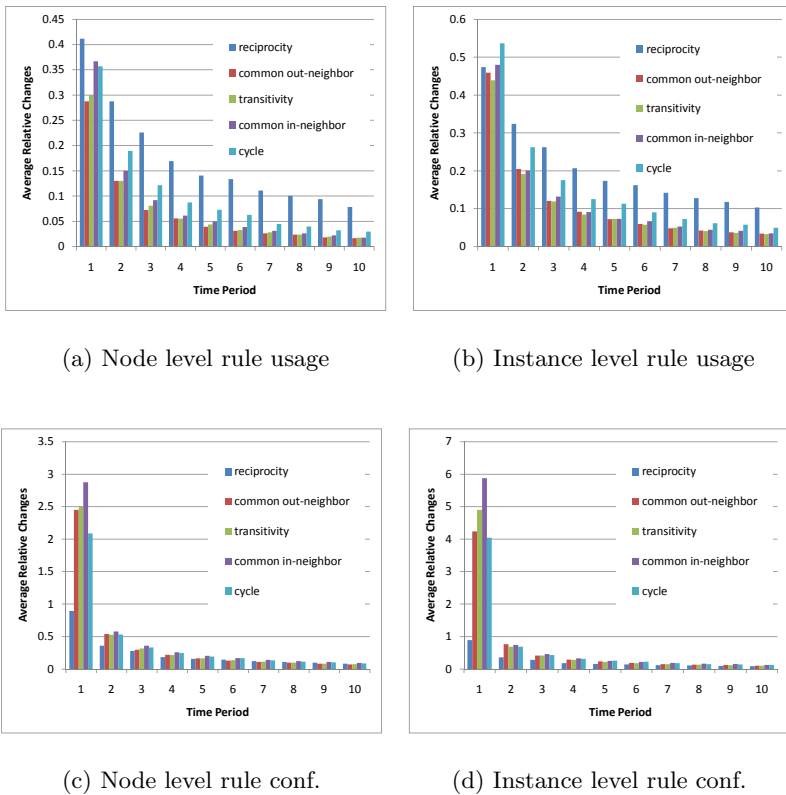
Fig. 7. Relative changes of different behaviors averaged over all users

behavior there is a time series of scores. Given a series of user behavior scores, we split it into 10 equal partitions and obtain a sub-series of scores at each splitting time. Each series of behavior scores of a user now is characterized by 11 representative values $x_0, x_1, \cdots, x_{10}$.

For each partition $i \in [1, 10]$ we compute the relative change as $\frac{x_i - x_{i-1}}{x_{i-1}}$. Figure 7 shows the average relative changes over all users for different behavior scores. As we can see, the average relative changes of all behavior scores decrease as time increases which shows that the users tend to follow the behaviors defined by our rules more consistently over time. In the case of rule confidence, the stability is even more obvious after the 1st partition. Thus, our proposed behavior scores can be used to effectively characterize how users form links in dynamic social networks which is useful in various tasks including user clustering and link prediction.

6   Conclusion

In this paper, we introduce a set of link formation behaviors derived from a set of link formation rules for dynamic social networks. Using an Epinions dataset, we show that active users have their links formed using rules with different rule usage and confidence behaviors. Reciprocity rule may not enjoy high usage but most users have relatively higher confidence using it. The node and instance level behaviors are found to be correlated. We further find the behaviors become more stable as users establish more links. With this knowledge, we believe that link behaviors can be a part of user profile which can be useful in social network applications such as link prediction and recommendation. This will also be the direction for our future research.

Acknowledgement

References

1. Faust, K.: Very local structure in social networks. Sociological Methodology 31(1), 209–256 (2007)
2. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: SIGKDD (2008)
3. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: SIGKDD (2005)
4. Leung, C.W.-K., Lim, E.-P., Lo, D., Weng, J.: Mining interesting link formation rules in social networks. In: CIKM (2010)
5. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: Simple building blocks of complex networks. Science 298(5594), 824–827 (2002)
6. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)

7. Nguyen, V.-A., Lim, E.-P., Tan, H.-H., Jiang, J., Sun, A.: Do you trust to get trust? A study of trust reciprocity behaviors and reciprocal trust prediction. In: SIAM SDM (2010)
8. Romero, D.M., Kleinberg, J.: The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. In: ICWSM (2010)
9. Snijders, T.A.B., Van de Bunt, G.G., Steglich, C.E.G.: Introduction to stochastic actor-based models for network dynamics. Social Networks, Special Issue on Dynamics of Social Networks 32(1), 44–60 (2010)
10. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)

Naming on a Directed Graph

Giorgio Gosti and William H. Batchelder⋆

Institute for Mathematical Behavioral Sciences (IMBS),
School of Social Sciences, University of California,
Irvine, CA 92697, USA
{ggosti,whbatche}@uci.edu

Abstract. We address how the structure of a social communication system affects language coordination. The naming game is an abstraction of lexical acquisition dynamics, in which N agents try to find an agreement on the names to give to objects. Most results on naming games are specific to certain communication network topologies. We present two important results that are general to any graph topology: the first proves that under certain topologies the system always converges to a name-object agreement; the second proves that if these conditions are not met the system may end up in a state in which sub-networks with different competing object-name associations coexist.

Keywords: Naming game, evolution of conventions, social networks, Markov chains.

1   Introduction

Communities of agents evolve agreements on how to name objects, and one formal approach to study this evolution is referred to as the naming game. The naming game refers to a situation where a collection of agents come to agreement on the name of an object through a dynamic process involving interactions between pairs of agents, where one agent proposes a name for the object and the other agent indicates success or failure. The game has been studied by a number of scholars in such areas as physics [1], artificial life [2], theoretical economics [3], theoretical biology [4], computer science [5]. In a typical study a communication structure is assumed, and rules are set up for selecting the sequence of agent pairs, for communication of proposals, for proposal agreement, and the consequences of success or failure of the proposals.

An important part of the naming game literature has focused on constraining the interactions to a communication network, and discussing how different network structures qualitatively effect the evolution of conventional naming systems. Among different network structures studied are completely connected networks [1], small world graphs [6], dynamic networks [7], and empirical social networks [3]. Some other works are more concerned with how agents invent new

---

words [8], or with different selection rules for the words that are proposed in the interactions [9]. Moreover, other works focus on different selection rules for sequences of agents-pair interactions [10].

This paper studies the naming game on an arbitrary directed graph, where on each of a series of discrete trials one agent communicates a proposed name for an object from their list to another agent, and following that communication each of these agent may update their list of possible names for the object. Convergence of the system occurs when each agent has a list of size one with the same name. We provide necessary and sufficient conditions on the directed graph structure to guarantee convergence with probability one, and we also discuss some possibilities that can occur if convergence is not guaranteed. In addition some simulations are provided that bear on some of the timing aspects of the game.

This paper is divided into five main sections. After the introduction, Sec. 2 presents the rules for the naming game played on a digraph. Then Sec. 3 shows that the naming game dynamics is a homogeneous Markov chain, and limiting theorems are provided as a function of the digraph. In Sec. 4 two simple simulation studies illustrate the limiting theorems, and finally some plans for future work are discussed.

2   The Naming Game Algorithm

2.1   Brief Review on Naming Games

The naming game [2,1,4,11] describes a set of problems in which a number of agents bootstrap a commonly agreed name to each object in a set representing the agents environment. Each naming game is defined by an interaction protocol that leads to a dynamical system that evolves over time. There are two main aspects of the system:

- One posits rules covering how the agents interact locally using simple deterministic or probabilistic rules to update their states.
- One attempts to discover if the dynamic system so defined converges to a consistent state consisting of a commonly accepted name for each object.

In this paper as in [1] there is only one object and time is discrete. We assume that a digraph determines the possibilities for agent communication, where each arc (ordered pair) represents, respectively, a potential speaker and listener. The focus of the paper will be on formal theorems on convergence as a function of the structure of the digraph.

