Samira El Yacoubi
Franco Bagnoli
Giovanna Pacini (Eds.)

LNCS 11938

Internet Science

6th International Conference, INSCI 2019
Perpignan, France, December 2–5, 2019
Proceedings

Springer

Lecture Notes in Computer Science 11938

More information about this series at http://www.springer.com/series/7409

Samira El Yacoubi · Franco Bagnoli ·
Giovanna Pacini (Eds.)

Internet Science

6th International Conference, INSCI 2019
Perpignan, France, December 2–5, 2019
Proceedings

Springer

Editors
Samira El Yacoubi
University of Perpignan Via Domitia
Perpignan Cedex 9, France

Franco Bagnoli
University of Florence
Sesto Fiorentino, Italy

Giovanna Pacini
University of Florence
Sesto Fiorentino, Italy

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-34769-7 ISBN 978-3-030-34770-3 (eBook)
https://doi.org/10.1007/978-3-030-34770-3

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the papers presented at the 6th International Conference on Internet Science (INSCI 2019), which was held during December 2–5, 2019, in Perpignan, France.

INSCI is a series of international conferences on Internet Science that were initiated in 2013 within the framework of the FP7 European Network of Excellence in Internet Science (EINS). The goal of EINS is to coordinate and integrate European research aimed at achieving a deeper multidisciplinary understanding of the development of the Internet as a societal and technological artefact which increasingly evolves with human societies. INSCI aims at pursuing the main EINS objective which consists of creating a real open and productive dialogue between all disciplines which study Internet systems under any technological or humanistic perspective.

INSCI 2019 was organized by IMAGES Espace-DEV from the University of Perpignan Via Domitia in association with the University of Florence and supported by the Institute of Research for Development (IRD). This conference built on the success of the first and second editions of INSCI which took place in Brussels, Belgium, and were organized as an EINS project, with the support of the European Commission. In its third year, the conference moved to Florence, Italy, with support from the Collective Awareness Platforms for Sustainability and Social Innovation (CAPS) initiative. In 2017, the conference was organized by CERTH Information Technology Institute in Thessaloniki, Greece. In 2018, the fifth edition took place in St. Petersburg, Russia, and was organized by the School of Journalism and Mass Communications and the Faculty of Applied Mathematics and Control Processes, St. Petersburg State University.

INSCI is a conference with a unique and precise scope; despite its wide-encompassing title, it focuses on interdisciplinary views on the Internet and Internet platforms as tools for people's engagement, opportunities, and better quality of life. Internet Science is a multidisciplinary field based on computer and communication science but covering many other disciplines like mathematics, sociology, anthropology, physics, complex systems analysis and control, economics, law, political science, literature, history, teaching techniques, etc.

Internet Science constitutes a very broad concept that includes the technical parts concerning protocols and connections, but also the contents and applications, and, more importantly, the people. The Internet is changing our lives, and therefore many crucial questions have arisen: How do people behave on the Internet? Are they changing their lifestyle and how? Can the Internet promote sustainability, cooperation, and collective intelligence? Can it support open democracy and policy making? How can awareness of possibilities and dangers of the Internet be promoted? What about topics like intellectual properties, privacy, reputation, and participation? What are the juridical aspect of the Internet? What about arts and humanities in general?

INSCI 2019 aimed at providing a good opportunity to bridge the academic and industrial communities in northern and southern countries in order to converge to a noble idea of the role that the Internet should play. The theme, "The Internet and Geopolitical Crossroads," dealt with the development of practices, platforms, and communication channels to promote the integration of southern European, Mediterranean, African, and Latin American countries with in a peaceful world, where the discussion of political freedom, migrant problems, health, and citizen participation can be promoted.

In the current volume, the best conference papers are published. From 45 submissions, 30 were selected for publication. They are organized into five thematic chapters.

The first chapter is devoted to the social aspects of the Internet, since most of the value of this network is given by humans and their contributions. The second chapter is devoted to the cognitive aspects of internet-mediated interactions. The third chapter contains papers related to some technical and technological aspects of communication, while the following chapter is related to data mining and analysis. Finally, the last chapter contains papers presented in the workshop about well-being and ageing.

This focused and intense program would not have been possible without the help and continuous encouragement of a number of people especially the members of Steering Committee who supported the organization of INSCI 2019 in Perpignan, France.

We express our gratitude to the Program Committee, as well as to the reviewers and local organizers, panel and workshop chairs, for their commitment and hard work. We are grateful to all the scientists that submitted their works, authors of published papers, and conference participants for their precious contribution.

Finally, the organization of INSCI 2019 was made possible thanks to the financial support of several institutions: the University of Perpignan Via Domitia, the Institute of Research for Development - IRD, the UMR Espace-DEV, the European Project InSPIRES, and the International Network of Systems Theory TDS. We are happy to see the conference expanding in its scope and coverage, and look forward to INSCI's future editions. We also thank EasyChair for the extremely important piece of publicly accessible software they developed.

September 2019

Franco Bagnoli
Samira El Yacoubi
Giovanna Pacini

Organization

Program Committee

Panayotis Antoniadis	ETH Zurich, Switzerland
Franco Bagnoli	University of Florence, Italy
Mohamed Bakhouya	International University of Rabat, Morocco
Stefania Bandini	University of Milano-Bicocca, Italy
Emanuele Bellini	Khalifa University, UAE
Hichem Ben Hassine	Institut Pasteur of Tunisia, Tunisia
Laure Berti-Equille	Espace-DEV, IRD, France
Leonardo Bocchi	University of Florence, Italy
Jonathan Cave	University of Warwick and UK Regulatory Policy Committee, UK
Sylvain Chatry	University of Perpignan, France
Florian Coulmas	University of Duisburg, Germany
Gabriella Cortellessa	CNR-ISTC, Italy
Sarah Cosentino	Waseda University of Japan, Japan
David Defour	LAMPS, University of Perpignan, France
J-Christophe Desconnets	Mission Infrastructures et données numériques, IRD, France
Driss El Ouadghiri	Moulay Ismail University, Morocco
Samira El Yacoubi	Espace-DEV, University of Perpignan, France
Allyx Fontaine	Espace-DEV, University of French Guyana, France
Begona Garcia Zapirain	University of Deusto, Spain
Carmen Gervet	Espace-DEV, University of Montpellier, France
Giorgio Gronchi	University of Florence, Italy
Alessio Guarino	Université de la Réunion, France
Andrea Guazzini	University of Florence, Italy
Setsu Ito	RCAST, The University of Tokyo, Japan
Salifou Nouhou Jangorzo	University of Maradi, Niger
Mohamed Khalgui	Xidian University, China
Olessia Koltsova	National Research University of Russia, Russia
Yiannis Kompatsiaris	ITI-CERTH, Greece
Morgan Mangeas	Entropy, IRD, France
Tamoo Matsuda	Mitsubishi Research Institute of Tokyo, Japan
Daniela Micucci	University of Milano-Bicocca, Italy
Isabelle Mougenot	Espace-DEV, University of Montpellier, France
Giovanna Pacini	University of Florence, Italy
Filippo Palumbo	Institute of Information Science and Technology CNR, Italy
Mario Paolucci	Institute of Cognitive Sciences and Technologies, Italy

Maria Jesus Pinazo Delgado	ISGlobal - Hospital Clínic, France
Raul Rechtman	Universidad Nacional Autonoma de Mexico, Mexico
Kavé Salamatian	University of Savoie, France
Panagiotis Sarigiannidis	University of Western Macedonia, Macedonia
Anna Satsiou	EYATH S.A., Greece
Frédérique Seyler	Espace-DEV, IRD, France
Thanassis Tiropanis	University of Southampton, UK
Žiga Turk	University of Ljubljana, FGG, Slovenia
Eloisa Vargiu	Eurecat Technology Center - eHealth Unit, Spain
Enrico Vicario	University of Florence, Italy
Daniele Vilone	Istituto di Scienza e Tecnologia della Cognizione (ISTC), CNR, Italy
Giuseppe Vizzari	University of Milano-Bicocca, Italy
El Hassan Zerrik	Moulay Ismail University, Morocco
Francesco Zurlo	Politecnico di Milano, Italy

Contents

Social Aspects of Internet

Assessing the Potential of Digital Collaborative Sharing Platforms in Fostering Neighbourhood Participation Through Volunteerism

Deniz Akin[1(✉)], Thomas Vilarinho[2], Jacqueline Floch[2], Trond Halvorsen[2], and Daniela Baer[2]

[1] Norwegian University of Science and Technology (NTNU), Trondheim, Norway
deniz.akin@ntnu.no

[2] SINTEF AS, Trondheim, Norway
{thomas.vilarinho, jacqueline.floch, trond.halvorsen, daniela.baer}@sintef.no

Abstract. This paper seeks to discuss whether and how digital collaborative sharing platforms can foster citizen engagement in urban neighbourhoods by addressing various challenges experienced by local initiatives. To that aim, we conducted a case study in the Saupstad neighbourhood of Trondheim in Norway.

Our case study includes qualitative interviews conducted with volunteers and general managers of a selection of central volunteering initiatives in Saupstad where we aimed to map the current state of neighbourhood volunteerism, and identify the challenges experienced both at the individual and organizational level. We also involved stakeholders in a co-creative dialogue meeting to discuss and develop scenarios to overcome the identified challenges and foster citizen engagement in volunteering activities.

Based on the empirical data we collected, we have identified challenges centralized around the interrelated themes such as 'volunteer motivation', 'volunteer recruitment', 'effective dissemination of information' and 'collaboration and communication with local actors'. In this paper we discuss to what extent collaborative platforms can address these challenges and be utilized to foster citizen participation.

Keywords: Volunteering · Community · Case study · Citizen engagement · Collaborative platforms · Community informatics

1 Introduction

This paper investigates how digital collaborative sharing platforms can address challenges encountered by neighbourhood volunteering organizations by conducting a case study in the Saupstad neighbourhood of Trondheim in Norway. First, we map the current state of neighbourhood volunteerism in Saupstad and identify various challenges experienced both at the individual and organizational level. Secondly, we discuss to what extent collaborative sharing platforms can address these challenges and be

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 3–17, 2019.
https://doi.org/10.1007/978-3-030-34770-3_1

utilized to foster citizen participation in neighbourhoods through volunteerism. In this context, we consider both well-established global social platforms, as well as new platforms tailored to the neighbourhood level.

The research is novel in the dual open approach that is employed. Rather than studying the introduction of a given platform, we employ qualitative interviews and community workshops to identify the needs of local non-profit organizations who depend on volunteerism. We are open to the choice of platform to adopt, and to which organization or group that it will be employed to. By providing examples of services that digital sharing platforms can offer, our interaction with prospective users helps them to identify needs that they did not realize beforehand, and we get rapid response on potential unintended consequences that the use of the platforms may entail.

The results are intended for social planners, such as local government officials and managers of non-profit organizations, who are considering adopting a digital collaborative sharing platform to foster citizen participation in a geographically defined neighbourhood. The findings are also relevant for programmers and designers of these platforms because we identify user needs that are not sufficiently covered by existing platforms, as well as social, organizational and technological barriers to adoption. Finally, we aim to add to the limited volume of research addressing the interaction between sharing economy and volunteerism at the local level.

2 Background

Social relations in urban neighbourhoods, and urban neighbourhood communities have largely been described as having undergone a massive transformation following modernization, industrialization, and urbanization. In respect to the effects of urbanization on social relations, Ferdinand Tonnies had famously formulated the concepts "Gemeinschaft" (communities built on personal relations) to "Gesellschaft" (communities built on instrumental relations) [1]. Accordingly, societies go through a transformation from Gemeinschaft to Gesellschaft. Communities that were once defined by a dense face-to-face interaction and interdependence of people with shared rituals and values, had vanished and gave its space to social relations based on impersonal ties defined by individual self-interest rather than a common good [1–3]. In respect to these conceptualizations, studies of urban neighbourhoods have been focusing on whether and how social relations among neighbours have been altered.

Keen on revitalizing social interaction and boost citizen participation in urban neighbourhood communities, there are several community-building activities organized by a variety of actors ranging from local policy makers, non-governmental organizations and volunteer-driven initiatives.

In this context citizen participation or citizen engagement is usually understood as citizen action where citizens take part in activities designed to gain some influence over conditions affecting their lives [4]. Citizen participation in neighbourhoods are specifically considered as a vital means of increasing the individual and collective capacities, as well as connections to the neighbourhoods, and therefore improving both the physical environment and the social conditions for the residents [5, 6]. Boosting citizen participation in urban neighbourhoods, undoubtedly, relies on volunteers who directly or indirectly willing to take part in neighbourhood activities, as an organizer or sole participant, generating social benefit for the other fellow residents.

However, as previous research has demonstrated, viability of volunteering organizations is dependent on sustainable volunteerism, which entail attentiveness to various, divergent and individual needs of the volunteers, and therefore has its own complexities in achieving long-term, sustainable participation [7].

The last couple of decades, a growing scholarly interest has also been centred around the impact of the use of information and communication technologies (ICTs) in social and civic life [3, 8, 9]. Some of these studies express concerns over the replacement of face-to-face interaction by cyber communication in virtual space, emphasizing on how increasing use of internet for communication results in isolation, lack of communicative skills in actual life, and a declining community engagement. In contrast to this, other studies emphasize the promises of ICTs, claiming that ICTs can facilitate broader social networks in ‘real life’, and therefore can be a solution to the modern times marked by a decline of traditional communities [10].

The dichotomic positioning of ICTs and real-world connectivity has been criticized broadly due to forcing a separation between online life and the off-line ‘real life’ [3]. It is easy to find examples of both the promises and limitations of ICTs, as these technologies can be conceived as both the cause of, and the solution to, the loss of communities in the traditional sense [3]. With many different ICTs to choose from, it’s crucial to also consider which technologies are used, and how they are being used.

In this article our aim is not to vouch for ICTs for citizen engagement in neighbourhood activities in Saupstad Norway as we acknowledge the cultural, political and ethical dimensions of faith in technological fix [11]. Drawing upon the complex social fabric of Saupstad, that hosts residents of multiple backgrounds, we aim to first map limitations and challenges experienced by the current organization of citizen activities. Second, we seek to discuss whether and to what extent these limitations can be addressed by digital collaborative sharing platforms that are designed to increase and encourage social connectivity via easier communication, peer-to-peer sharing, and therefore might have an implication for citizen engagement and social togetherness for local communities and neighbourhoods.

3 Methodology

The research was organized as a case study centred on the Saupstad neighbourhood in the city of Trondheim, Norway. This specific neighbourhood receives considerable attention and funding from the municipal government with efforts to promote neighbourhood participation and reduce social isolation in target groups. The municipal program focuses on improvements to the physical infrastructure, as well as financial and other support to local ideal organizations, and we saw a potential to investigate whether digital collaborative sharing platforms are suitable tools to support the municipal efforts to foster volunteerism in this area. We adopt John Wilson’s definition of volunteering as “any activity in which time is given freely to benefit another person, group, or organization. This definition does not preclude volunteers from benefiting from their work” [12].

We aimed to collect as much nuanced knowledge as possible to map the various challenges, as well opportunities, experienced by the volunteering organizations environment in Saupstad, and how the residents participate in their neighbourhood, and whether and how this participation can be boosted through use of digital collaborative sharing platforms. Neighbourhood participation through volunteerism is a broad field that cannot be restricted to one observational site. This is why our study has focused on a selection of local organizations, fully or partially run by volunteers, that aim to generate a social benefit to the residents of Saupstad through the activities they offer.

3.1 The Case Context

Saupstad is a suburb located in the south of Trondheim, Norway's third biggest city. The district is a part of the larger Saupstad-Kolstad region which is approximately 45 years old. Built in 1970s, Saupstad is the largest suburb in Trondheim. Having a relatively low record in respect to the indicators such as, income, education and disability benefits, Saupstad has captured the attention of politicians, architects, urban planners among other experts [13]. To address these challenges an 'area lift'[1] (in Norwegian: områdeløft) was initiated in the region [14]. Between 2012–2016 several measures were implemented in the area and there are new measurements that are planned to be completed in the region between 2017–2020.

The overall objectives of 'area lifting' are: to promote quality of life and health; building competence to the children and youth which prepares them to the challenges of the future; building a sustainable region with a variety of building structures and good quality of infrastructure and public space; to build a district with meeting areas and opportunities of citizen participation [15]. The area lift project is designed in a way that aims to mobilize the region's own resources which entails the inclusion of local community and incorporation of their needs and demands. In this respect a set of networks are established between housing cooperatives, residents, associations, unions, volunteer organizations [15].