2.2   Some Notation

In the digraph naming game there are N agents, $A = \{a_i | 1 \leq i \leq N\}$; and one object with M possible names, $W = \{w_j | 1 \leq j \leq M\}$. The communication structure is represented by a digraph, $\mathcal{D} = \langle A, V \rangle$, where $V \subseteq A \times A$ is a binary relation on A. The members of V are potential speaker-listener pairs of agents, $v = (a_s, a_l)$. Discrete time is denoted by $T = \{t | t = 1, 2, \ldots\}$. At each time point

t, each agent a_i has a list of possible words for the object denoted by $s_{i,t} \in \wp(W)$, where $\wp(W) = \{L|L \subseteq W\}$ is the power set of W. The global state of the system at any time point t denoted by $\mathbf{s_t} = \langle s_{i,t} \rangle$ is a vector of the lists of all agents drawn from the system state space $S = \{\mathbf{s}|\mathbf{s} \in X_{i=1}^{N} \wp(W)\}$ consisting of the N-fold Cartesian product of $\wp(W)$. Finally if at time t a pair $v_t = (a_{s,t}, a_{l,t})$, is selected (see **R2** below), for convenience we denote the speaker's and listener's lists, respectively by $s_{s,t}$ and $s_{l,t}$.

2.3   The Rules of the Game

We define the naming game algorithm as a set of five rules. The first rule defines the state of the agents at time $t = 1$, the second rule describes how the speaker-listener pairs are selected, the third describes how the speaker selects a word to communicate to the listener, the fourth describes listener feedback to the speaker, and the fifth describes how the system updates its state.

R1 Initial State. At $t = 1$ the state is $\mathbf{s_1} = X_{i=1}^{N} \emptyset$, where $\emptyset$ is the empty set.

R2 Selection. On each time t, the speaker-listener pair, $V_t \in V$, is determined independent of previous selections with uniform probability, $\forall v \in V, \Pr(V_t = v) = 1/|V|$.

R3 Speaker Rules. Suppose at time t the state is $\mathbf{s_t}$, and the selected pair is $v_t = (a_{s,t}, a_{l,t}) \in V$, then $a_{s,t}$ probabilistically selects a word B_t to propose to the listener $a_{l,t}$ as follows.

　Case 1 If $s_{s,t} = \emptyset$, then $\forall w \in W, \Pr(B_t = w) = 1/M$,*invention*.

　Case 2 If $s_{s,t} \neq \emptyset$ then the spoken word is selected at random from the speaker's list, namely $\forall w \in s_{s,t}, \Pr(B_t = w) = 1/|s_{s,t}|$.

R4 Listener Rules. Suppose at time t the state of the system is $\mathbf{s_t}$, and the selected pair is $v_t = (a_{s,t}, a_{l,t}) \in V$, and the proposed word is $b_t \in W$. Then:

　Case 1 If $b_t \in s_{l,t}$, the listener signals a success, and the speaker and the listener delete all the words in their lists, except b_t.

　Case 2 If $b_t \notin s_{l,t}$, the listener signals a failure. The listener and the speaker both add b_t to their list.

R5 State Change. On any time t, only the speaker and listener for that time point change their list according to **R4**, and all other agents maintain the list they had after the previous time point, $t - 1$.

3   Formal Results

In this section we will provide theorems concerning the limiting behavior of the naming game on a digraph. We will show that the naming game algorithm described by the five rules in 2.3 leads to a homogeneous Markov chain whose absorbing states correspond exactly to the set of solutions to the game, namely system states where each agent has the same single word for the object. Conditions on the digraph that guarantee absorption in one of these states with probability 1 will be given, and also the limiting behavior of digraphs where convergence to a solution state is not guaranteed will be presented. We assume only an understanding of finite state Markov chains as in [12].

3.1   Definitions

The convergence theorems in the next section require some fairly standard graph theoretic definitions as follows.

Definition 1. *Let* $\mathcal{D} = \langle A, V \rangle$ *be a directed graph and* $a, b \in A$. *Then there is a* path *from* a *to* b *in case either* $a = b$ *or there are nodes* $\{a_i \in A | \exists n \geq 2, \forall 1 \leq i \leq n\}$ *such that* $a_1 = a$; $a_n = b$; $\forall i = 1, \ldots, n-1$, $(a_i, a_{i+1}) \in V$.

Definition 2. *Let* $\mathcal{D} = \langle A, V \rangle$ *be a directed graph. The* path relation $P \subseteq A \times A$ *is defined by* $\forall a, b \in A$, $aPb \Leftrightarrow$ *there is a path from* a *to* b, *and the* path digraph *of* $\mathcal{D}$ *is denoted by* $\mathcal{P}_{\mathcal{D}} = \langle A, P \rangle$

Definition 3. *Let* $\mathcal{D} = \langle A, V \rangle$ *be a directed graph, and* P *be the path relation on* A. *Then*

  (i) $\mathcal{D}$ *is* strongly connected *in the case* $P = A \times A$.
 (ii) $\mathcal{D}$ *is* quasi strongly connected *in the case* $\forall a, b \in A$, $\exists c \in A$ *s.t.* $cPa \wedge cPb$.
(iii) $\mathcal{D}$ *is* weakly connected *in the case the digraph* $\mathcal{D}^* = \langle A, V \cup V^* \rangle$ *is strongly connected, where* $V^* = \{(b, a) | (a, b) \in V\}$.

It is easy to show that (i) implies (ii), and (ii) implies (iii).

Definition 4. *Let* $\mathcal{D} = \langle A, V \rangle$ *be a directed graph and* P *be the path relation on* A. *Then*

 (i) *A node* $a \in A$ *is a* dominant node *in case* $\forall b \in A$, aPb.
(ii) *A node* $a \in A$ *is a* maximal node *in case* $\forall b \in A$, $bPa \Rightarrow aPb$.

3.2   Convergence Theorems

In the interests of brevity, only the key ideas in the proofs will be mentioned.

Theorem 1. *Let* $\mathcal{D} = \langle A, V \rangle$ *be a digraph subject to the five rules of the naming game in Sec. 2.3. Then the resulting dynamic system can be described by a homogeneous Markov chain with state space* $S = \{\mathbf{s} | \mathbf{s} \in X_{i=1}^{N} \wp(W)\}$.

Proof. The key idea is that for any time point t, only the lists of the speaker and listener can change, and list changes are time invariant and independent of past states.   □

Theorem 2. *If* $\mathcal{D} = \langle A, V \rangle$, *is weakly connected then the absorbing states are exactly* $T = \{\mathbf{s} | \exists w \in W \text{ s.t. } \mathbf{s} = X_{i=1}^{N} \{w\}\}$

Proof. Suppose $\mathbf{s} \notin T$ is an absorbing state of the system. Then, if $\mathcal{D}$ is weakly connected there must be two agents, $a_i, a_j \in A$ with different lists in $\mathbf{s}$ and $(a_i, a_j) \in V$. There is a positive probability that at any time t this pair is selected and their lists change, thus $\mathbf{s}$ can't be an absorbing state.   □

Theorem 3. *Let $\mathcal{D} = \langle A, V \rangle$ be a weakly connected digraph subject to the rules of the naming game. Then the dynamic system converges to one of the absorbing states with probability one if and only if it is quasi strongly connected.*

Proof. Suppose $\mathcal{D}$ is quasi strongly connected. The proof of convergence to an absorbing state with probability one has two steps: (i) show that quasi-strongly connected digraphs have at least one dominant node, and (ii) show that at any time t there is a positive probability of a sequence of events starting with a dominant node that will result in a transition to one of the absorbing states.

(i) First select any two distinct nodes $a, b \in A$, and find a node $c \in A$ with a path to both. Then pick a new node $d \in A$, and find a node $e \in A$ with a path to both c and d. Continue this process recursively until all nodes are selected, and the final selected node is a dominant node with paths to all other nodes.