Despite being conceptualized as a so-called "problem district" where special measures to be taken to improve the residential and the living conditions in the neighbourhood, the residents of Saupstad are documented to be content living in the district [13, 16].

3.2 Data Collection and Analysis

The data collection followed a two-stage process. During the fall of 2018 we performed exploratory activities with local volunteer and non-profit organizations as well as individual inhabitants in order to identify the core needs of the organizations in the area. In the second stage (spring 2019), we narrowed down the number of organizations that we engaged with and began more targeted discussions about the potential of various sharing platforms.

[1] *Area lift* is a national social program that aims to improve the attractiveness of selected urban areas in the largest cities in Norway. Social inclusion and citizen participation are central aims of area lift projects.

This approach had several advantages. First, we wanted to avoid a situation where we ended up working with an organization that was not prepared to try out a new technology. Second, we wanted to identify the challenges where collaborative sharing platforms could make as large impact as possible. Third, it gave us time to establish a closer relationship with the managers of the organizations. Consecutive meetings gave us the opportunity to confirm or modify our impressions about the challenges. Fourth, we were able to follow the developments in the area over time to better understand the social and physical context that the technological solution should work in. Fifth, we had time to explore relevant emerging technologies that were brought to market in the same period.

The methods we employed were:

– **Qualitative Semi-structured Interviews with the Organizations' General Managers:** We have interviewed the general managers of VCS and KFC. Interviews focused on current state of volunteerism (activities offered, level of participation of residents and their motivations), recruitment strategies, and dissemination channels within each organization. We also asked questions about challenges and opportunities in respect to the usage of ICT for issues such as, planning, coordination and following of organized activities.
– **Qualitative Semi-structured Interviews with the Volunteers:** We have interviewed 4 volunteers from VCS. These interviews followed the same thematic as the ones with the organizations' leaders but with a volunteer and citizen perspective. Interviews focused on the informants' motivation to engage in volunteerism, their communication with the organization and the other volunteers, as well as various challenges and opportunities attached to their engagement as a volunteer.
– **Street Interviews:** We have conducted 22 interviews with people on the street. We have reached out a variety of residents in respect to age, gender, race, occupation. In these interviews we have asked people whether and how they follow social activities in Saupstad, either as a volunteer or as a participant. Additionally, we have also asked people to describe the kind of relation they have with their neighbours to understand whether they socialize with their neighbours and whether they would ask for help or other kind of assistance from them if they need it. We have also collected information on people's everyday use of ICTs and theirs attitude regarding using ICT to interact with theirs neighbours.
– **Exploratory Design Meeting (Dialogue Meeting)**: We arranged an exploratory design meeting in Saupstad. The design meeting was organized according to scenario-based approach which is quite common in the field of interaction design and development of ICT tools. Scenario-based approach is agreed to be a useful way to engage different stakeholders to elaborate on the design concept by enacting particular scenarios [17]. For the exploratory design meeting arranged in Saupstad, we have prepared scenarios based on the challenges of volunteerism we had identified in our research data material. We then, involved stakeholders to discuss these challenges through a game called "what if". In this game the stakeholders engaged in creation and discussion of possible scenarios to overcome the challenges. Although the meeting was open to the all residents of Saupstad, those who participated in the meeting were all active participants of the local community:

volunteers of the Saupstad Youth Council, volunteers of the VSC, leader of the neighbourhood school. Through different scenario-based games, volunteers explored a variety of problems and possibilities in respect to the volunteering activities they have been engaging.

The semi-structured interviews, the audios from the Exploratory Design Meeting and the answers from the street interviews were analysed in an inductive way following the principles of Klein and Myers [18]. Two researchers listened to the material and did a first independent round of thematic analysis resulting in a set of codes [19]. The researchers discussed and analysed together the different codes and parts of the coded data for establishing a set of overarching themes linked to our research questions.

3.3 Choice of Target Organizations

Among different local initiatives and organizations in Saupstad, we chose to focus on the Volunteering Centre of Saupstad (hereafter abbreviated as VCS), Kolstad Football Club (hereafter abbreviated as KFC) and Midteggen Housing Cooperative (hereafter abbreviated as MHC) given that they play an active role in neighbourhood community-development, and that they depend on volunteer engagement[2]. In addition to that these initiatives have different user groups and therefore offer a diverse view on the aspects of volunteering and social interactions in the neighbourhood.

The Volunteering Centre of Saupstad (VCS) was established in 2004, and forms one of the eleven "volunteering centres" in Trondheim[3]. The centre is own by Trondheim Red Cross and supported financially by the Municipality of Trondheim. It has one full-time employer who is the general manager of VCS. Currently, there are around 80 people volunteering for the centre that arranges various activities following local needs and demands. The need for an activity can be identified and suggested by a local neighbourhood partner, a volunteer or the general manager, which can further be discussed with other volunteers. The provision of activities relies on the willingness of volunteers to support and carry out the activities. Therefore, the range of the activities tend to focus on the user groups that engage in volunteering at the volunteering central. The major user-group of the VCS is retirees, and unemployed residents. VCS offers activities, run by volunteers, such as tour-group, dancing group (Friskusdans), choir, accompanion activity (følgetjeneste-for those who need accompany to go to shopping or hairdresser etc.).

Kolstad Football Club (KFC) was founded in 1972. The club is administrated by a board and the two full-time employers who are responsible for the club operations and day-to-day management. A great deal of people who engage with other tasks for the club are volunteers such as, trainers, people who stand in the food-counter during games and those who sell lottery tickets to raise money for the club. There are also

---

[2] In addition to these organizations listed, Saupstad is also home for other sport teams, Mevlana Cultural Association, The Retirees Association and Youth Council [ungdomsråd] which represents the youth in Saupstad in the municipality.

[3] The centres belong to the nationwide interest organization "Frivilligsentral". All centres have a common mandate.

volunteers called upon when the football field is to be cleaned from snow or other kind of tasks that requires collective action (in Norwegian 'dugnad'). The main user groups of KFC are children, youngsters and families who have children. KFC also appears to be a local actor that provides service and activities to everyone, not only children and youngsters. In cooperation with Trondheim Municipality, the club runs 'Utlånssentral' which lends out outdoor and sport equipments to anyone free of charge. In addition to that, the club organizes 17. May celebration in Husebyhallen, which intends to gather all the residents, and Blokk Rock Festival in the neighbourhood.

Midteggen is a housing cooperative [in Norwegian 'borettslag'] built in 1973. MHC is composed with 24 building comprising 472 apartments. The residents of the housing complex composed of families with and without children, retirees, people with a minority background. The management board of MHC, represented by volunteering residents, arrange a variety of activities for this mosaic of residents with different backgrounds.

As a site of analysis, we focused on VCS, KFC and MHC due to the active role they plan in arranging volunteering opportunities in Saupstad.

3.4 Identification and Assessment of Relevant Collaborative Sharing Platforms

Digital sharing platforms are internet-based applications for computers or handheld devices that allow users to share privately owned resources and information. Platforms for the sharing economy has had an exceptional impact on society over the past decade, but many of them are focusing on consumption of services in a strictly professional way, with only minimum of social contact between buyers and sellers. With the term collaborative sharing platforms, we refer to applications that have design features that directs it's use to the users' own local area. Typically, they will make an explicit statement in their marketing that they aim to strengthen communal work and inclusiveness.

Through online searches, and engagement with colleagues and app developers, we identified the following platforms that we assessed in more detail:

- Facebook: An American social media platform with a high penetration in Norway. Several ideal organisations establish so called Facebook groups for coordinating activities among members in a local area.
- Give and Take: A platform for social workers and volunteers developed in a Horizon 2020 research project (2014–2017). We arranged a meeting in Trondheim with associate professor Lone Malmborg from IT University of Copenhagen who coordinated the Give & Take project to discuss various aspects of the platform.
- Comoodle: A platform for helping inhabitants share the resources (classified as either stuff, space or skills) within a given community. Based on Kirklees, England, and funded by the Bloomberg Philanthropies organisation. We arranged a meeting in Trondheim with Duggs Carre form the Kirklees city council to hear about their experiences with Comoodle.

- Friskus: A Norwegian platform for disseminating information about local activities, such as walking tours, common breakfasts, ball games, etc. We engaged in dialogue with the developers to assess its relevancy for Saupstad.
- Bobleberg: Danish application that allows users to create interest groups, called bubbles, to find and meet others with similar interests. They have established a collaboration with the Danish Red Cross to combat loneliness.
- Nabohjelp: Nabohjelp allows users to post requests for services or items that they need or wish to offer. Developed by Obos BBL, the largest housing developer and homeowner association in Norway. It was launched in May 2017 and usage frequency is highest in Oslo.

The platforms were assessed by downloading them to our smart phones where we explored their user interface. We then proceeded to present the main features to volunteer organisations in the interviews and dialogue meeting. Dialogue with developers and users also provided useful insights.

4 Current State of Volunteerism in Saupstad: Opportunities and Challenges Identified

In what follows we discuss the challenges and opportunities central to the organization of volunteering in VCS and KFC by emphasizing on key interrelated themes namely: ‘volunteer motivation’, ‘volunteer recruitment’, ‘dissemination of volunteering activities’ and ‘collaboration and communication with local actors’.

4.1 Volunteering Motivation

Based on the data we collected, we observed that there are various personal interests that motivate individuals to volunteer for the neighbourhood organizations. Retirees volunteering for VCS, for example, are motivated by the need of developing further social ties and occupy their free time meaningfully. As one of the informants stated “*After you become a retiree, life gets boring*”. In this sense, volunteering plays a role in fulfilling the activity and social contact gap that many experiences after retiring. In a similar vein, one informant who was unemployed during the interview said: “*I could not find a job, but I could not think of just sitting at home. That was the reason why I became a volunteer*”. In addition to using volunteering as a means of utilizing free time, two other retirees we have spoken to mentioned that they see volunteering as a continuation of their work life. These two informants used to work with elderly when they were working for the municipality. In this context, to volunteer for elderly after retiring was natural for them as they had familiarity, interest and experience with this age group. VCS also attracts other user groups such as families with children. This user group prefers to engage in community activities which they can involve their children as well. These activities include, but not limited to, gardening, cleaning outside space or painting. Such activities, according to the interviewees, also provide an opportunity for children meet and play together.

Similarly, KFC has a profile of volunteers that is largely composed of people with young children. This is, in fact, a common tendency in Norway where many volunteers within sports are adults with younger children. That is, families who have a child playing football in the club are likely to be involved in volunteering activities arranged by Kolstad Football. According to our data material, the club is also actively working with remaking the image of Saupstad district which is often designated as a problem district by policy makers. In this sense, an additional factor that motivates people to volunteer for the club could be a sense of belonging to the neighbourhood and willingness to contribute to a better image.

Based on the data we collected, one can say that, similar to the results from various research on the motivations for volunteering in contemporary times, people engage in volunteering activities that responds their individual interest [7].

4.2 Dissemination and Recruitment

Both VCS and KFC utilize social media, their websites, the local newspaper and posters placed along the neighbourhood as a general means of disseminating information about the organizations and their activities. Although the general leader of VCS and KFC regard these communication tools important, they both expressed a need for improving their ways of spreading information. General manager of VCS states: *"There is a lot that happens in Saupstad, everyone agrees. But all agree that nobody knows what happens and when. People can have the posters in front of them and not read them."* According to the organizations, one of the reasons that people are not attentive to the information that is being disseminated in Saupstad might be that they find it too impersonal. Their presumption is supported by an extensive rapport written on the conditions and motivations for voluntary work in Norway, which states that the way people are being asked for volunteering and who asks them to volunteer is determinative for volunteering engagement as well as the amount of effort invested in the voluntary work [20].

During the dialog meeting we arranged in Saupstad, the stakeholders also drew attention to the ways in which arrangements and activities are being disseminated in the neighbourhood as posing barriers for participation. Inviting the whole neighbourhood through a single and all-encompassing invitation, according to stakeholders, prevent residents to feel personally welcomed, and make them feel not needed if the announcement is made to invite residents to take part for a sort of community work. As a solution to that, during the scenario-based games, the stakeholders came up with the idea of personalizing the invitations by adding the names of the residents on the announcements and sending it out individually as electronic post, or regular mail.

Barriers related to technology is also mentioned as a problem in respect to sending out information. Our data shows that while some elderly feels comfortable with having a smart phone and surfing on the internet, others prefer printed media as they can cut of the announcement and place it on their fridge with a magnet. Keen on being inclusive and reaching out to those who does not use internet, as well as to those who prefers face-to-face invitation, both organizations seek means to do so. For example, VCS prefers to meet the retiree's association in order to invite the elderly to the ongoing activities personally. Furthermore, many come to VCS via advices of friends who participate on the activities or volunteers. Most of the activities of the VCS target the retiree user group, as the activities are driven by the volunteers and most VCS volunteers are retirees. Consequently, the dissemination via word of mouth and the participation on activities reaches out retirees more than other user groups.

When it comes to the recruitment of volunteers, KFC usually reaches out to children through the schools. Adults, usually the parents of the children who play football, engage in KFC as they are willing to support the team and their children. For VCS, volunteers are often recruited subsequent to their participation on an activity arranged by VCS. If they enjoy the activity that is being organized, it is likely that they develop further interest to be involved and help the general manager for the upcoming arrangements.

We observed that despite having activities that target the whole neighbourhood and a willingness to involve multiple user groups, both organizations are quite homogeneous concerning their core user group (Parents and children at KFC and seniors at VCS). Consequently, that core-user group becomes somewhat associated with the organizations which provides advantages and disadvantages. It facilitates the word of mouth dissemination through their core user group, but it makes it more difficult for people outside of the core user group to feel belonging or for the organizations to consider ways of including other user groups. It hinders its potential to tap on the heterogeneity of the neighbourhood and spread social capital along different user groups. A volunteer from VCS tells about her friends with kids, who would not find a fitting activity in VCS despite activities for kids at VCS such as the Family Day. Similarly, when KFC leader was asked about how to include elderly residents into KFC volunteering, he understood that the elderly is more interested in activities where they are more direct beneficiaries: *"they (the elderly of the neighbourhood), unfortunately, are somewhat too busy with their own activities and personal development. That is very good. But, that means that those, who have wider time availability, look for possibilities within the VCS, which does a very good job. They engage into the retirees' association, as to some extent look for activities for themselves."*

4.3 Collaboration and Communication Among Local Actors

In addition to VCS and KFC, Saupstad is home to many initiatives of different sizes that are run by partly or totally by volunteers, targeting different groups of people in the neighbourhood. Our material shows that VCS and the KFC have partial and limited collaboration and communication with other local actors. For instance, KFC sometimes reach out to volunteers via the mosque located in the neighbourhood for activities, such as shoving the snow from the football pitch. Similarly, VCS collaborates with Associate for Retirees for many common arrangements and for reaching out to retirees. However, there is no common forum where the local actors periodically meet, develop collaborations and synchronize their activities.

It is hard to deny the fact that these organizations also compete to a certain degree due to targeting the residents of the same neighbourhood. Problems mostly arise, when their regular activities collide with each other. As mentioned before, each organization experience challenges in recruiting new volunteers. One reason seems to be that there are many activities arranged simultaneously and are forced to compete by being most interesting.

4.4 Internal Communication and Assignment of Volunteering Tasks

Internal communication is not much of an issue at KFC. There, most of the volunteering assignments are pre-defined and announced well in advance. There is no need of deliberation of volunteers to define the assignment and it only requires a bit of communication at the beginning of the season to get them sorted out. On the other hand, at the VCS, the activities depend on the interest of volunteers on performing them and the local needs. This mean that the VCS general manager facilitates the communication with volunteers for starting and developing activities and for connecting interested members.

In our interviews with the VCS, we talked with both the VCS leader and VCS volunteers, being some of them also leaders of activity groups within the VCS. Through those interviewees we managed to understand their communication practices and theirs challenges and limitations.

The general manager works as the main facilitator of the volunteering activities of the VCS, while the volunteers would be the ones who drive the activities. Those roles apply specially in activities such as the "following service" where the VCS general manager connects a volunteer with a person that needs a companion for running an errand and in activity groups such as the walking group. In those activity groups, the general manager supports them by arranging the infrastructure needed for the activity (room, coffee, etc.), help recruiting members and promoting the activities. While the participants, normally led by one or more activity group leaders, organize the walks.