(ii) Suppose $d \in A$ is a dominant node and select paths from d to each other node in the graph and concatenate all the arcs in all the paths in a sequence $dP'a_1 dP'a_2 \ldots dP'a_n$, where $dP'a_i$ is a sequence of arcs forming the path from d to a_i. Suppose there are L arcs in this sequence. Now at any time point t there is a positive probability $(1/|V|)^{2L}$ that each arc in the sequence starting from left to right is selected twice in succession. Next we show that such a sequence of selections will result in an absorbing state with a positive probability that exceeds a positive lower bound. Suppose at any time t the first arc in the first path, say (d, a_i) is selected twice. Whether or not the first communicated word $b_t \in W$ was a success, going into the second communication both agents have b_t on their lists, and there is a positive probability no smaller than $(1/M)$ that the speaker d on time $t + 1$ will speak b_t to the listener. If this happens, the lists of both agents d and a_i will be the singleton set $\{b_t\}$. Following this event, each successive speaker's list will consist of the single word $\{b_t\}$, and on that speaker's second consecutive communication the listener's list will also be $\{b_t\}$. Thus with probability no smaller than $p = (V^{2L} \cdot M)^{-1}$ all agents' lists will become $\{b_t\}$, and this corresponds to one of the absorbing states. Finally, since this sequence can occur starting on any time t, and it has positive probability, convergence of the chain to an absorbing state will occur with probability 1 since there is a geometric distribution of the trial to a successful occurrence of the aforementioned sequence of events.

Suppose $\mathcal{D}$ is not quasi strongly connected. The proof involves two steps: (i) show that there are at least two disjoint, strongly connected components of maximal nodes, M_1, M_2, with no paths between them, and (ii) show that there is positive probability that the elements of M_1 and of M_2 will absorb, respectively, with singleton lists with different words.

(i) Since $\mathcal{D}$ is not quasi strongly connected, there exist two distinct nodes i and j for which there is no $k \in A$ with $kPi \wedge kPj$. Set $C_i = \{k | k \neq i, \ kPi\}$ and $C_j = \{k | k \neq j, \ kPj\}$. Note that if $C_i = \emptyset$ ($C_j = \emptyset$) then set $M_1 = \{i\}$ (set $M_2 = \{j\}$) as one of the desired sets of maximal elements described above. Suppose $C_i \neq \emptyset$ and examine all the paths to node i that do not repeat nodes, and select a node $i^\star \in C_i$ which starts one of the longest of these paths. Clearly $i^\star$ is a maximal node because otherwise it would not begin a path with a maximal

number of non-repeating nodes. Next set $M_1 = \{k|\, i^\star Pk \wedge kPi^\star\}$ which is easily seen to consist of maximal elements. Similarly, if $C_j \neq \emptyset$ find a maximal element $j^\star \in C_j$, and set $M_2 = \{k|\, j^\star Pk \wedge kPj^\star\}$. Clearly M_1 and M_2 are disjoint, strongly connected, and there are no paths between them.

(ii) Select a path from $i^\star \in M_1$ that includes every element in M_1 with no repeats. Using an argument as in the first part of the proof, there is a positive probability that $i^\star$ can initiate a word and pass it along as a singleton list to all nodes in M_1, and further if this happens no further change in the lists of the nodes in M_1 are possible. Independently, there is a positive probability that a similar sequence of events can lead to all the members of M_2 absorbing on lists of a single word different from that of the nodes in M_1. If these two sequences occur, then the system will never reach one of the absorbing states described in Theorem 2. □

Theorem 3 provides necessary and sufficient conditions for the naming game on a weakly connected digraph to guarantee convergence to an absorbing state. In case the digraph is quasi strongly connected there are of course many other sequences of events that will lead to absorption much more quickly than the one described in the proof. However, because of the geometric distribution of the trial to absorb in the proof, we know absorption of the chain occurs with finite mean absorption time no larger than $p^{-1} = (V^{2L} \cdot M)$. If instead the digraph is not quasi strongly connected, then strongly connected components of maximal elements like M_1 and M_2 will absorb on singleton lists, but not necessarily with the same word. However in this case other nodes in the digraph with paths from two or more of these strongly connected components will keep switching their word list as illustrated in the next section by the digraph in Fig. 1b.

4 Computer Results

4.1 Three Nodes' Directed Graphs

To evaluate how agents reach agreement on an object-word assignment, we considered all thirteen possible weakly connected digraphs on three nodes equivalent under node permutation. We bounded the word set to $M = 5$ words, and ran our algorithm 1000 times for each of these structures. Our goal was to see if convergence occurred, and if so how long it took on the average.

We do not report all the results here; however, we noted that all the quasi-strongly connected digraphs converged to a solution as expected from Theorem 3. Convergence occurred between 2 and 139 iterations. The single weakly connected digraph that was not quasi-strongly connected is depicted in Fig. 1a. It converged to a solution only 195 times out of 1000 runs. In this graph, only two agents can be speakers and they speak only to the third agent. These two potential speakers are maximal elements, but neither is a dominating node. Thus convergence will occur only if both speakers happen to select the same word as their first proposal, and this occurs in theory with probability 1/5, approximately as found.

Fig. 1. (a) The three node weakly connected graph, which is not quasi strongly connected. (b) Nontrivial weakly connected digraph, which is not quasi strongly connected.

4.2 A Weakly Connected Digraph with Subcomponent Convergence

Figure 1b shows a weakly connected digraph that is not quasi-strongly connected. A priori, we expected that agents 1 and 3 and also agents 2 and 5 would converge to a common word; however, that word might be different. On the other hand we expected that node 4 might oscillate between the two words possessed by the two converging components of the digraph. We ran our algorithm on this graph 1000 times, and we set $M = 5000$, in this way the probability of two agents "inventing" the same word is negligible. We observed that no algorithm run converged to an agreement. We define the conditional probability that the interaction is a success, conditioned over the selection of a particular edge at time $t+1$, $\Pr(succ.|v_{t+1}) = |s_{s,t+1} \cap s_{l,t+1}|/|s_{s,t+1}|$. This probability can then be computed over all the edges in the graph, $\Pr(succ.) = \sum_{v_t \in |V|} \Pr(succ.|v_{t+1})/|V|$. At each time t, we measure the mean probability $\Pr(succ.|v_{t+1})$ over all the algorithm runs. For the edges $(1,3)$ and $(2,5)$, and the mean probability $\Pr(succ.)$ over all runs. As expected the mean of $\Pr(succ.)$ saturates below 1, because the algorithm did not converge to an agreement in any algorithm run. Contrarily, $\Pr(succ.|(1,3))$ and $\Pr(succ.|(2,5))$ converge to one, which shows us that we get stable local agreement.
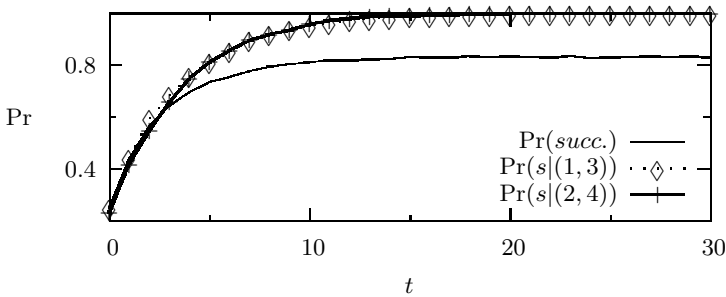
Fig. 2. Evolution of three mean probabilities of success over 1000 algorithm runs on the simplest nontrivial quasi strongly connected graph, $\Pr(succ.|(1,3))$, $\Pr(succ.|(2,5))$, $\Pr(succ.)$

5   Conclusions

We presented necessary and sufficient conditions and empirical evidence that depending on the structural properties of the communication network tell us if

the naming game will converge to a naming convention with probability one or if it has a non negligible probability to cycle trough states where a collective agreement is never reached.

In future work, we plan to develop more concrete assumptions in the naming game. We want to simulate agents with realistic memory assumptions and more reasonable selection rules. Also we want to allow speakers to communicate with multiple listeners, concerning multiple objects with possible structure.