However, most of the communication goes via the general managers. The centralization of communication puts a heavy burden on the general manager who uses a lot of time in communicating with the different members. Facilitating an internal activity group communication and arrangement could relief the activity leader. When asked about that possibility of having a message group or a telephone list, the interviewed VCS members thought that indeed that could be helpful. One of them suggested that the regular organization of the activity groups could be done by the group itself: *"I think that it can be fine if a group works alone without the general manager, but if there is something (out of the normal) it is natural that we involve her"*.

The communication at the VCS is done both via direct messages between the general manager and individuals or via broadcast messages to several members. Those are used for all purposes, inclusively to recruit members to engage into existing volunteering assignments. As one interviewee describes, the engagement of the leader in asking for help is crucial for them to volunteer in the assignments: *"She (the leader) is very good at sending messages and asking us (to engage in specific activities) and, I would not have worked as a volunteer without her"*. However, as another volunteer points, there are times when they feel like they could participate even more, but do not do it for feeling like there is nothing to do, while the reality is that the general manager says that she often has more assignments then volunteers and ends up taking care of many of those herself. The general manager asks members to inform her if they are available to engage in coming assignments, but the interviews with the VCS members suggest that this does not happen very often.

However, the interviews show that several considerations must be taken when considering a different communication channel, in special when engaging with an elderly audience. Although there is a perception (from the VCS general manager and the interviews on the streets) that few people read the posters placed along the neighbourhoods, the elderly volunteers we talked to said that they use those. Moreover, they mentioned that they occasionally delete or forget about old SMS messages or e-mails and, therefore, communication in paper works best: *"I brought (the leaflet of an activity) home with me. It is on my table. It is easier to remember. I do not seat in front of the computer all the time"*. Furthermore, some do not necessarily use computers, smart phones or messaging apps.

4.5 Responsibility Sharing

A central concern expressed by the general managers of the organizations is uneven distribution of responsibility in respect to recruitment, dissemination of information, and other practicalities related to the organization and coordination of activities. Although the general managers try to encourage volunteers to take more organizational responsibility, people who volunteer prefer not to get involved in administrative tasks, as we were told.

It is possible to argue that this is related to the changing nature of volunteerism in contemporary times, where individuals tend to engage in more sporadic, temporary forms of volunteering rather than traditional volunteering based on lifelong commitment and attachment to an organization [7, 21]. Our data material shows that volunteers have very little, if not any, attachment to the organization they volunteer for, and they appreciate being flexible in the volunteering effort they do. It is the activities they are engaged in and the social environment that the organization offers them, which attract them. One of our informants who volunteer for VCS mentioned that she appreciates the freedom she has to say 'no' when being asked to attend an activity. This can partly explain the lack of willingness volunteers have in taking organizational responsibility which would restrict their flexibility.

5 Discussion

Drawing upon the data material we collected and analysed, one can say that neighbourhood participation through volunteerism in Saupstad has a set of interrelated challenges, as well as untapped opportunities that could have been used to foster residents' engagement in the local organizations. In this regard, the main issues seem to be centralized around themes such as 'volunteer motivation', 'volunteer recruitment', 'effective dissemination of information' and 'collaboration and communication with local actors' and 'responsibility sharing'. In this respect we wonder whether and how digital collaborative sharing platforms can be of use in overcoming these challenges and contribute to neighborhood community development.

Our analysis shows that the ways in which local organizations disseminate their activities has an impact on recruitment of volunteers. In this sense a major concern seems to be whether the information spread reaches out everyone, as well as whether it is attractive to encourage participation. As mentioned before, general invitations run the risk of going unnoticed and being neglected. In contrast, personalized invitations and word of mouth appears to have a bigger potential to mobilize engagement and recruiting new people. Given that preparing personalized invitations in printed form would be costly as well time inefficient, utilization of ICTs which offer personalized invitations could support and strengthen engagement. Social network sites, that supports mainly messaging and discussions, can potentially support the effective personalized dissemination. As previous research has shown, individuals are often motivated to participate when they are asked by someone they know, like a family member, friend or acquaintance [20]. In this context, and based on its enormous popularity, Facebook appears as a potential facilitator and therefore being used by many organizations. It is a platform that makes it possible for individuals to invite the list of 'friends' they have listed in the platform. In this sense organizations can utilize the friend-list of their active members to recruit more. Facebook, however, has multiple purposes that might preclude the main intention of organizations. The intense flow of information on the application might also re-generate the issue of being indifferent to the announcements, a problem that the residents of Saupstad already experience.

In this sense, we argue that platforms that are local and specific in terms of their scope of use could be of more use. What we mean by aim-specific is that the application should have mainly the aim of connecting a local community and lowering down the threshold for those who are willing to be more active in their neighbourhood. Platforms such as Bobleberg allows individual users to form groups for people with similar interest, such as a knitting or fishing group. Nabohjelp and Comoodle can be designated as classic platforms of sharing economy where skills and assets are shared. Through Nabohjelp one can narrow down the area where a particular help/service offer is being disseminated. None of these applications requires the so-called surveillance of an organization, but rather allow individuals to reach out others in an independent manner. In giving individuals more freedom to post and disseminate, these platforms suit for those who wish to engage in volunteering but do not want to belong to an organization. As we discussed under Sect. 4.5, it is quite common that people appreciate the flexibility as a character in the volunteering they do. That is, they would like to decide the time and the kind of activity they wish to be a part of. Lack of admin status, however, leaving these platforms in a precarious situation concerning one-to-one matching services. For instance, an elderly person might not feel comfortable to ask a total stranger to come over her apartment and change the light bulb. When it comes to this kind of one-to-one services, VCS, for example, runs a sort of background check of the volunteers and provide assurance. When it comes to the generation of social capital through these platforms, as the literature on sharing economy has shown, there are promises of an increasing trust among user groups who share skills and assets [22]. Yet, it is hard to predict as well as measure whether these applications do result in meaningful, long-term social relations among the users which can be translated into long-term local engagement in the area one resides in.

In respect to our research question, FRISKUS appears to be the most useful platform designed in a way to offer digital solutions to coordination and arrangement of volunteering activities locally. FRISKUS offers a local common activity calendar where the ongoing arrangements in a certain region or neighbourhood are gathered. The organizations, that purchase the license, can use the main features of FRISKUS which can be listed as, posting activities to the common calendar, sending personalized invitations to the members in a local area, as well as generating and sending certificates and other documents digitally to the participants. The activity calendar is available to everyone regardless of the fact that one has a user-profile. In order to register attendance and send private messages, however, one needs to sign up free of charge. Given that FRISKUS specifically working with local organizations, one receives information that is solely related to the activities in their neighbourhood. In that sense, users could both feel personally welcomed, and that they would not get lost in a pool of out-of-context information. Furthermore, one chooses the region, municipality or neighbourhood they would like to follow, to narrow down the announced activities. Needless to say, the common activity calendar makes it easier for the general managers of the organizations to schedule their activities in order to avoid overlaps. Also, the platform offers a database of voluntary registry, which makes it easier for the general managers to search through and get in touch with volunteers. It is important to underline that one without an admin or coordinator status is not allowed to post an announcement on FRISKUS. In that sense, peer-to-peer activities are somewhat limited on this platform. As mentioned before, general managers carry the heavy burden of practical tasks, and often fail to engage the volunteers to take over some organizational task. In this sense, using FRISKUS might add to the burden of the general managers if they fail to recruit volunteers who would like to take on admin/coordinator title on the app and post announcement on behalf of the organization. Another challenge that remains with FRISKUS, as well as any digital platform, is that they cannot be utilized to reach out people out of the cyberspace. That is, anyone who does not have access to digital tools or lack willingness of literacy to use technology, would be excluded. It is, on the other hand, possible to transfer the common calendar of FRISKUS to any public digital screen, which could have increased the visibility of ongoing activities.

Our research has shown that neighbourhoods, and the local communities are composed of residents of different needs and demands. In this socially, culturally and demographically complex setting, fostering neighbourhood participation through volunteerism remains a hard task carried by various local organizations. Mapping the current state of volunteerism in Saupstad, Trondheim in Norway, and identifying the challenges experienced by the local organizations, we discussed whether and how collaborative sharing platforms can be utilized. We argued for potential use of platforms that are local and specific in respect to their scope of use. Needless to say, social life is far more complicated to be envisaged by the design of technological tools. However, digital platforms such as FRISKUS seems to bear a great potential in having features that would respond many, if not all, of the challenges connected to neighbourhood participation.

References

1. Tjora, A.: Hva er fellesskap. Universitetsforlaget, Oslo (2018)
2. Henning, C., Lieberg, M.: Strong ties or weak ties? neighbourhood networks in a new perspective. Scand. Hous. Plann. Res. **13**(1), 3–26 (1996)
3. Bell, D.: An introduction to cybercultures. Routledge (2006)
4. Wandersman, A., Florin, P.: Citizen participation and community organizations. In: Rappaport, J., Seidman, E. (eds.) Handbook of Community Psychology. Springer, Boston (2000). https://doi.org/10.1007/978-1-4615-4193-6_11
5. Chavis, D.M., Wandersman, A.: Sense of community in the urban environment: A catalyst for participation and community development. Am. J. Commun. Psychol. **18**(1), 55–81 (1990)
6. Ohmer, M.L.: Citizen participation in neighborhood organizations and its relationship to volunteers' self-and collective efficacy and sense of community. Soc. Work Res. **31**(2), 109–120 (2007)
7. Hustinx, L., Lammertyn, F.: Collective and reflexive styles of volunteering: a sociological modernization perspective. Voluntas Int. J. Voluntary and Nonprofit Organizations **14**(2), 167–187 (2003)
8. Stoecker, R.: Cyberspace vs. face-to-face: community organizing in the new millennium. Perspect. Global Dev. Technol. **1**(2), 143–164 (2002)
9. Mandarano, L., Meenar, M., Steins, C.: Building social capital in the digital age of civic engagement. J. Plann. Lit. **25**(2), 123–135 (2010)
10. Hampton, K.N.: Neighborhoods in the network society the e-Neighbors study. Inf. Communi. Soc. **10**(5), 714–748 (2007)
11. Johnston, S.F.: The technological fix as social cure-all: origins and implications. IEEE Technol. Soc. Mag. **37**(1), 47–54 (2018)
12. Wilson, J.: Volunteering. Annual Rev. Sociol. **26**(1), 215–240 (2000)
13. Rørtveit, H.N.: Genialt planlagt? Drabantbyen som hjem i spennet mellom boligfaglige idealer og levde landskap. Norwegian University of Science and Technology (2015)
14. Trondheim Kommune. In: Områdeprogram for Saupstad-Kolstad 2013–2020 (2013)
15. Trondheim Kommune. In: Program for Områdeløft Saupstad-Kolstad Handlingsdel 2017–2020 (2017)
16. Rambøll, M.C.: Kvantitativ undersøkelse for områdeløft Saupstad-Kolstad. In. Trondheim (2015)
17. Brandt, E., Messeter, J.: Facilitating collaboration through design games. In: Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices, vol. 1, pp. 121–131. ACM (2004)
18. Klein, H.K., Myers, M.D.: A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Q. **23**(1), 67–94 (1999)
19. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006)
20. Wollebæk, D., Sætrang, S., Fladmoe, A.: Betingelser for frivillig innsats: motivasjon og kontekst. Rapport fra Senter for forskning på sivilsamfunn og frivillig sektor (2015)
21. Rehberg, W.: Altruistic individualists: motivations for international volunteering among young adults in Switzerland. Voluntas Int. J. Voluntary Nonprofit Organ. **16**(2), 109–122 (2005)
22. Bucher, E., Fieseler, C., Lutz, C.: What's mine is yours (for a nominal fee)–exploring the spectrum of utilitarian to altruistic motives for internet-mediated sharing. Comput. Hum. Behav. **62**, 316–326 (2016)

Social Network Sentiment Analysis and Message Clustering

Alexander A. Kharlamov[1,2,3], Andrey V. Orekhov[4(✉)], Svetlana S. Bodrunova[4], and Nikolay S. Lyudkevich[4]

[1] Institute of Higher Nervous Activity and Neurophysiology of Russian Academy of Sciences, 5A Butlerova street, Moscow 117485, Russian Federation
kharlamov@analyst.ru

[2] Moscow State Linguistic University, 38 Ostozhenka street, Moscow 119034, Russian Federation

[3] Higher School of Economics, 20 Myasnitskaya street, Moscow 101000, Russian Federation

[4] St. Petersburg State University, 7/9, Universitetskaya embankment, St.Petersburg 199034, Russian Federation

a_v_orehov@mail.ru, s.bodrunova@spbu.ru, nikolai.lyudkevich@gmail.com

Abstract. Till today, classification of documents into negative, neutral, or positive remains a key task within the analysis of text tonality/sentiment. There are several methods for the automatic analysis of text sentiment. The method based on network models, the most linguistically sound, to our viewpoint, allows us take into account the syntagmatic connections of words. Also, it utilizes the assumption that not all words in a text are equivalent; some words have more weight and cast higher impact upon the tonality of the text than others. We see it natural to represent a text as a network for sentiment studies, especially in the case of short texts where grammar structures play a higher role in formation of the text pragmatics and the text cannot be seen as just "a bag of words". We propose a method of text analysis that combines using a lexical mask and an efficient clustering mechanism. In this case, cluster analysis is one of the main methods of typology which demands obtaining formal rules for calculating the number of clusters. The choice of a set of clusters and the moment of completion of the clustering algorithm depend on each other. We show that cluster analysis of data from an n-dimensional vector space using the "single linkage" method can be considered a discrete random process. Sequences of "minimum distances" define the trajectories of this process. "Approximation-estimating test" allows establishing the Markov moment of the completion of the agglomerative clustering process.

Keywords: Message clustering · Sentiment analysis · Markov moment · Tonality · Lexical mask · Approximation-estimation test · Short texts · Twitter

This research in the part of data interpretation and literature review has been supported by the Russian Science Foundation grant 16-18-10125-P (2016–2018, prolongued for 2019–2020).

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 18–31, 2019.
https://doi.org/10.1007/978-3-030-34770-3_2

1 Introduction

Sentiment analysis is a set of methods in computational linguistics for automatic identification of emotionally loaded lexicon in texts [1]. Tonality, or sentiment, refers to the emotional attitude of the author to some object of the real world, event, process, or their properties. The lexical sentiment is the emotional substance of the text expressed at the level of lexemes or their combinations. The tonality of a text can be defined as a function of the lexical tonalities of its constituent sentences [2].

Till today, classification of documents into negative, neutral, or positive remains a key task within the analysis of text tonality/sentiment. In modern systems of automatic determination of the emotive evaluation of the text, the one-dimensional emotive space is most often used: positive/negative (good/bad). One of the disadvantages of this approach is that the emotive component of the document cannot always be unambiguously determined. A document may contain signs of both positive and negative evaluation [1,3–5]. Today, successful use cases of multidimensional spaces are known [1,6]. It is possible to classify the polarity of documents and on a scale having a large number of measurements [7,8].

Initially, a fragment of unstructured text is examined using processing tools and algorithms of natural language. Then the objects and terms selected from this text are analyzed to understand their meaning [9]. The definition of tonality can be made using specific scale systems.

Another direction of the study of the sentiment of texts is the identification of subjectivity or objectivity. In this case, it is necessary to determine to which class of documents the text under investigation belongs (subjective/objective) [10]. This task can sometimes be more complicated than the classification of polarity. The subjectivity of words and phrases may depend on their context. For example, an objective document may contain subjective sentences (for example, a news article quoting differing opinions of people) [10,11].

2 Automated Sentiment Analysis: A Review of Approaches

The studies of tonality can be divided into two categories: tonality analysis by experts and automated sentiment analysis [12]. The most noticeable differences between them can be defined by the metrics of efficiency, accuracy, and performance. There are four main approaches to automatic sentiment analysis: using rules and dictionaries, supervised machine learning, non- or semi-supervised machine learning, and graph-theoretic (network) models; some researchers, though, put, e.g., neural networks inside supervised approaches [13], but we consider graph-theoretic (network-based) approaches a separate branch in sentiment detection research; it may include neural networks or may not, depending on the opinion of the researcher.