References

1. Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., Steels, L.: Sharp transition towards shared vocabularies in multi-agent systems. J. Stat. Mech., P06014 (2006)
2. Steels, L.: Self-organizing vocabularies. In: Langton, C.G., Shimohara, K. (eds.) Artificial Life V, Nara, Japan, pp. 179–184 (1996)
3. Lu, Q., Korniss, G., Szymanski, B.: The Naming Game in social networks: community formation and consensus engineering. Journal of Economic Interaction and Coordination 4(2), 221–235 (2009)
4. Nowak, M.A., Plotkin, J.B., Krakauer, D.: The evolutionary language game. Journal of Theoretical Biology 200(2), 147–162 (1999)
5. Avesani, P., Agostini, A.: A peer-to-peer advertising game. In: Orlowska, M.E., Weerawarana, S., Papazoglou, M.P., Yang, J. (eds.) ICSOC 2003. LNCS, vol. 2910, pp. 28–42. Springer, Heidelberg (2003)
6. Liu, R., Jia, C., Yang, H., Wang, B.: Naming game on small-world networks with geographical effects. Physica A: Statistical Mechanics and its Applications 388(17), 3615–3620 (2009)
7. Nardini, C., Kozma, B., Barrat, A.: Whos talking first? Consensus or lack thereof in coevolving opinion formation models. Physical Review Letters 100(15), 158701 (2008)
8. Brigatti, E., Roditi, I.: Conventions spreading in open-ended systems. New Journal of Physics 11, 023018 (2009)
9. Tang, C., Lin, B., Wang, W., Hu, M., Wang, B.: Role of connectivity-induced weighted words in language games. Physical Review E 75(2), 27101 (2007)
10. Barrat, A., Baronchelli, A., DallAsta, L., Loreto, V.: Agreement dynamics on interaction networks with diverse topologies. Chaos: An Interdisciplinary Journal of Nonlinear Science 17, 026111 (2007)
11. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. Journal of Mathematical Psychology 51(6), 359–382 (2007)
12. Ross, S.: Introduction to probability models, 9th edn. Academic Pr., London (2007)

Synthesis and Refinement of Detailed Subnetworks in a Social Contact Network for Epidemic Simulations

Huadong Xia, Jiangzhuo Chen,
Madhav Marathe, and Henning S. Mortveit

Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA

Abstract. In recent years, individual based epidemic simulations on synthetic social contact networks have been proved useful in supporting public health epidemiology. In use of such models, it is often desirable to understand how disease spreads through a small subset of the population; furthermore often data is available that allows us to develop more refined representations of social contact networks that span these sub-populations. Examples include: schools, office campuses, military bases, etc.

Here, we develop a multi-level refinement and analysis technique to synthesize and embed subnetworks of an existing contact network when more detailed activity data about a small community is available. The methodology is illustrated by embedding detailed high school networks in a synthetic social contact network of the New River Valley, Virginia.

Keywords: sensitivity analysis, network refinement, epidemic simulation, social network synthesis.

1 Introduction

Network-based models of computational epidemiology are being increasingly used as decision support tools by policy analysts [4,5,9]. In contrast to aggregate differential equation based models [7], these network models describe diffusion of an infectious disease over a synthetic social contact network, which is constructed from individual level synthetic demographic and activity data. In a contact network, nodes represent individuals, and edges represent physical proximity. The usefulness of such network-based models relies in a large part on the *quality* and the *realism* of the social contact networks.

There are two approaches for constructing *realistic* synthetic social contact networks. The first approach is based on using explicit measurements to infer edges in the network. For example, in [8,10], electronic devices were used to record people's physical proximity in order to construct a realistic social contact network. However, these methods are typically limited to inferring small networks due to the fact that they are extremely time consuming and costly.

The second approach constructs a synthetic social contact network for a large population by combining diverse data sets, including census, land use, information on built infrastructure, activity patterns, etc. Dynamic networks are then synthesized using these data sets in conjunction with well established social and behavioral theories; see [1]. These synthetic social networks are statistically similar to realistic social contact networks. A number of situations require a more refined representation of a small subnetwork of the underlying synthetic social network. For example, consider the situation when one would like to understand the spread of infections on a military base, a college campus, a large health care facility or an office campus. In such situations, a more refined representation of these subnetworks is desirable.

Our contributions. In this paper, we present a methodology to synthesize and embed subnetworks in an existing social contact network. The subnetworks are constructed using detailed data about the subpopulation, their demographics and their activity structure. We illustrate our methodology by embedding detailed representations of high schools in a synthetic social contact network of New River Valley (NRV), Virginia. We will use NRV_n to denote the original network and NRV_r to represent the refined network. Using fast epidemic simulations over NRV_n and NRV_r, we study the situations when such refinements affect the epidemic dynamics. For flu like diseases, our results suggest that the dynamical measures are insensitive to in-class (sublocation) contact models but are sensitive to in-school and between-school social contact models. The basic methodology is general and can be applied to other types of *campus locations*, such as a college campus, a military base, a health care facility, etc. Specific contributions of our paper include:

- First, we propose a methodology to refine an existing social contact network by replacing normative networks representing social interactions in a campus by subnetworks based on more realistic activity data pertaining to individuals and locations on the campus. The methodology is modular and comprises of multiple levels of refinement, including sublocation (class), location (school), and between-location levels. Analyst can choose a model or synthesize a network based on specific measurements for each of the levels. In our case study on the NRV network, we use real activity data. But our methodology is applicable with synthetic activity data too. This is helpful when more realistic (although not real) activity data become available.
- Second, using the NRV network, we show that refinement of even a small fraction of the network *can have significant impact*. More generally, we present a methodology to study whether the refinement over the existing network is significant, in terms of its influence on the epidemic dynamics; and to identify the type of refinements that are likely to matter more than others. This methodology is based on simulations.
- Finally, we present a systematic methodology for sensitivity analysis and uncertainty quantification as it pertains to the subnetwork refinement process. We stress that network structure sensitivity is different than classical normal parameter sensitivity analysis. A detailed sensitivity analysis involves

executing a large factorial experiment design involving thousands of simulation runs. In this work we have used *EpiFast*, an agent-based high performance epidemic simulation system [2].

2   Methods for Subnetwork Construction and Embedding

In this section, we first describe our general methodology to construct a social contact network and apply the methodology to generate a contact network (NRV_n) for the New River Valley region of Virginia based on synthetic population data; and contact networks for the three high schools in NRV region, based on real population data. Next we describe embedding of the three high school networks into NRV_n to improve its quality.

2.1   Contact Network Construction

People in the society perform their own activities everyday, including work, study, shopping, etc. These activities move people between locations. We divide locations into sublocations. For example, a school is a location while classrooms are sublocations in a school. Contacts between people may occur if they stay at the same sublocation at the same time. Activities of people can be represented by a series of time-labeled people-location bipartite graphs G_{PL} [4]. Each G_{PL} corresponds to a time interval during which no one moves.

Next we project each graph G_{PL} into a people-people graph G_{PP}. Since contacts only happen between people who are co-located, we group people in the same sublocation together. These are the people who are within distance of two to each other in G_{PL}.

To construct the contact network within a sublocation, we apply one of several graph models, called *sublocation models*, including complete graph, $G(n, p)$ graph, and random geometric graph. For example, if we use the $G(n, p)$ graph model, then any two people in the same sublocation at the same time will be connected with probability p. We are interested in whether there is significant difference between different sublocation models and between different parameter settings of the same model (e.g. p in $G(n, p)$). This is studied in Section 3.

We put the sublocation contact networks together to form the people-people contact network G_{PP} for the time interval. In G_{PP}, each edge has a weight equal to the duration of the corresponding time interval.

Finally we aggregate the G_{PP} graphs for all time intervals to form the long-term G_{PP} (contact network for the whole day). The construction process is illustrated in Figure 1 of the full paper [11] (omitted here due to space limit).

Using the above methodology, we first construct a synthetic social contact network for the NRV region, based on the synthetic population data, including demographics and activities, that our group generated from the US 2000 census and activity survey data. Details about the synthetic population construction for a US region can be found in [1]. The synthetic NRV contact network contains a subnetwork for each of the three high schools in the NRV region. We also

construct contact networks for the three schools based on real in-school activity data. The differences between the subnetworks within the NRV_n network and the high school networks created separately include the following:

1. The activities used to construct the whole NRV_n network are synthesized and coarse; while the three high school networks are based on real class schedules of all students (with their ID's anonymized). In the former case, most high school students have only two classes, one for the whole morning and one for the whole afternoon. In the latter case, the students have more classes and shorter class durations. Therefore, the students are mixed more fully in the latter case; between them there are more edges with smaller weights.
2. In the synthetic subnetworks, the sublocation model is complete graph. That is, any two students taking the same class are connected. In the "real" school networks, we use various sublocation models including $G(n, p)$ random graph and random geometric graph.

The real class schedules of the students in the three NRV high schools are collected from the 2009 school registration data. Detailed statistics of the real NRV high school class schedules can be found in [11]. We highlight that the students in the three high schools are only 1.7% of the whole NRV population. It is worth mentioning that to our best knowledge, it is the first time the real class schedule is used to capture student activities. Works in [6] constructed school network in a similar way, but they used artificial data.

2.2   Subnetwork Embedding

We would like to refine the synthetic NRV contact network with the three more realistic high school contact networks. To this end we need to embed the three high school networks into the whole NRV_n network. The embedding involves three steps as illustrated in Figure 1.

(a) **Identifying and mapping.** First, identify synthetic high school subpopulation from the global synthetic population. Unfortunately the synthetic high school subpopulation is smaller than the real high school subpopulation.[1] We have to pick extra people from the NRV synthetic population to fill the gap. This process is illustrated in Figure 1(a).
(b) **Removing school edges.** Delete from the whole NRV_n network all edges among the identified synthetic subpopulation that are due to in-school activities.[2] We keep edges that are generated due to non-school activities.
(c) **Embedding.** Embed edges in the real high school networks into the global contact network based on the vertex mapping in the first step.

---

[1] The inconsistency is mainly due to the fact that NRV high school students increased about 10% from the year 2000 (when the census data was obtained) to the year 2009 (when the class schedule data was collected).

[2] Each edge in the contact network also has an "activity type" label. It is used to identify school type edges between people in the identified high school subpopulation.
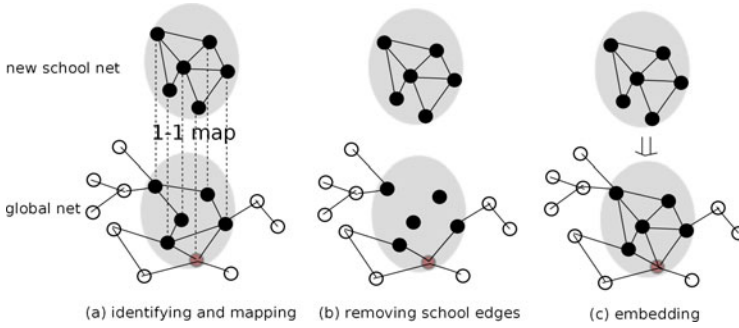
Fig. 1. Procedures for subnetwork embedding. Real subpopulation is larger than the synthetic subpopulation; we pick the adjacent gray node to fill the gap.

3   Epidemic Dynamics in the Refined Contact Network

For our subnetwork synthesizing and embedding methodology, we are interested in two questions: (1) Does the refinement make much difference, in terms of network dynamics? (2) Which parameters is our methodology sensitive to? We present methodologies for exploring these questions, consisting of epidemic simulations on the original and refined contact networks and statistical analyses on the simulation results. In Section 3.1, we show that although the refinement changes only a very small part of the whole network and does not seem to have a noticeable impact on the overall network structure, the epidemic dynamics of the whole population does change significantly. In Section 3.2, we conduct a series of sensitivity tests to quantify uncertainties in our methodology.

Simulations. We run EpiFast [2] to simulate a strong flu-like infectious disease on the NRV synthetic contact network and various refined networks. The disease has an *attack rate* (fraction of the population infected with the disease during an epidemic) of about 30% without any intervention. For each network the simulation has 30 replicates and in each replicate we allow the disease to propagate for 300 days. We characterize the epidemic dynamics by the average over 30 simulation replicates.

3.1   When and Which Details Matter

We compare two contact networks: the original synthetic NRV network (NRV_n) and the refined NRV network (NRV_r) with real class schedule data and complete graph sublocation model. Since the synthetic network also uses complete graph sublocation model, the only source of difference between two networks is the high school students' class schedules. Does this difference have any significant impact on the network dynamics? We find that network structural analysis may not be able to give appropriate answer to this question. Instead, we introduce a simulation methodology. We run epidemic simulations on both networks with the

same configuration. The epidemic dynamics over time can be characterized by an average epidemic curve, which represents the average number of new exposures on each day.

We first compute the distributions of node degree and contact duration in both networks. We find that the distributions are very different between in-school subnetworks: a *real* school network has a much larger average degree, because students mix more due to frequent class changes; and much shorter contact durations, due to shorter classes, than the corresponding synthetic school network. The degree distributions are almost identical, however, for the whole synthetic NRV network and the refined NRV network, as shown in Figure 2. This is probably because the involved high school subpopulation is only a small fraction (1.7%) of the whole NRV population. More structural analysis results can be found in the full version of this paper [11].

By running epidemic simulations on NRV_n and NRV_r, we find that they have significantly different dynamics. In Figure 3, the synthetic NRV network has an earlier outbreak and a higher peak. *This implies that the details of a small part of a contact network can have a large impact on the dynamics of the whole network.* It means that our refinement on the synthetic contact network does make a difference.
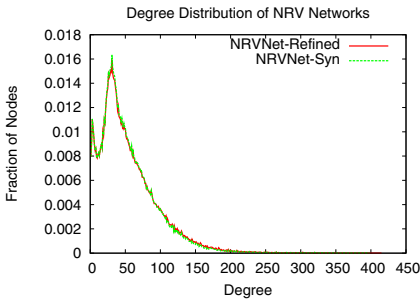
Fig. 2. Global networks (synthetic and refined) have similar degree distributions
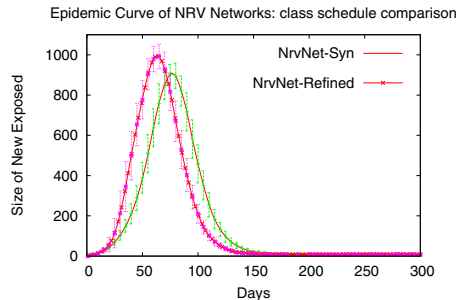
Fig. 3. Global networks (synthetic and refined) have different epidemic dynamics

3.2   Sensitivity Test

We consider two types of sensitivity: (i) numerical sensitivity that measures the effect of random seed on the structural properties of the resulting network and epidemic dynamics on the resulting network and (ii) network structure sensitivity that tries to capture the effect of structural variation of the network on the epidemic dynamics. The first one is standard in the literature; we study the second one by using a *Switch Markov Chain* to vary the structure of the contact network and evaluate the changes on the epidemic dynamics [11].

We study the following questions: (1) How sensitive is the refined network to the in-class network construction model? (2) How sensitive is the refined network

to the choice of the 1-1 map? We characterize the epidemic dynamics by three response variables: (`attack rate, peak, peak day`) and consider three factors: (*sublocation model: L, connectivity parameter: C* and *switching operations: S*). We have two levels for factor L: $G(n, p)$ and geometric random graph; three levels for factor C: $\{0.2, 0.5, 0.8\}$ for p in $G(n, p)$.[3] For factor S we have 4 levels: initially we map each real student to a synthetic student in the same school; then we perform *switch Markov chain* [3] to randomize the 1-1 map. Switching is either within class, within school, between certain two schools, and between all three schools. We now have a $2 \times 3 \times 4$ design and generate 20 network instances for each of the 24 cells. We run epidemic simulations for 30 replicates on each of the 480 networks and compute averages over replicates. For each of the response variables we perform a 3-factor ANOVA analysis.

Results in Table 1 indicate that the specific choice of sublocation model does not seem to be a significant factor when studying flu like diseases. The connectivity parameter and switching operations, however, are significant factors. To precisely measure the sensitivity of the two factors across each level, we further conduct multiple comparisons for each level pair through Tukey Tests. Analyses show that the epidemic dynamics when edges are flipped within a school is not significantly changed so far as the overall density of the network is kept invariant. Detailed results can be found in [11].