Methods for the classification of tonality based on rules and dictionaries consist in the search for emotive lexis according to previously compiled tonal dictionaries and rules. In this case, the text is analyzed according to the following scheme. First, each word in the text is assigned a value of its tonality from the dictionary (if it is present in the dictionary), and then the overall tone of the whole text is calculated by summing the value of the tonalities of each word [14]. The main problem of methods based on dictionaries and rules is the complexity of the process of compiling a dictionary. In order to obtain a method that classifies a document with high accuracy, the terms of the dictionary must have a weight adequate to the subject area of the document. For example, the word "huge" concerning the hard disk storage capacity is a positive characteristic, but negative for the size of a mobile phone. Therefore, this method requires considerable effort, as it is necessary to make a large number of rules. There are several approaches to automate the compilation of dictionaries and rules for a specific subject area [15].

Machine learning with a teacher (supervised learning) is a popular method for sentiment analysis. Its essence lies in the fact that, at the first stage, a machine classifier is trained (for example, a Bayesian one on previously marked texts), and then the resulting model is used in the analysis of new documents. The basis of machine learning without a teacher (unsupervised learning) is the idea that the terms that are more common in the corpus of texts have the highest weight in each text. By highlighting these terms and then determining their tonality, one can conclude about the tonality of the entire text [16].

The methods based on graph-theoretic models (or network models), too, use the assumption that not all words in the corpus of texts are equivalent. Some words have more weight and have a higher influence on the tonality of the text. We consider this assumption a key one that brings sentiment analysis closer to real-world construction of meaning in texts. When using this method, the analysis of tonality is divided into several stages. First, the graph is constructed based on the text under study. Then, the ranking of its vertices and the classification of the found words are made [17]. Then, when classifying words, a tonal dictionary is used, in which each word is assigned a grade, e.g. "positive", "negative" or "neutral."

The overall quality of the sentiment analysis system is assessed by comparing how well it agrees with expert opinion on the emotional evaluation of the text under study. For this, metrics such as accuracy and completeness can be used.

The formula for finding completeness is: $R = C/T$, where C is correctly agreed opinions of automatic analysis system and expert opinion, T is the total number of opinions both found and not found by the automated system.

A similar formula calculates accuracy: $P = C/\mathbf{T}$. However, in this case, $\mathbf{T}$ is the total number of opinions found by the automatic system. It is obvious that $\mathbf{T} \leq T$ [18].

Thus, the accuracy expresses the proportion of the studied texts for which, in the evaluation of tone, the opinion of the automatic sentiment analysis system coincides with human expert opinion. At the same time, according to the study, experts usually agree on the tonality of a particular text not in 100% of cases e.g., 79% of agreement has been reported [19]. Consequently, a program that determines the tonality of a text with an accuracy of 70% does so almost as well as a person. As the sentiment for a given word is context-dependent and dictionaries cannot always help in disambiguation, this, in its turn, poses the yet unresolved question of the baseline accuracy measurement to which the sentiment detection results are to be compared. Disambiguation studies, though showing considerable success [20], have their own limitations and do not resolve the issue completely. But this should not prevent the scholars from seeking better instrumentation for sentiment detection, having in mind the inequality of words within texts and the ambiguities of their meanings.

Adding to this, our paper discusses an approach to sentiment analysis in which unsupervised learning (cluster analysis) and graph-theoretic (network) models are combined.

3 Previous Works on Combinations of Clustering/sentiment Methods

Graph-theoretic approaches have become traditional for community detection in real-world communication networks [21,22]. Later, they have been applied to sentiment detection, based on the assumption that a text can be represented as a network of inter-related and structurally linked lexical units, or a graph.

Uniting cluster analysis with other approaches, including the network one, has, too, already grown into a small but distinct tradition in sentiment detection studies. But earlier studies mostly used community detection in real-world social media (see, e.g., [23]), topic modelling, hard or fuzzy clustering by k-means and c-means, and other types of classifiers (like naive Bayes, maximum entropy, and support vector machines see [24]), or several of these algorithms [25] on the level of a text, user, or the overall discussion network, and then applied sentiment extraction tools to the identified user, topics, or communities, respectively. Some of them tried to algorithmically enhance the resulting sentiment detection, e.g. by making it community-sensitive [23]. Coming closer to our approach h, network-based studies enhanced sentiment analysis via translation of sentiment from one language to another [26].

Only several works have successfully tried to unite several clustering/classification and network-based instruments to enhance the resulting extraction of classes of sentiment-loaded units of analysis. One paper [27] detects news frames in newspaper headlines, it combines k-means clustering on the level of a word and automated sentiment analysis. But it also uses manual labeling for cluster centroids, which we try to avoid. Another work [13] has suggested a hybrid cuckoo method to better detect centroids (called 'cluster-heads' in the paper) to improve sentiment detection.

Beyond these approaches, the one that deals with concept-oriented semantic networks has also produced promising results, as they allow to overcome the sematic weaknesses of statistical methods of text analysis [28]. They identify clusters of concepts rather than analyze frequencies of isolated words [29, p. 3623] (for earlier reference, see [30]). We follow this approach in essence, but use the idea of a lexical mask.

Also, as in most clustering techniques, stabilization of cluster centers (centroids) remains an unresolved problem. We here provide an approach that suggests cluster stabilization instrument using the Markovian moment.

4 Homogeneous Semantic Network and Analysis of Text Tonality

One of the ways of text representation for sentiment analysis is via using the idea of a lexical mask of tonality, built based on a homogeneous semantic network of the text.

First, we will formally describe the homogeneous (associative) semantic network [31]. Let us assume that are dealing with a pragmatically oriented corpus of X texts belonging to a particular subject area.

In another work of ours [32], we introduce a way to represent such a corpus as a directed graph of interrelated concepts, or a semantic network of subgraphs linking the basic concepts with their peripheries. The pre-procedure for defining this semantic network includes formation of the first-level dictionary as a linearly ordered alphabetical set of words with their defined frequencies; introduction of binary relations of "associativity" (different from the algebraic property of associativity) between the words and formation of paired syntagmas; representation of the paired syntagmas as directed graphs [33] that is, minimal oriented subgraphs with two vertices, one of which is the main word and the other is a word-associate, and the edge is directed from the main word to the word-associate; and introduction of the "asterisk" subgraphs that consist of the main vertices and all its vertices-associates where the edges are directed from the main words to the words-associates.

Mathematically, an asterisk $< g_i, < g_j >>$, is a directed subgraph that includes all ordered $< g_i, g_j >$ pairs, where the index i for main word g_i is fixed, and the index j of words-associates g_j runs through all its possible values. The number of edges in any asterisk subgraph is equal to the number of the vertices-associates. The set of all g_j (vertices-associates of the asterisk $< g_i, < g_j >>$) is called its periphery.

Given this, we may form the following definitions.

Definition 1. *The homogeneous (associative) semantic network N extracted from the text is a directed graph whose vertices are the concepts in the text. The edges of this graph represent the connections between concepts. These links correspond to paired syntagmas in sentences. Each edge is directed from the first concept to the second. Both the vertices of the network (graph) and the edges connecting them have ranks of importance in the text.*

Thus, a network can be represented in two ways, either as a set of pairs of words (syntagmas) or as a set of asterisks.

Definition 2. *The homogeneous semantic network N is the set of paired syntagmas (stable phrases) $< c_i, c_j >$ from the corpus of texts X, the weight characteristics of which are obtained during the iterative ranking procedure (see below)*

$$N = \bigcup < c_i, c_j > .$$

Definition 3. *A homogeneous semantic network N is the set of asterisks $z_i =< c_i, < c_j >>$:*

$$N = \bigcup_i z_i = \bigcup_i < c_i, < c_j >> .$$

The latter definition of a homogeneous semantic network is essential for ranking its vertices. The stronger the connection between the vertices of the graph, the higher the weight of the edge.

The vertices of the network (graph) are ranked during an iterative procedure [34], the initial state of which is the frequency of occurrence of words in the text and the frequency of the syntagmas in the text.

A lexical mask is a set of lexical markers corresponding to the vertices of the network. Lexical markers are determined by an expert in a given subject area. They characterize the tonality of the text on a scale from very negative to very positive. This corresponds to the following definitions:

Definition 4. *A lexical marker is a word or phrase from a text that characterizes the tone of a text of a particular direction. The lexical marker has a tonality of weight assigned by the expert on a scale from very negative to very positive.*

Definition 5. *A lexical mask is a set of markers identified based on the semantic network of the text, with weights of tonality ranked by the semantic network.*

Definition 6. *Text tonality is the sum of the weights of the tonality of all lexical markers of a lexical mask, ranked by their ranks in the semantic network.*

The initial statistical representation of the text is formed, when the network of words with their connections is re-arranged using the iterative procedure [34] similar to the Hopfield network algorithm. This allows moving from the frequency portrait of the text to the associative network of key text concepts:

$$\omega_i(t+1) = \sigma(\overline{E}) \sum_i \omega_i(t)\omega_{ij};$$

here, $w_i(0) = p_i,\ w_{ij} = p_{ij}/p_j\ (i \neq j)$ and $\sigma(\overline{E}) = (1 + e^{-k\overline{E}})^{-1}$ is a function normalizing to the average energy of all vertices of the network , where p_i is the frequency of occurrence of the i-th word in the text, p_{ij} is the frequency of joint occurrence of the i-th and j-th words in text fragments (sentences). The resulting numerical characteristic of words is their semantic weight. This weight characterizes the degree of their importance in the text.

For the typology of texts (with the help of cluster analysis), their mapping to the vector space of tonality is constructed using the lexical mask. In the vector space of tonality, the basis vectors correspond to a set of lexical markers. The norms of the basis vectors are equal to the ranks of the corresponding lexical markers [35].

5 Stability of Clustering

For cluster analysis of texts by their lexical tonality, we suggest to apply the hierarchical agglomerative algorithm of "single linkage" we have described previously elsewhere [36,37] developed from [38].

The main advantage of the "single linkage" method is its mathematical properties. The results obtained using this method are invariant to monotonic transformations of the similarity matrix. The method of "single linkage", in comparison with other clustering methods, performs with high stability. This method is especially useful in vector spaces. On the other hand, its main disadvantage is the occurrence of the so-called "chain effect" in the final stage of clusterization. As the end of the process approaches, one large cluster forms, as either previously formed clusters or isolated points join it. In the end, if there is no criterion for stopping the clustering process, all points of the set X merge into one cluster [38].

The decision on the moment of completion of the clustering algorithm (determination of the preferred number of clusters) can be approached within the framework of the theory of optimal stopping rules [39,40].

Consider the set of minimum distances obtained after $m-1$ iterations of the "single linkage" algorithm, to use this set to derive the statistical criterion for the completion of an agglomerative clustering process in the vector space $\mathbb{R}^n$. The set of minimum distances has the form $\{F_1, F_2, \ldots, F_{m-1}\}$ and is linearly ordered with respect to the numerical values of its elements: $0 \leq F_1 \leq F_2 \leq \ldots \leq F_{m-1}$.

When clusters merge or when an isolated point joins them, a sharp jump in the numerical value of the minimum distance should occur. The moment of this jump coincides with the moment of completion of the clustering process. This moment, on the graph of the minimum distance function, is more accurately approximated by an incomplete quadratic parabola (without a linear term), and not a straight line [36,37]. To determine the moment when the nature of the monotonous increase of a numerical sequence change from linear to parabolic, we use the previously constructed approximation-estimated test [41,42], described in detail for other purposes here [32].

To be used for sentiment analysis, the statistical criterion for the completion of the agglomerative clustering process, based on the method of "single linkage", can be formulated as follows.

Let $\{F_1,\, F_2,\, \ldots,\, F_k\}$ be a linearly ordered set of minimum distances. The set $\{y_1,\, y_2,\, \ldots,\, y_k\}$ is the "trend set" obtained using the $y_i = F_i + q \cdot i$, where q is the "trend coefficient", i is the iteration number of the $\mathcal{A}$ agglomerative clustering algorithm. The clustering process is considered complete at the k-th iteration, if for the nodes $y_{k-4}, y_{k-3}, y_{k-2}, y_{k-1}$, the inequality $\delta^2 \leq 0$, and for the set of points $y_{k-3}, y_{k-2}, y_{k-1}, y_k$ the inequality $\delta^2 > 0$.

When clustering the sample set $X \subset \mathbb{R}^n$ by the "single linkage" method, for a random sequence of minimum distances F_t, the natural filtering consistent with the process will be "a sample σ-algebra" $\mathcal{S}\,(\mathbb{R}^n)$ [36]. Then, by definition, the statistics τ will be the Markov moment of stopping the agglomerative clustering process: $\tau = \min\{t \in T \ \mid\ \delta_t^2 > 0\}$.

That is, the Markov moment of stopping the agglomerative clustering process is the minimum value of t, at which the null hypothesis of H_0 is rejected — the values of elements of a linearly ordered trend set increase linearly and an alternative hypothesis is accepted H_1 — the values of the elements of a linearly ordered trend set increase parabolic [36].

The clustering process is completed using an approximation-estimation test that evaluates the jumps of a monotonically increasing sequence of minimum distances. The magnitude of a significant jump sufficient to stop the clustering process depends on the sensitivity of the stopping criterion. The sensitivity of the criterion is given by the non-negative coefficient q [36,37]. The larger the value of q, the lower the sensitivity of the criterion for stopping the clustering process. The stopping criterion has the greatest sensitivity at $q = 0$. In this case, the clustering will result in the largest number of clusters. Increasing q can reduce the sensitivity of the stopping criterion so that the process continues until all m vectors are combined into one cluster.

Cluster analysis, in a certain sense, has a high degree of subjectivity. Therefore, the interpretation of its results depends mostly on the researcher him-/herself. In the works [43–45], the strict definition of stable clustering is not given, but only its intuitive notion is introduced. "The stability of clustering shows how different the resulting groupings are obtained after repeatedly applying clustering algorithms for the same data. A small discrepancy between the results is interpreted as high stability" [43, p. 87]. In our case, a quantitative measure of clustering stability can be considered the value of the interval of change of the coefficient q, at which the same result is obtained for the sample set X.

The main disadvantage of the "single linkage" method is the high probability of the occurrence of a "chain effect" and the formation of large elongated (in one or several dimensions) clusters. Therefore, the use of this method is appropriate for the typology of well-conditioned data. It is essential to note in this respect that the "chain effect" occurs at the final stage of the clustering process when already formed clusters are added one after another to some other cluster. In this case, the correct choice of the sensitivity of the stopping criterion due to the non-negative coefficient q is essential.

In the general case, a sequence of intervals of stable clustering, for different values of q, is denoted by: Q_1, Q_2, ..., Q_{e-2}, Q_{e-1}, Q_e, where Q_e is the set of values of the coefficient q for which all the m points are combined into one cluster.

6 Numerical Simulation

A series of numerical experiments were carried out to test the clustering algorithm with a cosine metric. To this end, a Python 3.7.0 program was written in the PyCharm integrated development environment.

The specificity of clustering, in this case, is that clusters are formed "in azimuth", i.e., for typology, the norm of a vector is entirely unimportant, while the angle between vectors and some fixed axis is decisive. For a flat case, this is usually the abscissa axis. As illustrative examples, consider the results of computational experiments for three different data sets.

The first set is C_1 (Fig. 1a).

A significant for clustering in vector space using the "single linkage" method with a cosine metric is the high sensitivity of the "approximation-estimating test". The division into small subclusters occurs at $q = 0$ in four azimuths, with a large number of isolated points (Fig. 1b).

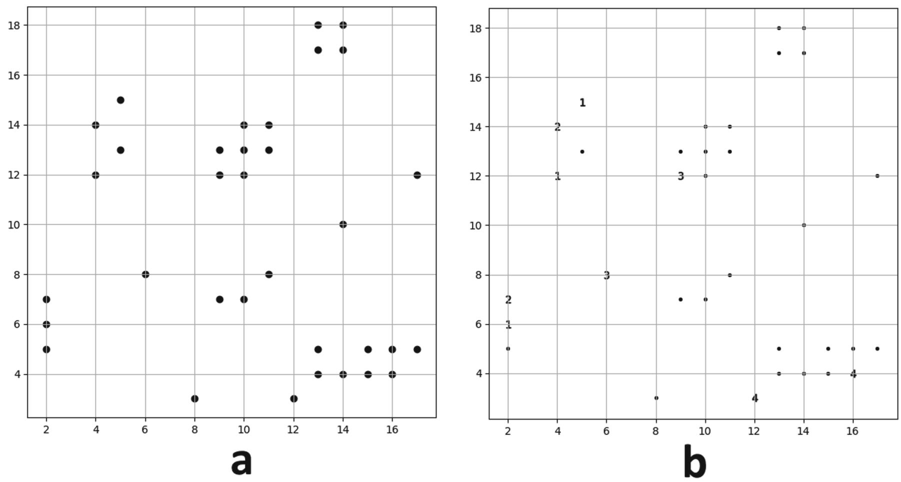


Fig. 1. **a** — is a set of points C_1; **b** — partition into subclusters at $q = 0$.

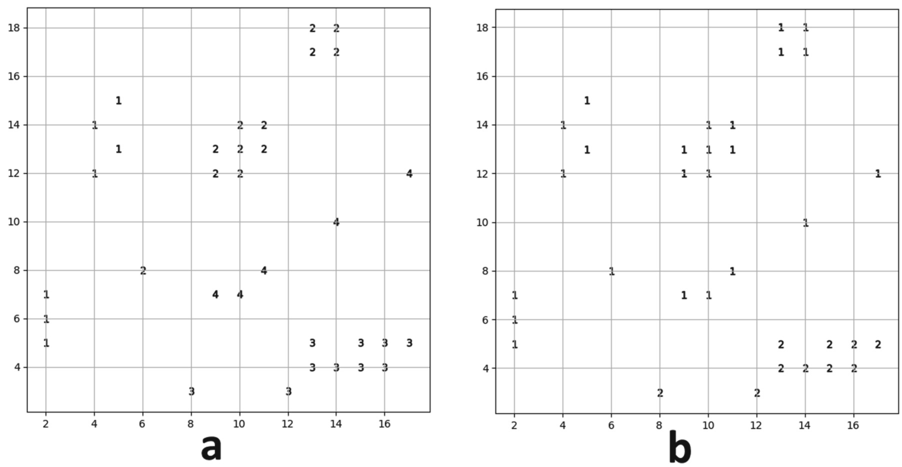


Fig. 2. **a** — splitting into an acceptable number of clusters without isolated points; **b** — occurrence of a "chain effect".

For $q \in [0.01, 0.045]$, we get a partition into an acceptable number of clusters without isolated points (Fig. 2a), and for $q \in [0.046, 0.49]$, the "chain effect" occurs (Fig. 2b). The Fig. 2b clearly shows that the clusters with labels 1, 2 and 4 (in the Fig. 2a) have merged into one cluster, the elements of which are marked 1 in this figure.

A similar picture of the formation of subclusters, clusters, and the "chain effect" is observed for the C_2 and C_3 datasets (Figs. 3, 4, 5, 6).

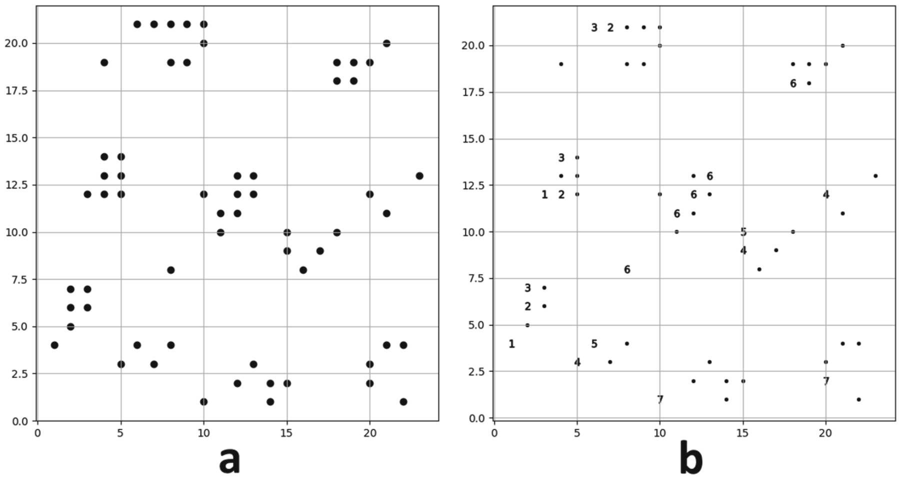


Fig. 3. **a** — is a set of points C_2; **b** — seven azimuth subclusters, $q = 0$.

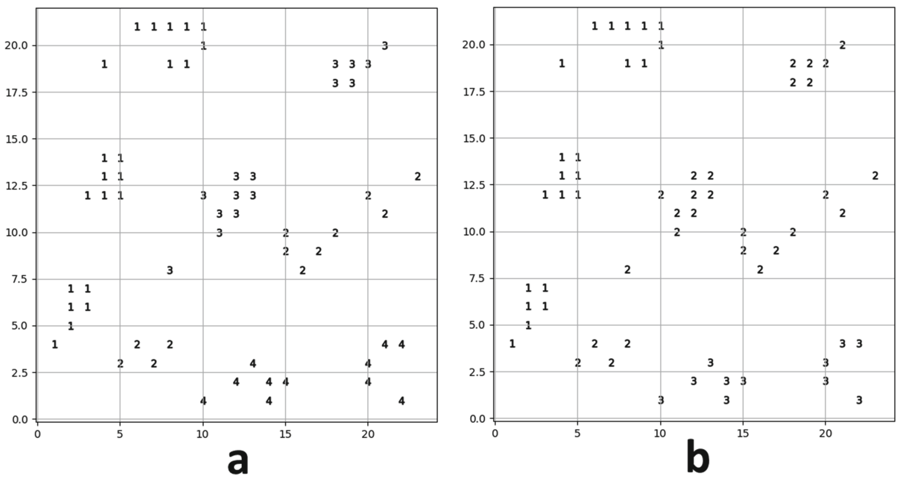


Fig. 4. **a** — acceptable number of clusters $q \in [0.01, 0.03]$; **b** — occurrence of a "chain effect" $q \in [0.04, 0.11]$.

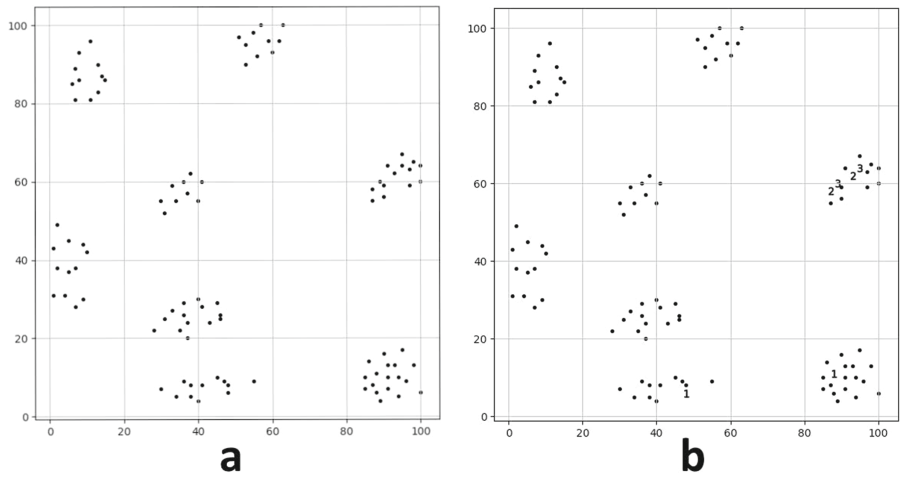


Fig. 5. **a** — is a set of points C_3; **b** — three azimuth subclusters, $q = 0$.

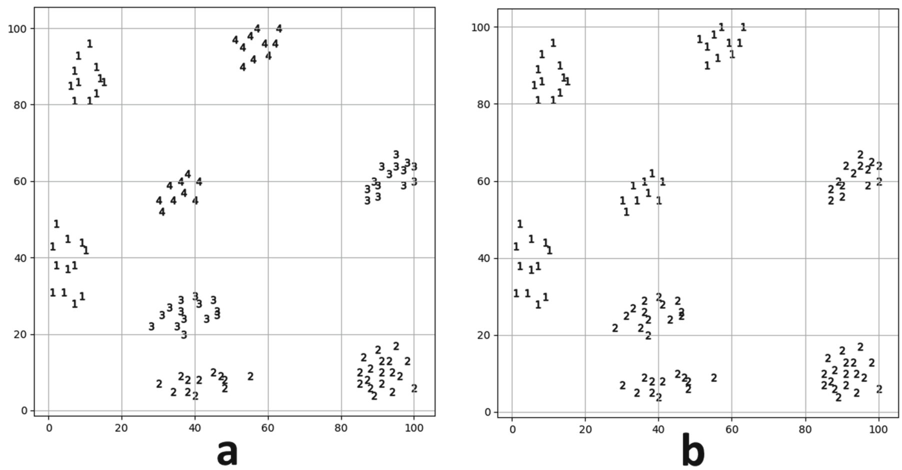


Fig. 6. **a** — acceptable number of clusters in four azimuths $q \in [0.01, 0.09]$; **b** — occurrence of a "chain effect" $q \in [0.10, 0.15]$.

7 Conclusion

The proposed combination of the clustering algorithm and the lexical network mask instead of the vector representation of the text allows for hoping for the efficiency of the suggested mechanism for typing texts according to their tonality. The transition to the network representation is natural in the problems of text analysis, as the use of a pure frequency vector representation may be incorrect from a linguistic point of view [32]. Clustering with a Markov moment, in turn, allows automating the procedure for determining the number of clusters in a given corpus of texts. Our work falls within the segment of academic literature that combines clustering algorithms with detection of sentiment, but suggests a stopping criterion for clustering as well as the use of lexical masks for the clustered texts, thus adding to the existing options for sentiment-based text classification.

A series of numerical experiments were carried out to test the clustering algorithm with a cosine metric. To this end, a Python 3.7.0 program was written in the PyCharm integrated development environment. Above, we have demonstrated by illustrative examples the results of computational experiments for three different datasets.

Based on the analysis of the above numerical experiment results and general considerations, we can formulate the following hypothesis for further studies:

Conjecture 1. "An acceptable number of clusters is formed when $q \in Q_{e-2}$" [37].

Of course, this hypothesis requires earnest study and discussion for various real-world datasets. This is especially true for short texts such as tweets, search engine queries, or news headlines where the number of edges in asterisks is highly limited. Further experiments are required to compare the quality of the suggested approach for longer- and shorter-text datasets.

Acknowledgements. This research in the part of data interpretation and literature review has been supported by the Russian Science Foundation grant 16-18-10125-P (2016–2018, prolongued for 2019–2020).

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**, 1–135 (2008)
2. Liu, B.: Sentiment analysis and subjectivity. In: Indurkhyaé, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, vol. 2, pp. 627–666 (2010)
3. Turney, P.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: Proceedings Association for Computational Linguistics, pp. 417–424 (2002). arXiv:cs.LG/0212032
4. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. Proc. EMNLP **10**, 79–86 (2002)
5. Bodrunova, S.S., Blekanov, I.S., Smoliarova, A., Litvinenko, A.: Beyond left and right: real-world political polarization in Twitter discussions on inter-ethnic conflicts. Media Commun. **7**(3), 119–132 (2019)
6. Bollen, J., Mao, H., Zeng, X.-J.: Twitter mood predicts the stock market. J. Comput. Sci. **2**(1), 1–8 (2011)
7. Pang, B., Lee, L.: Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings 43rd Annual Meeting Association for Computational Linguistics, University of Michigan, USA, June 25-30, 115–124 (2005)
8. Snyder, B., Barzilay, R.: Multiple aspect ranking using the good grief algorithm. In: Proceedings Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL), pp. 300-307 (2007)

9. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. J. Am. Soc. Inf. Sci. Technol. **61**(12), 2544–2558 (2010)
10. Su, F., Markert, K.: From words to senses: a case study in subjectivity recognition. In: Proceedings 22nd International Conference on Computational Linguistics, vol. 1, pp. 825–832, ACL (2008)
11. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity. In: Proceedings 42nd Annual Meeting, Association for Computational Linguistics, pp. 271–278 (2004)
12. Washington, E.: Human Sentiment Analysis. Growingsocialmedia.com. Last accessed 14 Nov 2013
13. Pandey, A.C., Rajpoot, D.S., Saraswat, M.: Twitter sentiment analysis using hybrid cuckoo search method. Inf. Process. Manag. **53**(4), 764–779 (2017)
14. Pazelskaya, A., Soloviev, A.: Method of determining emotions in texts in Russian. In: The International Conference on Computational Linguistics and Intellectual Technologies "Dialogue-2011", Moscow, 510–522 (2011) [in Russian]
15. Chetviorkin, I., Loukachevitch, N.: Extraction of russian sentiment lexicon for product meta-domain. Proc. COLING **2012**, 593–610 (2012)
16. Klekovkina, M.V., Kotelnikov, E.V.: The method of automatic classification of texts by tonality, based on the dictionary of emotional vocabulary. Trudy 14–i Vserossiyskoi nauchnoi konferentsii "Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kollektsii", pp. 81–86 (2012) [in Russian]
17. Ustalov, D.A.: Extraction of terms from the Russian texts using the graph models. In: Proceedings Conference Graphs Theory and Applications, pp. 62–69 (2012). [in Russian]
18. Kobayashi, N., Iida, R., Inui, K., Matsumoto, Y.: Opinion mining on the web by extracting subject-aspect-evaluation relations. In: AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, pp. 86–91 (2006)
19. Ogneva, M.: How Companies Can Use Sentiment Analysis to Improve Their Business. Mashable.com. Last accessed 13 Dec 2012
20. Rentoumi, V., Giannakopoulos, G., Karkaletsis, V., Vouros, G.A.: Sentiment analysis of figurative language using a word sense disambiguation approach. In: Proceedings International Conference RANLP-2009, pp. 370–375 (2009)
21. von Stieglitz, S., Dang-Xuan, L., Bruns, A., Neuberger, C.: Social media analytics-an interdisciplinary approach and its implications for information systems. Bus. Inf. Syst. Eng. **6**(2), 89–96 (2014)
22. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. Nat. Acad. Sci. **99**(12), 7821–7826 (2002)
23. Deitrick, W., Hu, W.: Mutually enhancing community detection and sentiment analysis on twitter networks. J. Data Anal. Inf. Process. **1**(03), 19–29 (2013)
24. Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci. **181**(6), 1138–1152 (2011)
25. Li, N., Wu, D.D.: Using text mining and sentiment analysis for online forums hotspot detection and forecast. Dec. Support Syst. **48**(2), 354–368 (2010)
26. Scheible, C.: Sentiment translation through lexicon induction. In: Proceedings ACL 2010 Student Research Workshop, pp. 25–30 (2010)
27. Burscher, B., Vliegenthart, R., Vreese, C.H.D.: Frames beyond words: applying cluster and sentiment analysis to news coverage of the nuclear power issue. Soc. Sci. Comput. Rev. **34**(5), 530–545 (2016)

28. Cambria, E., Poria, S., Bajpai, R., Schuller, B.: SenticNet 4: a semantic resource for sentiment analysis based on conceptual primitives. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2666–2677 (2016)
29. Kang, G.J., et al.: Semantic network analysis of vaccine sentiment in online social media. Vaccine **35**(29), 3621–3638 (2017)
30. Doerfel, M.L.: What constitutes semantic network analysis? a comparison of research and methodologies. Connections **21**(2), 16–26 (1998)
31. Kharlamov, A.A., Yermolenko, T.V.: Neuronetwork Environment (Neuromorphic Associative Memory) for Information Complexity Overcoming. Searching of Sense in Weak Structured Information Arrays. Part II. Information Processing in Hyppocampus. Informational Technology. 21(11), pp. 883–889 (2015) [in Russian]
32. Orekhov, A.V., Kharlamov, A.A., Bodrunova, S.S.: Network presentation of texts and clustering of messages. In: Proceedings International conference on Internet Science 2019 (INSCI2019), in this volume (2019)
33. Ore, O.: Theory of Graphs. American Mathematical Society, Providence (1983)
34. Kharlamov, A.A., Yermolenko, T.V.: Text analysis: linguistics, semantics, pragmatics in the cognitive approach. Int. Sci. J. "Manag. Syst. Mach." **6**(250), 29–33 (2015). [in Ukrainian]
35. Kharlamov, A.A.: Formation of the n-grammatic thematic model of the text. Speech Technol. Number **1–2**, 15–23 (2016). [in Russian]
36. Orekhov, A.V.: Markov moment for the agglomerative method of clustering in Euclidean space. Bulletin of Saint Petersburg University, Series Applied Mathematics. Computer Science. Control Processes, 15(1), 76–92 (2019). https://doi.org/10.21638/11702/spbu10.2019.106
37. Orekhov, A.V.: Agglomerative method for texts clustering. In: Bodrunova, S.S., et al. (eds.) INSCI 2018. LNCS, vol. 11551, pp. 19–32. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17705-8_2
38. Aldenderfer, M.S., Blashfield, R.K.: Cluster Analysis. Sage Publications, Thousand Oaks (1984)
39. Wald, A.: Sequential Analysis. John Wiley & Sons, Hoboken (1947)
40. Sirjaev, A.N.: Statistical sequential analysis: optimal stopping rules. Am. Math. Soc. **38**, 174 (1973)
41. Orekhov, A.V.: Criterion for estimation of stress-deformed state of SD-materials. AIP Conf. Proc. **1959**, 070028 (2018). https://doi.org/10.1063/1.5034703
42. Orekhov, A.V.: Approximation-evaluation tests for a stress-strain state of deformable solids. Bull. St. Petersburg Univ. Ser. Appl. Math. Comput. Sci. Control Process. **14**(3), 230–242 (2018). https://doi.org/10.21638/11702/spbu10.2018.304
43. Granichin, O.N., Shalymov, D.S., Avros, R., Volkovich, Z.: A randomized algorithm for finding the number of clusters. Automation and remote control **86–98**, (2011). [in Russian]
44. Shalymov, D.S.: Randomized method for determining the number of clusters on a data set. Sci. Tech. Bull. St. Petersburg State Univ. Inf. Technol., Mech. Optics **5**(63), 111–116 (2009). [in Russian]
45. Shalymov, D.S.: Algorithms for sustainable clustering based on index functions and stability functions. Stoch. Optimizat. Comput. Sci. **4**(1–1), 236–248 (2008). [in Russian]