Table 1. Multi-factor ANOVA: * shows that the F-test is significant at 99% confidence level

factor	attack rate			peak			peak day		
	DF	SS	F value	DF	SS	F value	DF	SS	F value
L	1	0.0000004	1.0389	1	24	0.3797	1	0.1	0.0327
C	2	0.0088561	10628.4552*	2	498078	3976.3839*	2	13482.1	4236.0993*
S	3	0.0019159	1532.8931*	3	73668	392.0810*	3	2591.7	542.8744*
$L : C$	2	0.0000013	1.5858	2	87	0.6981	2	2.1	0.6690
$L : S$	3	0.0000019	1.5305	3	500	2.6622	3	6.1	1.2721
$C : S$	6	0.0011373	454.9766*	6	150700	401.0359*	6	172.0	18.0094*
$L : C : S$	6	0.0000041	1.6268	6	329	0.8745	6	12.8	1.3428
Residuals	456	0.0001900		456	28559		456	725.7	

---

[3] Factor values are chosen for the geometric random graph model accordingly, so that the constructed school networks have the similar density as those constructed with $G(n, p)$.

References

1. Barrett, C., Beckman, R., Khan, M., Kumar, V.A., Marathe, M., Stretz, P., Dutta, T., Lewis, B.: Generation and analysis of large synthetic social contact networks. In: Proceedings of the 2009 Winter Simulation Conference, pp. 1003–1014 (2009)
2. Bisset, K.R., Chen, J., Feng, X., Kumar, V.A., Marathe, M.V.: EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In: Proceedings of the 23rd International Conference on Supercomputing (ICS), pp. 430–439 (2009)
3. Cooper, C., Dyer, M., Handley, A.J.: The flip markov chain and a randomising P2P protocol. In: Proceedings of the 28th ACM Symposium on Principles of Distributed Computing (PODC), pp. 141–150 (2009)
4. Eubank, S., Guclu, H., Kumar, A.V.S., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. Nature 429(6988), 180–184 (2004)
5. Ferguson, N.M., Keeling, M.J., Edmunds, W.J., Gani, R., Grenfell, B.T., Anderson, R.M., Leach, S.: Planning for smallpox outbreaks. Nature 425(6959), 681–685 (2003)
6. Glass, R.J., Glass, L.M., Beyeler, W.E., Min, H.J.: Targeted social distancing design for pandemic influenza. Emerging Infectious Diseases 22(11) (2006)
7. Hethcote, H.W.: The mathematics of infectious diseases. SIAM Review 42, 599–653 (2000)
8. Hlady, C., Curtis, D., Severson, M., Fries, J., Pemmaraju, S., Segre, A., Herman, T., Polgreen, P.: A near-real-time method for discovering healthcare worker social networks via wireless devices. In: 47th Annual Meeting of the Infectious Disease Society of America (October 2009)
9. Longini, I.M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D.A.T., Halloran, M.E.: Containing pandemic influenza at the source. Science 309(5737), 1083–1087 (2005)
10. Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S., Chong, S.: CRAWDAD data set ncsu/mobilitymodels (v. 2009-07-23) (July 2009)
11. Xia, H., Chen, J., Marathe, M.V., Mortveit, H.S.: Synthesis and embedding of subnetworks for individual-based epidemic models. Technical Report 10-139, NDSSL, Virginia Tech. (2010), http://staff.vbi.vt.edu/chenj/pub/NDSSL-TR-10-139.pdf

Technosocial Predictive Analytics for Illicit Nuclear Trafficking

Antonio Sanfilippo, Scott Butner, Andrew Cowell, Angela Dalton, Jereme Haack,
Sean Kreyling, Rick Riensche, Amanda White, and Paul Whitney

Pacific Northwest National Laboratory, Richland, WA
{antonio,scott.butner,andrew,angela.dalton,jereme,
sean.kreyling,rmr,amanda,paul.whitney}@pnl.gov

Abstract. Illicit nuclear trafficking networks are a national security threat. These networks can directly lead to nuclear proliferation, as state or non-state actors attempt to identify and acquire nuclear weapons-related expertise, technologies, components, and materials. The ability to characterize and anticipate the key nodes, transit routes, and exchange mechanisms associated with these networks is essential to influence, disrupt, interdict or destroy the function of the networks and their processes. The complexities inherent to the characterization and anticipation of illicit nuclear trafficking networks requires that a variety of modeling and knowledge technologies be jointly harnessed to construct an effective analytical and decision making workflow in which specific case studies can be built in reasonable time and with realistic effort. In this paper, we explore a solution to this challenge that integrates evidentiary and dynamic modeling with knowledge management and analytical gaming, and demonstrate its application to a geopolitical region at risk.

Keywords: Illicit trafficking, nuclear proliferation, predictive analytics, modeling, knowledge management, analytical gaming, decision making.

1   Introduction

The illicit trafficking of nuclear materials is one of the greatest threats to world security. Although relatively few incidents related to the theft or loss of nuclear and radioactive materials have been documented during the last two decades, with only a handful involving weapon-grade fuel [1], the potential threat is enormous. A single successful attempt to create a nuclear smuggling network, such as the international proliferation network orchestrated by A. Q. Khan from the 1970s through the 1990s, can have highly destabilizing effects on several geopolitical regions at once, with disastrous consequences for international security. The ability to reason proactively about the formation of illicit nuclear trafficking networks is therefore crucial to anticipate strategic surprises, such as North Korea's sudden launch of its *Taepo-Dong* intercontinental ballistic missile over Japan on August 31, 1998.

So far, most of the computer modeling work on anticipating and interdicting illicit nuclear trafficking has focused on radiation detection and to a lesser extent on methods of how to deploy these detectors. This work has made a valuable contribution to the

development and effective use of technologies such as radiation portal monitors that have greatly improved the nation's ability to prevent nuclear weapons from crossing our borders. However, in order to understand illicit nuclear trafficking as a complex system to be able to inform preemptive action, we need to

1. assess the propensity of state and non-state actors to proliferate within a geopolitical region of interest
2. anticipate how actors at risk of engaging in proliferation strategy may behave in order to achieve their objectives.

In this paper, we show how these two tasks can be addressed in a systematic, effective and efficient manner using a Technosocial Predictive Analytics (TPA) approach that integrates probabilistic evidentiary reasoning, knowledge management, analytical gaming and dynamic modeling. We describe an application of this approach to a geopolitical region at risk that includes six countries from Sub-Saharan Africa, and outline the workflow process associated with the modeling, knowledge and gaming technologies utilized.

2   Background

The use of computer modeling to anticipate and interdict illicit nuclear trafficking has mainly focused on radiation detection. For example, [2] describe a radiation detection approach that uses correlation analysis to identify isotopes from spectra, obtained via short measurement times with relatively small detector crystals, with reference to a library of known isotope spectra. This method can be used to develop handheld radiation devices that are easily deployed in high volumes in the field for isotope identification purposes to detect illicit trafficking of nuclear materials and radioactive sources.

Some effort has also been devoted to using modeling to help guide the placement of radiation detectors. For example, [3] describe a stochastic interdiction model relative to a transportation network that estimates detection probabilities for a specific set of parameters (e.g. material being sensed, geometric attenuation, shielding, cargo and container type, and time allotted for sensing). Such a model can help increase effective deployment of portal detector installations, which the Department of Homeland Security and the US Department of Energy have been pursuing in the US and overseas to reduce the risk of illicit trafficking of nuclear material through international airports, seaports and border crossings.

Modeling work in radiation detection and the deployment of radiation detector devices has made a valuable contribution to the development and effective use of technologies such as radiation portal monitors and handheld radiation detectors that have greatly improved the nation's ability to prevent nuclear weapons from crossing our borders [4]. However, the approach pursued is primarily tactical in nature as it addresses detection and interdiction of an ongoing act of illicit nuclear trafficking. If we wish to preempt illicit nuclear trafficking, a more strategic approach is needed in which we can assess the propensity of state and non-state actors to proliferate and for those actors at risk anticipate the networks they might create to acquire the know-how, materials and technology necessary to develop nuclear weapon capabilities.