Migration Issues in Russian Twitter: Attitudes to Migrants, Social Problems and Online Resources

Olga Borodkina(✉) and Vladimir Sibirev

Saint Petersburg University, Saint Petersburg, Russia
oiborodkina@gmail.com

Abstract. Social networks are a specific space for communication characterized by a high level of freedom of expression. International migration is one of the topics that attract the interest of Russian-language Twitter users. One of the most important issues is the attitude of the population towards international migrants, which impacts all spheres of Russian social networks; however, at the present time there is a lack of research on the Russian segment of social networking concerning migration issues. This paper deals with a reflection of the discussion on international migration issues in the Russian-language Twitter. The data was collected on a platform developed and supported by the Russian company Beensaid. Tweets were selected on the basis of the subject "migration" and its associated keywords and then analyzed. The results demonstrate that many Russian users have a negative attitude towards international migrants. The findings show that there are different kinds of social problems associated with international migration. On the Russian Twitter, the migration flow from Central Asia is mainly discussed; the cultural risks, the security risks connected with illegal migration and terrorist attacks are the main issues for many Russian people. Among the key problems also discussed on Twitter, human rights and violations of the immigrants' rights in the economic and social spheres in Russia are popular. At the same time, the research findings point to the significant role Twitter plays in disseminating information about various legal and social services, though online resources are not sufficiently used as social support resources for migrants in Russia.

Keywords: Twitter · International migration · Social problems

1 Introduction

International population movements characterize the modern world [5]. In the post-soviet space, as a result of globalization, the number of international migrants moving to Russia primarily for economic reasons is growing, which has led to the development of "superdiverse" metropolises, similar to those in many other countries [1, 21]. Russia is one of the leading countries receiving migrants. In 2017, Russia hosted the fourth-largest number of immigrants worldwide (around 12 million) [7]. Every year more than 500 thousand migrants come to Russia; according to official statistics, 575,158 international migrants arrived in Russia in 2016 [13]. If short-term visitors and illegal

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 32–46, 2019.
https://doi.org/10.1007/978-3-030-34770-3_3

migrants are also taken into account, experts estimate that on average there are between 10 and 14 million migrants in Russia at any one time [12].

Most immigrants are from CIS (Commonwealth of Independent State) countries, since Russia has a visa-free agreement with them. The main countries of origin for Russia are Ukraine, Armenia, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Tajikistan, Turkmenistan, and Uzbekistan. International migration influences the demographics of the country and covers the shortages in the labor market, especially in the sphere of low-skilled and low-wage labor. At the same time, international migration is associated with a number of social problems, primarily connected with illegal migration.

International migration in Russia has been actively studied by Russian researchers over the past 20 to 30 years. The most discussed issues are migration legislation and policy, labor migration, and the problems of social integration and cultural adaptation of migrants [10, 12, 14, 20]. At the same time, the migration situation in Russia is outside of the aegis of international studies. However, the issues related to migration and social networking have been little investigated. The studies on online social networks have at least three important directions. Firstly, online social networks allow migrants to maintain strong connections with relatives and friends; online social networks compensate for the lack of personal social networks and construct new virtual communication networks. Secondly, it is obvious that the role of social networks in the social adaptation of migrants is growing, and online social networks are beginning to fulfill the role of educational and informational resources, especially with regard to finding housing and employment. And, thirdly, online social networks have become the most important way to express attitudes towards migrants and migration issues openly, which impacts on migration and social policy. This study investigates the migration issues on Russian-language Twitter. This social network is a platform for personal interaction and information resources. While exploring the social network, special attention was paid to the following key points: attitudes towards migrants and the main social problems that users associate with migration.

2 Twitter as a Social Network

Twitter was chosen for the study for several reasons. Among the foreign social networks in Russia, Twitter is second after Facebook, and for this study Twitter is essential at least for two reasons. Firstly, it has a fairly wide network of Russian-speaking users outside of Russia, and secondly, as in many other countries, Twitter involves the most educated and politically active expert users; this last characteristic is typical not only for Russia; for example, Maireder [9] and Ausserhofer [9] argue the situation is the same in Australia. Individuals as well as groups are able to act as users on Twitter. In fact, Twitter has a lot in common with traditional media; it is focused primarily on the dissemination of information rather than only on communication.

Over the past few years, Twitter has become the subject of many international studies [4, 16, 22]. It provides a specific communicative space, which researchers connect with a new type of publicness, that is, the "personal public" [16]. Bruns and Burgess [4] identify three key ways of using Twitter. Firstly, "Twitter is used for … porting of events as they occur; the simple format of Twitter messages and the near-ubiquitous accessibility of the Twitter network …combine to make live tweeting a more important practice on Twitter than comparable live activities have been for previous social media platforms". Secondly, "Twitter is also used widely for ongoing discussion – and instant evaluation – of newsworthy events", and thirdly, "Twitter's coverage of newsworthy events also consists of significant amounts of broader commentary on current events, reflecting mainly the senders' own perspectives and intended more as markers of those perspectives than as formal contributions to debate" [4]. Schmidt [16] pays attention to three main elements of Twitter as a communicative space that enables the emergence of personal publics: (1) technological features and affordances; (2) social and textual relations; and (3) shared rules.

In Russia, as in other countries, Twitter is increasingly used for interpersonal communication as a source of real-time information. It has become a space for debating news, politics, sport, recreation, etc. On the Russian Twitter, 1.2 million authors "generate" 78 million tweets per month. Twitter is the only Russian social network where male authors predominate; 55.4% of tweet authors are men [18]. In this study the focus was on the discussions concerning migration issues only on the Russian-language Twitter.

3 Research Methods

The analysis of the content of the Russian segment of Twitter was carried out in order to evaluate public opinion on the international migration and social problems associated with migration. The social network (Big Data) was monitored on a platform developed and supported by the Russian company Beensaid. A total of 13,200 entries were made on Twitter between Nov. 2017 and Feb 2018[1]. The tweets were selected on the basis of the subject "migration" and its associated keywords (migrant, immigrant, international migration). The location was determined according to the Twitter API (application programming interface) then a selective control of the degree of compliance of these data with the content of the tweet was carried out. Additional results of Twitter were achieved by monitoring tags. The correlation matrix was constructed based on dichotomous communication measures to analyze the relationships between the tags. Pareto's principle was applied to remove the insignificant weak links from the network graph: 20% of the strongest links determined 80% of the content of the graph; the similarity between tags was measured on base of cosine similarity. The sentiment analysis of tweets was determined by the SVM (Support Vector Machine) machine classifier, pre-trained on public texts of communicative messages on this topic. In addition to the standard information about each tweet (time of publication, author, place), the keywords (tags) characterizing its content were highlighted during the process of text mining. To visualize the relationships between the tags, network analysis methods were used and the classification of tweets by content based on the modularity method was also conducted. To assess the relationship between the tone of the tags, as well as the peculiarities of the content of tweets from different regions of the

[1] The empirical data were collected in the project "Social risks of the international youth migration in contemporary Russia" (No. 16-18-10092) supported by the Russian Science Foundation.

Runet (Russian-language Internet), as the method of multidimensional statistical analysis, correspondence analysis was used. Correspondence analysis allows the exploration of conjugation tables (tags key tone, tags country, etc.) by graphically representing rows and columns of the table as points in a low-dimensional space.

4 Research Results

4.1 Attitude Towards Migrants Among Twitter Users

The collapse of the Soviet Union led to an intensification of migration processes and the formation of a migration system in the post-Soviet space. During the post-Soviet decades, the configuration of migration flows has changed significantly. On the one hand, Russia is one of the biggest recipient countries. Labor migration, largely consisting of citizens of Ukraine, Kazakhstan, Uzbekistan and other former Soviet republics, has come to replace the forced mass migration of the Russian-speaking population of those countries. On the other hand, a significant percentage of Russian citizens left Russia; the main destinations were the US and Western European countries. According to official data on the 1st of January, 2015, a total of 2,162,230 Russians were living in America, 726,900 in Europe, and 15,385,600 in the countries of the former USSR [13]. Wilken [22] argues that "of crucial importance in the present context, is that questions of location and location-awareness are increasingly central to our contemporary engagements with the Internet and mobile media". There were the reasons to expect that the attitudes of Russian Twitter's users towards migrants in different countries were varying, since in Russia they are host community, and in others countries there are immigrants. The data in Fig. 1 concerning the perception of migrants by Russian Twitter users living in Russia, Europe, the United States and the countries that were former Soviet republics demonstrate opposite situation.

The attitude of Russian users of Twitter living in different countries is very similar. The sentiment analysis demonstrates mostly negative attitudes toward migrants. (36.3% of tweets in Russia, 36.2% in the USA, 34.8% in ex-USSR countries, 34.4% in the countries of the European Union). As for the positive attitude, this indicator is also very close in different countries: in Russia 29.1% of messages, in the EU countries 27.8%, in the ex-USSR countries 27.7%, and 20.0% in the United States, the lowest score. However, the reasons for the intolerant attitude among Russian-speaking users of Twitter living in Russia and those living in other countries are different.

Previous studies show that in Russia, among the social problems associated with international migration, cultural risks are considered the most problematic [3]. The cultural distance between the host community and migrants is a very important issue, and it is manifested in the different languages, in the differences of everyday behavior culture, traditions and religions and in increased attention to the perceived threat to the norms and values of the local population while the risks associated with employment and violence remain on the periphery of public attention. In addition, the negative attitude towards migrants in Russian society is connected with security; increasing levels of crime and ethnic conflict situations as well as terrorism attacks are strongly connected with immigration [2]. As for the Russian-speaking population of other

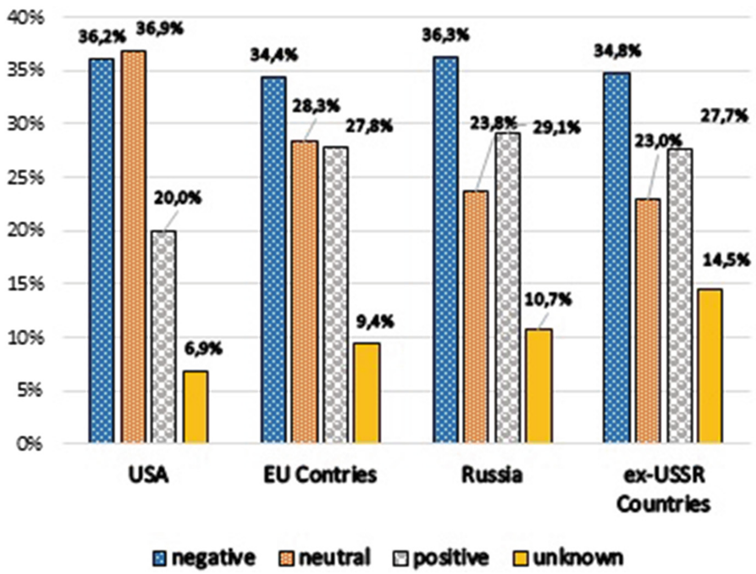


Fig. 1. The attitudes of Twitter users living in different countries towards migrants.

countries, particular in the US and EU, they themselves are migrants and perceived primarily as competitors in the labor market and in everyday life, and this is a key reason for negative relationships with international migrants.

4.2 Russian Twitter in Ukraine: Migration Issues

To analyze the Russian-language Twitter, it is important to look at how migration problems are perceived by Ukrainian Twitter users. Ukraine is one of the leading countries of origin for international migration to Russia. In 2017, 150,182 migrants from Ukraine arrived in Russia for the period of 9 months and longer [13]. In addition, the events of recent years, especially those related to the military conflict in such Ukrainian regions as Donbass and Lugansk, led to an increasing number of refugees from that country. On January 1, 2018, of the 125,442 persons to receive asylum in the Russian Federation, 123,434 were from Ukraine [19].

The continuing migration outflow of the population from Ukraine is associated with the economic and social policies that have led to a significant level of poverty, the reduction of the well-being of the population, lack of social opportunities, social uncertainty, and deterioration of health care and education. After the introduction of visa-free travel between Ukraine and the EU countries (June 2017), the labor migrant flow from Ukraine to Russia has greatly decreased. Many Ukrainian citizens now choose to migrate to EU countries, primarily Poland, Germany and Italy. However, Russia still remains the second-most popular country for Ukrainian citizens, and the most international migrants into Russia between 2014 and 2017 were Ukrainians: about 200,000 in 2014 and 150,000 in 2017 [17].

As for the attitude to migrants among Russian Twitter users in Ukraine, 34.9% of them are negative. A significant proportion of users have positive (26.2%) or neutral (23.8%) attitudes (Fig. 2). This is probably due to the fact that a lot of citizens of Ukraine themselves have the status of being migrants.

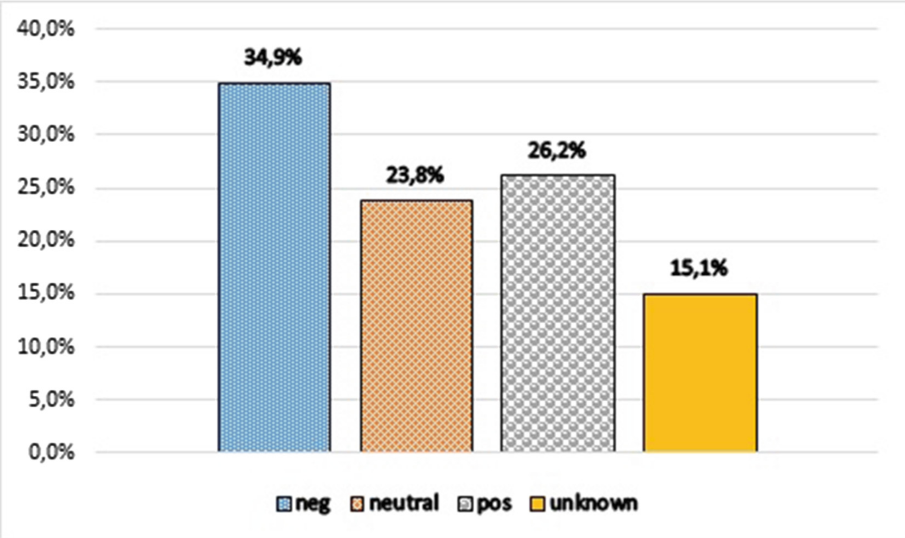


Fig. 2. Attitudes towards migrants among Twitter users in Ukraine

It is important to note that the attitude towards migrants among Twitter users in Ukraine does not differ significantly from the Russian Twitter. To determine the differences in the messages in the Russian and Russian-speaking Ukrainian users of Twitter, correspondence analysis was carried out (Table 1, Fig. 3).

Table 1. Key tags on migration in Russian Tweeter (% of Russian segment messages containing relevant tags).

Teg	%	Teg	%
Islamic State	↑ 100,0%	Job	↑ 93,6%
CIS	↑ 100,0%	Help	↑ 93,2%
Damage	↑ 100,0%	Migrant	→ 90,0%
Language	↑ 100,0%	Crime	→ 89,3%
Process documentation	↑ 98,5%	Police	→ 87,1%
Passport	↑ 97,7%	Terrorist act	→ 84,5%
Payment	↑ 96,0%	Homeland	↓ 78,7%
Citizen	↑ 95,9%	Refugee	↓ 78,0%
Central Asia	↑ 94,4%	Ukraine	↓ 62,8%
Family	↑ 93,9%		

As expected, the most common keyword in Ukrainian tweets is "Ukraine" (38% of messages with this keyword). Since, despite the conflict between them, Russia and Ukraine are closely connected and the flow of migrants from Ukraine to Russia remains significant, 62% of the messages with the tag "Ukraine" are in Russian tweets. The tags "refugee" and "homeland" (22%) are also quite common in Ukrainian messages, which reflect the complex political processes taking place in Ukrainian society. The problems of migrants' employment, their registration and the increasing number of crimes associated with migrants are of primary concern to Twitter users in Russia, although the last topic is relevant also for Ukraine.

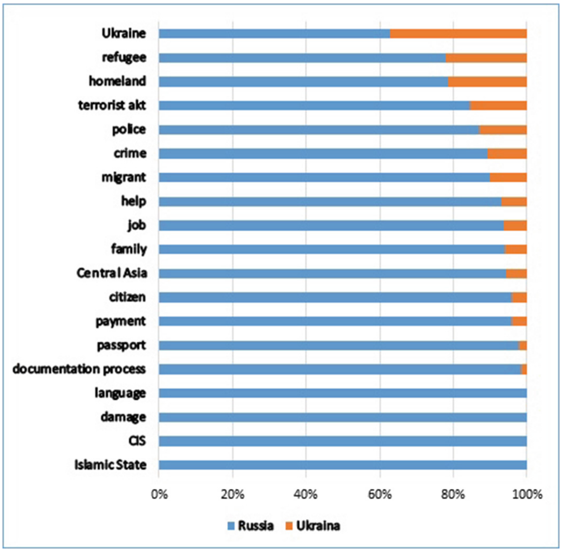


Fig. 3. The difference in Russian and Ukrainian Twitter messages on migration (The percentage of messages in Russia and Ukraine).

It should be noted that four tags are present only in Russian tweets. The first is *CIS* (Commonwealth of Independent States). This is a really important issue for Russia since the majority of international migrants arriving in the Russian Federation come from CIS countries but for Ukrainians, this is an irrelevant geopolitical association. Ukraine did not initially ratify the CIS Charter so it was never a de jure CIS member state, although it did take part in the work of its collective bodies. On May 19, 2018, the president of Ukraine Petr Poroshenko signed a decree revoking all representatives of Ukraine from all CIS statutory bodies. The National Security and Defense Council of Ukraine instructed the Ministry of Foreign Affairs of Ukraine to take, in accordance with the established procedure, measures to terminate the operation of certain interstate agreements. So, CIS is not important issue for Twitter users in Ukraine.

Language is the second tag that is only in Russia's Twitter. For many Russians, the cultural difference of foreigners, especially from the countries of Central Asia, is one of the main reasons for their intolerance towards migrants. Their ignorance of the Russian language is interpreted as an unwillingness to accept the culture of the host society and as imposition of their own culture, which is alien to Russians. In Ukraine, the migration situation is different. Since Ukraine does not attract migrants, there are very few international migrants and, therefore, the issue of language in connection with migration is not discussed in social networks.

The third tag is *damage*. According to public opinion in Russia, international migration is connected with "damage" to Russian society. Many Russian scholars and experts support this point of view, and highlight the negative social consequences of mass immigration to Russia – the failure of cultural integration, the negative impact on the education system, especially schools, the additional burden on the health care system, the impact on the consumer sphere, and the growth of ethnic crime and conflict tension [12]. For Ukrainians, migration is a short or long-term way to improve their financial condition and fulfill career expectations, and as a result, there is no tag "damage" in Ukrainian tweets.

The last tag that is present only in the Russia's Twitter is *ISIS* (The Islamic State is a terrorist organization banned in Russia). This fact can be explained by the social-political context. For Russians, the threat of terrorism has been one of the main social risks for at least the last two decades, and in recent years, Russian public opinion, partly owing to the influence of media, associates terrorist attacks with Islamic migrants. In Ukraine, the main content of public discussions is related to the confrontation with Russia, therefore the topic of the Islamic State is not a key tag in social networks either.

Thus, despite the considerable similarity of the content of tweets in Russia and Ukraine, there is big difference: – in Russian society, international migration is the source of much stronger social tension than in Ukraine. This is due to the socio-political conditions of the two countries. Russia is a host country, and for the Russian population, migration is strongly associated with immigration from Central Asia. In contrast, Ukraine is a country of origin, and Ukrainian citizens interpret migration as emigration and a way to improve their well-being.

4.3 Social Problems of Migration Identified on Russian Twitter

To identify the main problems associated with migration, an additional study was conducted. On the platform https://netlytic.org/home/, all the messages on the Russian part of Twitter, including # V Migrant # Migration, were collected. During February and March 2018, 57 users and 154 hashtags related to migration were identified, and then a content analysis of the tweets was done. As result of this analysis, a cloud of hashtags describing the main problems in migration was constructed (Fig. 4). The size indicates the frequency of use of this hashtag in the analyzed text.

Fig. 4. The tags on the key problems of migration on Twitter.

The findings show that mainly the migration flows from Central Asia are discussed (# Uzbekistan, # Tajikistan, # Kyrgyzstan) on Russian Twitter. Migration from Ukraine is also one of the key topics, and it is the result of the significant increase in the number of refugees from Ukraine to Russia (# Ukraine, # refugees). The content of the tweets demonstrates that migration has a great influence not only on the receiving country (Russia) but also on the countries the international migrants have left, and the political and economic issues of these counties are the focus of much discussion (# society, # politics, # economy, # employment). Twitter users pay a lot of attention to changes in migration legislation (# laws, # court, # immigration), human rights and the violation of the rights of migrants in the Russian Federation (# Russia rights, # human rights). Sensitive issues such as crimes committed by migrants are highlighted and, according to public opinion, special attention is given to illegal migration and terrorist attacks (# Crime, # terrorism, # illegal migration). Religion is also mentioned in migration discussion (# religion) but at a low frequency. The topic occupies one of the central places since a significant proportion of migrants in Russia are from such states as Uzbekistan, Tajikistan, Kyrgyzstan, and Kazakhstan (these countries are key tags), where the Muslim religion dominates [2].

4.4 Correspondence Analysis of Tweets About Migration Issues

To describe the relationship of different keywords with the attitude to migrants on Twitter, it is appropriate to use correspondence analysis [6]. The main purpose of the matching analysis is to move from the original data matrix to a new simpler matrix while losing as little information as possible. The correspondence analysis allows the presentation of the obtained results graphically ("maps of correspondence"), which greatly facilitates the interpretation of the links. As a rule, in the practical use of correspondence analysis, a two-dimensional solution, which is the most visual and not too difficult to interpret, is chosen. In this case, the categories of the analyzed variables having similar distributions are represented by points lying close to each other; and, in contrast, categories with widely differing distributions form points far apart from each other. Accordingly, close tags are located next to each other but very different tags are far apart. A similar strategy is applied to relationships with migrants. At the same time, the keywords, tinged with "negative", "positive" or "neutral" meaning, are concentrated around these categories. The result of the correspondence analysis on migration issues on Twitter is presented in Fig. 5.

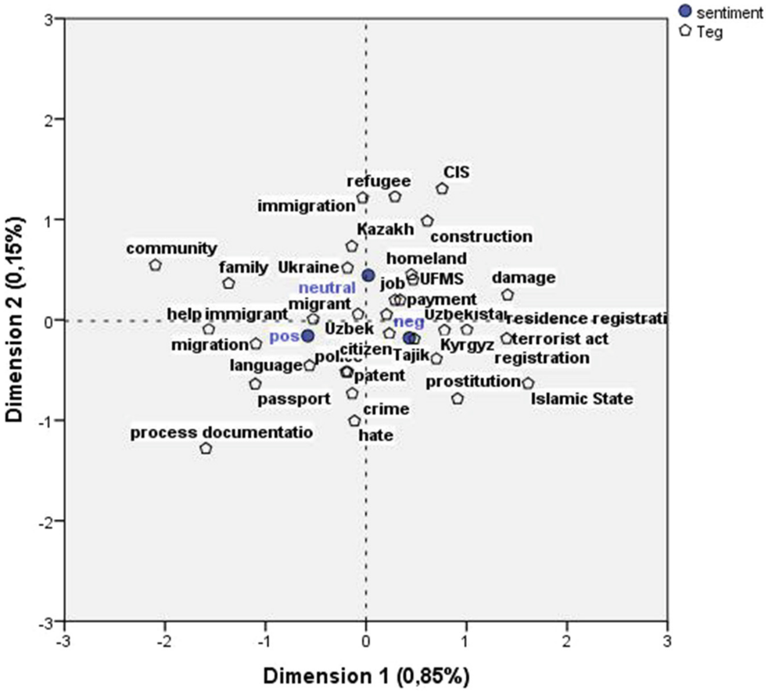


Fig. 5. The results of correspondence analysis ('symmetric' scaling); tags and attitudes towards migrants are represented in a two-dimensional space.

This map of correspondences describes the original data rather well. The first horizontal axis (85% of the total variations) can be interpreted as an emotional attitude towards migrants; on the left there is positive, on the right there is negative. Accordingly, the second vertical axis (15%) reflects an assessment of the effect of international migration on Russian society; at the top the positive effect is presented, at the bottom is the negative one.

In the first upper right quadrant, there are tags describing the negative aspects of attracting migrants to the Russian labor market. In the second right lower quadrant came the tags that characterize the most important negative problems that worry and concern all segments of Russian society. The main concerns are the Islamic State and related terrorist attacks, as well as criminal offenses related to migrants. The third lower left quadrant contains tags describing the organizational problems of regulating migration flows, including sanctions for violations of migration legislation. The fourth upper left-hand quadrant describes the existing social integration practice of international migrants in Russian society, and particular attention is paid to the problem of the observance of their rights.

4.5 Migration Issues on Twitter: Russian Regions

According to the geographical distribution of Twitter in Russia, Saint Petersburg (2.28%) is in first place, Moscow (2%) is second with a slight lag, and the Republic of Sakha (1.5%) is third [18]. There are significant regional differences in Russia: in the level of economic development, the welfare of the population. One of the research questions was how the population of different Russian regions relates toward international migrants (Fig. 6).

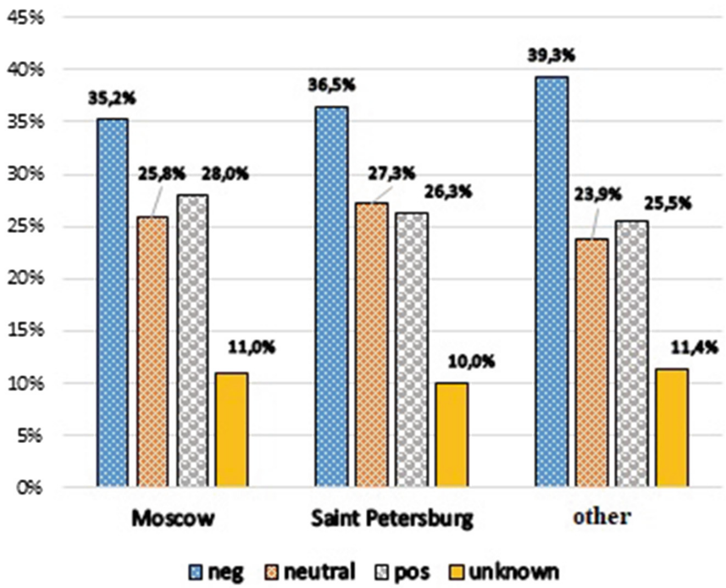


Fig. 6. The attitudes of Twitter users living in different Russian regions toward migrants.

An analysis of the attitude towards migrants of Twitter users in Moscow, Saint Petersburg and other Russian cities demonstrates that migration is very sensitive issue for all Russian regions. The socio-cultural and economic difficulties associated with the arrival of foreign citizens to earn money are supplemented by problems of interregional migration in modern Russia. The trend of displacement of the inhabitants of the northern part of the country, Siberia and the Far East to the central and southern regions is rather strong now, and makes the economic and social problems connected with migration deeper.

More than half of the internal and external migrants try to settle in the central part of Russia, firstly in Moscow and Saint Petersburg. Many people believe that international migration is connected with employment problems, increasing levels of crime, and the expansion of different cultures and traditions. For these reasons the majority of Twitter users in Moscow, Saint Petersburg and other regions demonstrate a negative attitude towards migrants (35.2%, 36.5% and 39.3%, respectively). The positive impacts of migration processes, including contributions to improving the demographic situation in Russia and satisfying the existing needs for labor resources, are not evaluated by the majority of Russian citizens. The differentiation of Twitter content across Russian cities is demonstrated by results correspondence analysis in Fig. 7.

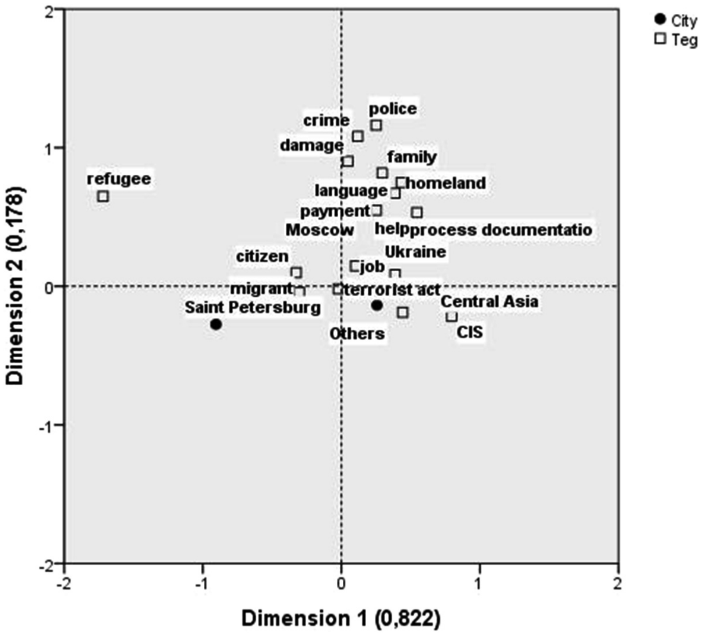


Fig. 7. Results of correspondence analysis of tweets across Russia ('symmetric' scaling).

Moscow plays the dominant role in the information sphere (almost more than 28% of total number of twitters); which is to be expected since the largest proportion of migrants are in Moscow. Saint Petersburg accounts for only 8.5% of the tweets for the analyzed period. The most discussed problems (particularly in Moscow) can be divided into two main groups: on the one hand, the socio-economic adaptation of migrants (family, help), and on the other hand, the negative issues connected with migration (crime, damage). In other Russian cities, the main topics are labor migrants and Central Asia. The problem of refugees from Ukraine stands alone and it is mainly discussed in Moscow and Saint Petersburg. This is expected since these cities were a significant number of refugees from there. Understanding the capabilities, resources and needs of international migrants in different Russian regions is essential for the development of regional and national public policies.

5 Discussion

Social networks allow the identification of the key social problems that are pressing concerns in the contemporary world, and international migration is one of them. Social networks need to be thoroughly studied since they reflect public opinion about the migration issue, and it is important that online social networks as a communicative space are characterized by freedom of expression as it is particularly important for countries with limited democratic institutions.

The monitoring of the Russian-language Twitter indicates a set of social problems connected with international migration. Firstly, Russian society is not tolerant enough of the international migrants; and for Russian citizens international migration is strongly associated with different social risks, such as cultural expansion, economic problems, competition in the labor market, and criminal activities. It should be noted that public opinion does not always reflect the real situation. For example, the level of crimes committed by migrants is no higher than the level of crimes committed by Russian citizens. In this regard, migration experts quite often blame the media for creating a negative image of migrants that influences public opinion. Secondly, that discrimination against migrants and violations of their human rights are widespread practices in Russia is recognized by the majority of social networks users. This issue is closely related to the topic of undocumented migration since the lack of a certified identity gives opportunities for frequent violations of human rights [15]. In Russia the existence of a sizable segment of undocumented residents highlights not only the limits of immigration control measures but also police corruption [8] and bureaucratic barriers. Thirdly, the analysis of messages in the social networks has demonstrated the similarity of attitudes towards migrants and migration from the Russian population and the Russian-speaking communities in other countries. This is an interesting point that requires further research.