Some exploratory work has been done on modeling proliferation propensity using social, economic and political factors. For example, [5] describe a probabilistic evidentiary model based on Dempster-Shafer theory that assesses the likelihood that a country's security culture to foster nonproliferation as a function of factors related to trade and geography, industrial infrastructure, exports trends, export control and commitment to strengthen nonproliferation credentials (e.g. adherence to international nuclear protocols and treaties). [6] describe a model that assesses the propensity of a state to engage in nuclear weapon proliferation through a Bayesian network analysis of social factors. Our objective in this paper is to extend the reach of these more strategic approaches to develop a software platform for illicit trafficking capable of supporting anticipatory analysis and preemptive decision-making.

3 Approach

We use the Technosocial Predictive Analytics framework (TPA) [7, 8] to extend the reach of previous work on modeling proliferation propensity [5, 6]. TPA help analyst and policymakers manage strategic surprises and opportunities through the exploration and manipulation of plausible futures. This objective is achieved through the development and integration of methods and capabilities for analysis and decision making that

- Enable reasoning with models that combine human and physical factors (*Technosocial Modeling*)
- Facilitate the acquisition of knowledge inputs (*Knowledge Encapsulation Framework*)
- Enhance the user's experience in the modeling and decision making tasks (*Analytical Gaming*).

Technosocial Modeling (TM) The TM component supports the development of models that combine human and physical factors. It enables rigorous model validation through quantitative comparison of dynamic models to observations [9, 10], formal model calibration through aggregation of subject matter expertise [10, 11], and automatic support for the integration of diverse modeling frameworks such as system dynamics (SD) and Bayesian nets (BN) models [12].

Knowledge Encapsulation Framework (KEF) KEF [13, 14] allows users to collect, store, share, and annotate heterogeneous information objects (e.g. text, speech, images, video), and it provides the means to link these to computational models. KEF is capable of learning which information objects are of interest and mine online sources for new and related information across both traditional media (e.g. reports, newswires, and academic articles) and social media such as blogs, wikis and web forums. KEF is implemented within a semantic wiki environment, which provides a distributed and collaborative working framework. Models can be imported into KEF and evidence found within the wiki content can be interactively and collaboratively aligned with the model structure.

Analytical Gaming (AG) AG [15, 16] provides an environment in which analysts and decision makers can engage in interactive role-play to critique each other's ideas and action plans in order to achieve preparedness in real-world situations. AG facilitates

creation and execution of games analogous to traditional tabletop simulation exercises. One application of the AG approach is to generate virtual evidence, by recording the behaviors of players, for calibrating model parameters in the absence or sparseness of real-world evidence. AG may also be configured as a collaborative and interactive interface to computer models [16].

3.1 Integrating TPA Components to Support Analysis and Decision Making in Anticipating and Countering Illicit Nuclear Trafficking

Our TPA-based workflow and integration strategy comprises four major aspects:

1. Combine social models of proliferation with physical models of nuclear fuel cycle detection
2. Create active input-output links between modeling and knowledge management processes and capabilities
3. Use role-play to generate data on how actors at risk of engaging in proliferation strategy may behave in order to achieve their objectives.
4. Use role-play to let human interface with models.

Figure 1 provides a graph view of the Bayesian net model we have developed to measure the propensity of a (non-)state actor to engage in nuclear weapon proliferation. A detailed description of the model and its development is given in [10]. Presently, we shall focus on those elements that are more salient for integration with other TPA components.

The model integrates technical, social, and geopolitical factors. It is initially developed using insights elicited from subject matter experts (SMEs) through interactive communications channeled through the wiki environment within KEF. These interactions
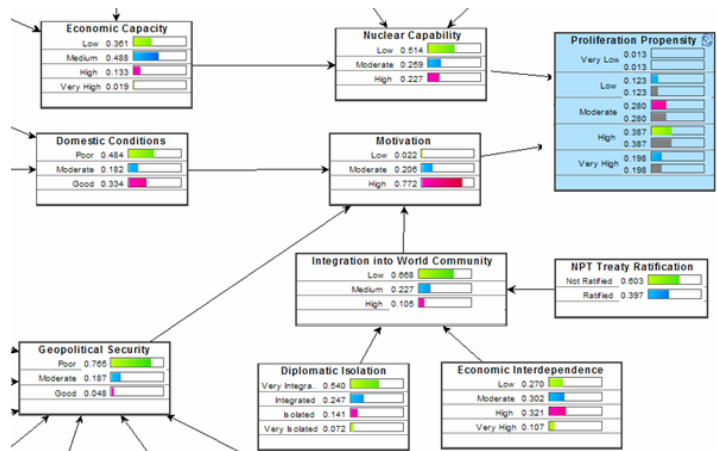
Fig. 1. Bayesian net model of intent to proliferate [10]

are stored together with any additional materials of interest (e.g. model fragments, text, pictures, videos) and linked to model nodes so as to provide transparency and reach-back into the model development process.

The model is then calibrated in two ways: (a) using crowd-sourcing techniques and aggregation algorithms such as conjoint analysis [11] to elicit weights for model parameters from SMEs, and (b) applying the model to ground-truth data and make adjustments to increase goodness of fit between the model and the reference ground

truth data. When real-world evidence is not available, or it is sparse and in need of corroboration, artificial evidence is generated through role-playing in the AG components to facilitate model-calibration (see below).

Once the model is calibrated, it is instantiated with evidence relative to the real-world scenarios of interest, which presently concern the propensity of countries in Sub-Saharan Africa (South Africa, Nigeria, Uganda, Gabon, Senegal, and Ghana) to engage in nuclear proliferation activities. The nodes within the model are associated with queries that a search engine or database front-end can interpret to return relevant documents. For example the model node named `Nuclear Capability of Neighbors` in Figure 1 can be associated with a broad search query such as `(NeighborCountryName) has nuclear capability` where the expression `(NeighborCountryName)` functions as a variable over countries.

The calibrated model with the added queries is then imported into KEF via an xml-based exchange protocol. Upon loading the model, KEF reads the queries associated with model nodes and lets the user instantiate query variables. For example, if we are working on a model of Nigeria as a potential proliferator and wish to retrieve evidence about the nuclear capabilities of Nigeria's neighboring countries, then we would substitute the search variable `(NeighborCountryName)` with country names such as *South Africa*, *Uganda*, *Gabon*, *Senegal*, and *Ghana*.

Queries with instantiated variables are fired by the user and relevant documents are returned for each model node for which a search query was specified. Users quickly inspect the documents collected using tools such as keyword-based in-document search to verify document relevance. Relevant documents are then analyzed using a variety of text mining and content analysis capabilities [13, 14, 17], so that elements of content that provide evidence relevant to model nodes can be easily linked to the model nodes, as shown in Figure 2. When content is chosen as evidence for a model node, an evidence rating dialogue box allows the user to score the content selected for credibility and biases as shown in Figure 2, following the approach described in [10].

KEF can be easily configured to mine documents from a variety of data sources available on the Internet (e.g. Google Scholar, opensource.gov, the CNS Nonproliferation Databases, blogs dedicated to discussions on nuclear proliferation). Evidence is collected from heterogeneous data sources, including newswires, scientific publications, specialized databases and social media.

Fig. 2. Harvesting evidence in KEF

Following this approach, we have used the same abstract model shown in Figure 1 to assess the proliferation propensity of several countries in Sub-Saharan Africa: South Africa, Nigeria, Uganda, Gabon, Senegal, and Ghana (see also [10]). This

assessment is needed to understand which country within a region of interest is most a risk of developing an illicit nuclear trafficking network and should consequently be studied into further details. For example, out of the six Sub-Saharan African countries we considered, South Africa and Nigeria emerged as the most likely to engage in nuclear proliferation activities [10]. Since South Africa has been extensively studied since the 1980s in this regard [18] due to its past research activities of nuclear explosive [19], we decided to focus on Nigeria and develop a model of how Nigeria would go about developing an illicit nuclear network.

Since there is no evidence that Nigeria has started to engage in illicit nuclear network creation activities, we used the AG environment to generate virtual evidence that could then be used to calibrate an illicit nuclear trafficking model. The game was configured as a transaction game where multinational companies, superpowers (e.g. USA and China) and countries in the Sub-Saharan African region interact to sell, acquire or regulate the exchange of materials and capabilities related to the development of a full nuclear fuel cycle. The pursuit of nuclear proliferation was broken down into a series of steps summarizing the major stages in developing nuclear weapons, e.g. acquire uranium ore, computer and fissile core fabrication capabilities, nuclear reactor equipment, and weapon delivery systems. A game master monitors the development of the game and intervenes when players have special requests

Fig. 3. A player's interface to the game

(e.g. perform an activity which is currently not in their scope). Otherwise, all interactions are handled by the AG environment.

Figure 3 provides a view of player's AG application screen half-way through playing a game session. The player represents *Niiland* (a country alias for Nigeria). His/her aim is to acquire assets (lower left in Figure 3) which would enable the construction of nuclear weapons. In carrying out this aim, the player communicates with the other players through instant messaging to (1) use finances and other resources available to the player (e.g. crude oil) to acquire nuclear material and capabilities, and (2) cover his/her real intents to escape interception by controlling actors (e.g. the US). Each player has specific roles in the game and each plays to his/her own advantages in making alliances tactically or strategically, as needed.

Occasionally, in a random fashion, intelligence is leaked by the game master that reveals covert/deceptive operations. Players can negotiate with the game master authorization to perform new activities and be rewarded with new assets in the event the new activity is successfully carried out.

The same illicit nuclear trafficking game is played a number of time and the results are stored and analyzed to characterize the behavior of players. The analysis of players' behavior is then used to calibrate an agent-based model of how the exchange of goods and know-how may play out through time series simulations with reference to developments of ongoing behaviors and the emergence of new behaviors. Agent-based modeling (ABM) provides an ideal tool to capture the evolution of networking structure emerging from proliferation activities and knock-on effect from the behaviors of specific actors involved, such as the observables from the role-play activity described above. While a general clear and rigorous methodology for integrating data with ABMs is largely missing, TPA has made advances in developing algorithms that enable data integration for ABMs using sequential Monte Carlo methods, and in particular a particle filter, for developing a maximum likelihood-based approach to parameter estimation in ABMs from data [20]. Once the ABM is sufficiently calibrated, it is integrated with the game. Model parameters, roles and activities are matches with roles, assets and activities in the game so players' behavior is both regulated by and perturbs the model's simulations as discussed in [7, 16].

4 Conclusions and Further Work

Reaching an understanding of illicit nuclear trafficking that can successfully inform preemptive action requires that we (a) assess the propensity of state and non-state actors to proliferate within a geopolitical region of interest, and (b) anticipate how actors at risk of engaging in proliferation strategy may behave in order to achieve their objectives. In this paper, we have described a predictive analytics approach that achieves this objective by integrating evidentiary and dynamic modeling with knowledge management and analytical gaming, and demonstrated its application to a geopolitical region at risk. This approach can be instrumental in enabling analysts and policymakers to plan strategic action in influencing, disrupting, interdicting or destroying the function of illicit nuclear networks networks and their processes, and can be integrated with a radiation detection approach to address medium and short medium analysis and intervention objectives.

References

1. International Atomic Energy Agency, Illicit Trafficking Database, Fact Sheet: January 1993–December 2006 (2009)
2. Arlt, R., Brutscher, J., Gunnink, R., Ivanov, V., Parnham, K., Soldner, S., Stein, J.: Use of CdZnTe detectors in hand-held and portable isotope identifiers to detect illicit trafficking of nuclear material and radioactive sources. In: Nuclear Science Symposium Conference Record, vol. 1, pp. 4/18–4/23. IEEE, Los Alamitos (2000)
3. Nedialko, B.D., Gonzalez, M.A., Michalopoulos, D.P., Morton, D.P., Nehme, M.V., Popova, E., Schneider, E.A., Thoreson, G.G.: Interdiction Modeling For Smuggled Nuclear Material. In: Proceedings of the 49th Annual Meeting of the Institute of Nuclear Materials Management (July 2008)

4. McDonald, J., Coursey, B., Carter, M.: Detecting Illicit Radioactive Sources. Physics Today, 36–41 (November 2004)
5. Peterson, D., Sanfilippo, A., Baddeley, B., Franklin, L.: Proactive Intelligence for Nuclear Nonproliferation. In: Proceedings of the Annual Meeting of the ISA's 49th Annual Convention, San Francisco, CA, USA, March 26 (2008),
   http://www.allacademic.com
6. Coles, G.A., Brothers, A.J., Olson, J., Whitney, P.D.: Assessing State Nuclear Weapons Proliferation: Using Bayesian Network Analysis of Social Factors. In: Conference Proceedings from Pacific Northwest International Conference on Global Nuclear Security - The Decade Ahead (2010)
7. Sanfilippo, A.P., Cowell, A.J., Malone, E.L., Riensche, R.M., Thomas, J.J., Unwin, S.D., Whitney, P.D., Wong, P.C.: Technosocial Predictive Analytics in Support of Naturalistic Decision Making. In: Proceedings of The 9th Bi-annual International Conference on Naturalistic Decision Making, Covent Garden, London, UK, June 23-26 (2009)
8. Technosocial Predictive Analytics Initiative website,
   http://predictiveanalytics.pnl.gov
9. Whitney, P.D., Walsh, S.J.: Calibrating Bayesian Network Representations of Social-Behavioral Models. In: Social Computing, Behavior Modeling and Prediction, pp. 338–345
10. Whitney, P., White, A., Walsh, S., Dalton, A., Brothers, A.: Bayesian Networks for Social Modeling. In: Proceedings of SBP-11 (2011)
11. Walsh, S., Dalton, A., White, A., Whitney, P.: Parameterizing Bayesian Network Representations of Social-Behavioral Models by Expert Elicitation: A Novel Application of Conjoint Analysis. In: Proceedings of Workshop on Current Issues in Predictive Approaches to Intelligent and Security Analytics, pp. 227–232. IEEE Press, Los Alamitos (2010)
12. Jarman, K.D., Brothers, A.J., Whitney, P.D., Young, J., Niesen, D.A.: Integrating System Dynamics and Bayesian Networks with Application to Counter-IED Scenarios. In: 10th International Probabilistic Safety Assessment & Management Conference (2010)
13. Cowell, A.J., Gregory, M.L., Marshall, E.J., McGrath, L.R.: Knowledge Encapsulation Framework for Collaborative Social Modeling. In: AAAI Spring Symposium on Technosocial Predictive Analytics. AAAI Press, Menlo Park (2009)
14. Cowell, A.J., Gregory, M.L., Marshall, E.J., McGrath, L.: Knowledge Encapsulation and Collaboration. In: LP Pipeline for Special Track on Applied Natural Language Processing at 22th International FLAIRS 2009 (2009)
15. Riensche, R.M., Paulson, P.R., Danielson, G.R., Unwin, S.D., Butner, R.S., Miller, S.M., Franklin, L., Zuljevic, N.: Serious Gaming for Predictive Analytics. In: AAAI Spring Symposium on Technosocial Predictive Analytics. Association for the Advancement of Artificial Intelligence (AAAI), San Jose (2009)
16. Sanfilippo, A.P., Riensche, R.M., Unwin, S.D., Amaya, J.P.: Bridging the Gap between Human Judgment and Automated Reasoning in Predictive Analytics. In: International Probabilistic Safety Assessment & Management Conference, Seattle, WA (2010)
17. Sanfilippo, A., Cowell, A.J., Tratz, S., Boek, A., Cowell, A.K., Posse, C., Pouchard, L.: Content Analysis for Proactive Intelligence: Marshaling Frame Evidence. In: Proceedings of the 2007 AAAI Conference (2007)
18. South Africa: Nuclear Case Closed? National Security Archive. 1993-12-19 (1993),
   http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB181/sa34.pdf (retrieved 11/15/2010)
19. http://www.globalsecurity.org/wmd/world/rsa/nuke.htm
20. Jarman, K.D., Walsh, S.J., Whitney, P.D.: Parameter Estimation for Agent Based Models using Sequential Monte Carlo Methods. PNNL-SA-75391, Pacific Northwest National Laboratory, Richland, WA (2010)

Author Index