There are some limitations of our study. The sample presents Russian-speaking users of Twitter, so it did not reflect the opinions of the significant part of immigrants in Russia from the countries of Central Asia who do not use Russian in online communication. Similarly, the comparison of data for Russia and Ukraine is based on the Russian-speaking segment of Ukrainian social networks. Nonetheless, examining how Russian users relate to international migrants and with which social problems they have associated migration informs our understanding of the migration processes and their social affects as well as the multi-faceted place of social networking in the migration dynamic.

6 Conclusion

Today's Russian migration policy meets modern demographic, social and economic realities as well as the prospects of national development [14], and this situation is reflected in the migration discussion in social networks. At the same time, the Russian authorities do not use social networks to construct a more tolerant society with regards to migrants. The negative image of migrants constructed by mass media [11], is still strong and it is been disseminated in social networks. To change public opinion, it is necessary at least to provide information about the contribution of international migrants to the economic and social development of the country. It might be well for Russian policy to use the social networks and their structure to prevent and reduce the social risks associated with international migration, specifically cultural, economic and security risks. The most popular social networks, including Twitter, could be providers of information about human rights, job opportunities and social assistance for international migrants.

The findings of this research contribute to the literature on migration debates in social networks by identifying the main problems and resource systems for international migrants on the Russian Twitter. This study offers a new dimension connecting with social networking for further research on the way international migration processes are considered on the global networks.

Acknowledgments. The research was conducted in Saint Petersburg University with the support of the Russian Science Foundation, the project № 19-18-00246 “Challenges of the transformation of welfare state in Russia: institutional changes, social investment, digitalization of social services”.

References

1. Bilecen, B., Gamper, M., Lubbers, M.: The missing link: social network analysis in migration and transnationalism. Soc. Netw. **53**, 1–3 (2018). https://doi.org/10.1016/j.socnet.2017.07.001
2. Borodkina, O.I., Sokolov, N.V., Tavrovskiy, A.V., Pankratova, L.S.: Social work with Muslim migrants in Russia: public opinion issues and practical perspectives. J. Relig. Spirit. Soc. Work. Soc. Thought **36**(1–2), 215–246 (2017). https://doi.org/10.1080/15426432.2017.1311243
3. Borodkina, O., Romanenko, V.: The construction of risks by residents of St. Petersburg in the context of international migration. In: Proceeding of the 4th International Multidisciplinary Scientific Conference on Social Sciences and Arts. Book, vol. 3, no. 3, pp. 767–774. SGEM, Albena (2017)
4. Bruns, A., Burgess, J.: Researching news discussion on Twitter: new methodologies. J. Stud. **13**, 801–814 (2012). https://doi.org/10.1080/1461670X.2012.664428
5. Castles, S., Miller, M.: The Age of Migration: International Population Movements in the Modern World. Palgrave Macmillan, Basingstoke (2003)
6. Greenacre, M.: Correspondence Analysis in Practice. CRC Press, Florida (2017)
7. International Migration Report. United Nations. New York (2017). https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2017_Highlights.pdf
8. Light, M.: Policing migration in Soviet and Post-Soviet Moscow. Post-Sov. Aff. **26**, 275–313 (2010). https://doi.org/10.2747/1060-586X.26.4.275
9. Maireder, A., Ausserhofer, J.: Political discourses on Twitter: networking topics, objects, and people. In: Weller, K., Bruns, A., Jean Burgess, J., Mahrt, M., Puschmann, C. (eds.) Twitter and Society, pp. 305–318. Peter Lang, New York (2014)
10. Mukomel, V.: Migration from central Asian countries: the challenges for the receiving societies. East Eur. Stud. **7**, 76–92 (2016)

11. Romanenko, V.: Image of international migration in Russian online mass media. In: Bodrunova, Svetlana S. (ed.) INSCI 2018. LNCS, vol. 11193, pp. 66–77. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01437-7_6
12. Remizov, M.: Needinaya Rossiya. Etnograficheskij otchet. Non-united Russia. Ethnic policy reports. Knizhnyj mir, Moscow (2015). (in Russian)
13. Rossiya v cifrah. Naselenie. Demografiya. Migraciya. Russia in Figures. Population. Demography. Migration. http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/demography/. Accessed 10 Sept 2018
14. Ryazantsev, S.V.: Foreign migration policy of Russia. Int. Trends **14**(47), 22–29 (2016). https://doi.org/10.17994/IT.2016.14.4.47.2
15. Sciortino, G.: The regulation of undocumented migration. In: Rath, J., Martiniello, M. (eds.) An Introduction to International Migration Studies, pp. 351–378. Amsterdam University Press, Amsterdam (2012)
16. Schmidt, J.-H.: Twitter and the rise of personal publics. In: Weller, K., Bruns, A., Burgess, J., Mahrt, M., Puschmann, C. (eds.) Twitter and Society, pp. 3–14. Peter Lang, New York (2014)
17. Shcherbakova, E.M.: Migraciya v Rossii. Predvaritel'nye itogi 2017. Migration in Russia. Preliminary results of 2017. http://www.demoscope.ru/weekly/2018/0763/index.php
18. Social networks in Russia, summer 2017: numbers and trends. https://blog.br-analytics.ru/sotsialnye-seti-v-rossii-leto-2017-tsifry-i-trendy/
19. Statistics: people with the status of temporary asylum in the Russian Federation (2018). https://refugee.ru/news/statistika-v-2017-godu-vremennoe-ubezhishhe-poluchili-vdvoe-menshe-lyudej/
20. Tyuryukanova, E.: Labor migration in Russia. DemoskopWeekly **8**, 315–316 (2008). https://doi.org/10.32609/0042-8736-2017-3-40-57
21. Vertovec, S.: Superdiversity and its implications. Ethn. Racial Stud. **30**(6), 1024–1054 (2007). https://doi.org/10.1080/01419870701599465
22. Wilken, R.: Twitter and geographical location. In: Weller, K., Bruns, A., Burgess, J., Mahrt, M., Puschmann, C. (eds.) Twitter and Society, pp. 155–167. Peter Lang, New York (2014)

Personal Illness Experience in Russian Social Media: Between Willingness to Share and Stigmatization

Victoria Dudina[1(✉)], Darja Judina[2], and Konstantin Platonov[2]

[1] Department of Sociology, Saint Petersburg State University, Saint Petersburg, Russia
v.dudina@spbu.ru

[2] Center for Sociological and Internet Research, Saint Petersburg State University, Saint Petersburg, Russia
d.yudina@spbu.ru, konplatonov@gmail.com

Abstract. Social media creates a public space where people with different illnesses can communicate with others experiencing the same conditions and discuss issues that were discussed only in private before. The willingness to share personal health information and private experience is important for successful peer-to-peer health communication but often is prevented by the fear of being stigmatized. The goal of our research was to understand the relationships between the willingness of people with different illnesses to share personal experiences in online communities and the extent of illness stigmatization. We compared online communities devoted to cancer, diabetes, hepatitis B and C, HIV and tuberculosis on the most popular Russian-language social networking site "Vkontakte". We selected open groups with memberships of more than 1000 users. For selected communities we measured the share of personal stories, the proportion of deactivated or anonymous profiles among users sharing personal experiences and community engagement rates. We found some dependencies between stigmatization of disease and users' willingness to discuss their personal experiences online. The share of personal stories in the hepatitis and HIV communities was smaller, than in oncology and diabetic communities. But tuberculosis, which occupies a middle position on the scale of stigmatization, showed the largest share of personal stories. The most anonymity was revealed in the communities of highly stigmatized illnesses. We revealed also that a greater willingness of users to share personal experience is associated with a higher engagement rate. Our results contribute to understanding the effects of social media on health.

Keywords: Online health community · Illness · Stigma · Personal experience

1 Introduction

Wide spreading of Internet technologies and social media use transforms communication practices in many domains including health communication. In the US, 87% of adults use the Internet, 72% of Internet users say they searched for information about

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 47–58, 2019.
https://doi.org/10.1007/978-3-030-34770-3_4

health on the Internet over the past year, 26% of Internet users read about someone else's experience in self-management of health, and 18% of users go online to find other people who have similar health problems [25]. At the beginning of 2018, the weekly Internet audience in Russia was 70% among the adult population, and the daily Internet audience was 64% [12]. The popular Russian search engine and "health" was one of the biggest search topics [32]. According to Google research, one in 20 Google searches are for health-related information [16] and 44% of Russians if experience symptoms of illness and pain, first of all, find the solution on the Internet [15].

For a long time for the majority of people, the illness experience was a private experience, shared only with relatives, doctors, and sometimes with some close friends. The Internet transforms the illness experience from a private to public experience for many people [8]. In the 1980s, only five people out of 80 interviewed people with epilepsy had ever in their life talked to people who had the same illness [26]. A 2011 publication reports that every fourth Internet user living with a chronic condition is looking for other people with a similar health condition [13]. Experience of illness increasingly becomes public, because people with certain conditions have the opportunity to unite with others experiencing similar problems, for discussion, mutual assistance, protection the rights of patients, etc.

Descriptions of personal illness experience in social media are important because they allow people with certain illnesses to obtain the information that doctors usually do not provide. This information concerns ways of coping with the disease in everyday life and creating the subjective meaning of the condition. Participation in online health communities is usually complementary to professional medical help and meets the needs that cannot be satisfied by medical professionals. When communicating with health professionals, patients often lack empathy and emotional support and experience the gap between professional health information and everyday life. Doctors are perceived by patients as experts who do not have personal experience of living with the disease [28]. At the same time, patients need to build their own daily life with disease, and to understand how other people manage in similar situations [14]. Sometimes people access online health communities because they believe that doctor does not inform them about all possible treatment options or think that doctor is not familiar with the latest achievements in medicine [7].

The most common effect of patients using social media for health is patient empowerment, i.e. enhanced subjective and psychological well-being, improved self-management and control. Peer-to-peer online health communication has such important functions as social support, consisting of emotional, esteem, informational, and network support, emotional expression and social comparison [28]. The exchange of personal experiences is the key feature of online health communities. The willingness to share personal health information and personal stories is very important for these communities to perform their functions. Researchers identified seven domains through which online patients' experiences could affect health: finding information, feeling supported, maintaining relationships with others, affecting behavior, experiencing health services, learning to tell the story and visualizing disease [34]. They also concluded that "the act of participating in the creation of health information (e.g., through blogging and contributing to social networking on health topics) also influences patients' experiences and has implications for our understanding of their role in their own health care management and information" [34].

At the same time, online health communities are characterized by information imbalance in supply and demand where "there is much demand for high quality, personal experiences, but minimal supply of personal experiences is provided" in online health communities [20]. Most users came to online health communities to read what others post, comment or discuss and only a small number of users actively and regularly participate in discussions and share their personal experience.

Although the willingness to disclose personal health information is determined by a variety of individual and non-individual factors [2], at the same time, we can assume that to a certain extent willingness to share personal experience depends on the perception of illness in society.

If the willingness to share personal experience depends on the social characteristics of particular illness, then we will probably see differences in the shares of personal narratives in online communities of different illnesses within the same Internet resource. We assume that the stigmatization of illness is an important factor influencing willingness to share personal narratives and affecting willingness to disclose the personal health information in online health communities. The goal of our research is to understand the relationships between the willingness of people with different illnesses to share personal experiences in online communities and the extent of illness stigmatization.

2 Background

Existing research of online health communities focuses either on benefits, and limitations of social media use for health communication regardless of the disease [1] or on specific communities devoted to particular diseases. There is a lot of research devoted to different aspects of communication in online communities for people living with different chronic conditions, for example, cancer [3, 4], diabetes [17, 18, 33], HIV [11, 22], tuberculosis [19, 31] etc. The other type of research focuses on the common features of online health communication [27].

At the same time, there is a lack of research comparing communities devoted to different illnesses. There are only a few examples of such research. The analysis of different types of online health communities' users demonstrated that depending on the illness, certain types of users (caretakers, opportunists, scientists, and adventurers) were more popular than others [20]. The comparative study of online discussion forums for breast cancer, diabetes and fibromyalgia revealed emotional and temporal aspects of the illness experience and differences in discussion content between health conditions [6]. Research of differences among breast and prostate cancer online support groups revealed that the type of support varied by cancer type. Support messages on the breast cancer sites mostly offer emotional support, and messages of support on the prostate cancer sites focus on informational support [5]. The main features of communication in online health communities devoted to specific diseases differ and should be studied in order to better understand the needs of their participants in informational and emotional support.

Considering problems of personal health information disclosure, researchers usually do not pay attention to differences between illnesses. In accordance with communication privacy management theory, [23, 24] the willingness of people to share personal information with others depends on the perceived benefits and costs of information disclosure. One of the costs of disclosure is the perceived risk to be stigmatized or rejected. Research shows that the level of stigmatization is different for different illnesses. Crandall & Moriarty revealed dimensions of stigmas which are critical for social rejection: the severity of the illness, behaviorally causality, avoidability, contagiousness and sexual transmission. They concluded that diseases perceived to be severe or under personal control are most likely to lead to stigmatization and social rejection [9].

In order to understand how online communities of different illnesses differ in terms of the willingness of users to share their personal experiences, we decided to select illnesses comparable by their perceived severity but differ by the level of stigmatization. For our analysis, we selected 5 diseases of public health concern in modern Russia [10] which are characterized by high levels of primary disability and mortality and reduced life expectancy: cancer, diabetes, hepatitis B and C, HIV, tuberculosis. All these illnesses are considered as quite severe, but different by the level of stigma that can make sufferers hide their illness from others in many contexts. Comparing these illnesses in accordance with the dimensions of stigmas proposed by Crandall and Moriarty [9], we see that HIV, tuberculosis, hepatitis B and C are highly stigmatized diseases unlike cancer and diabetes which are less stigmatized.

Table 1. Dimensions of stigma for selected illnesses

	Contagiousness	Behaviorally caused	Avoidability	Sexual transmission
Cancer				
Diabetes				
Hepatitis B, C	+	+	+	+
HIV	+	+	+	+
Tuberculosis	+	+	+	

We proposed several hypotheses.

Hypothesis 1 (H1). Stigma pushes sufferers to conceal their illness, and, therefore, prevent them from sharing their personal health information with others. The higher the degree of stigmatization of the illness, the smaller the share of personal stories in an online community devoted to the illness.

Hypothesis 2 (H2). The anonymity of internet communication eliminates the fear of stigma. The higher the stigma of the illness, the greater the proportion of deactivated or anonymous profiles among users sharing personal experiences in an online community devoted to the illness.

Hypothesis 3 (H3). Personal stories encourage users to actively comment and to share their personal experiences; therefore, a greater amount of personal stories makes communities more active and contributes to a higher level of engagement.

The higher the number of personal stories in the content of the online community, the higher the engagement rate (ER).

3 Method and Results

In order to compare online communities devoted to different illnesses in terms of users' willingness to share personal experiences, we should minimize the influence of the internet platform configuration, so we have to analyze communities situated on the same internet platform. For our purposes, we chose the social networking site Vkontakte (vk.com). It is the most visited social networking site in Russia [21], which contains more than 500 million registered accounts [30]. Moreover, this site provides open access to public pages through the API (https://vk.com/dev/methods). User's publications in Vkontake groups have three forms: posts, comments to posts and messages on boards. Boards are similar to online forums since they are composed by threads, addressed to different topics. The sampling of online communities was carried out by the API method "group.search" applying key words, which were contained in the titles of groups, public pages or events [29]. This method revealed 7331 groups. This dataset was cleaned manually; we removed both groups irrelevant to our purpose as well as closed groups. Then we excluded groups where the number of members was less than 1000 accounts and which did not have any activity during 30 days before the date of data collecting. The remaining groups were checked for spam and activity; we excluded groups, which did not demonstrate any interactions among users. The final sample included 34 groups (Table 2).

Table 2. Sample of groups.

Illness	Key words	Number of groups selected by keyword	Number of manually selected groups	Number of groups in the final sample
Cancer	Oncology, cancer, chemotherapy	5054	76	8
Diabetes	Diabetes, insulin	644	237	11
Hepatitis B, C	Hepatitis	240	71	5
HIV	Antiretroviral, ARVT, HIV, AIDS	1329	126	8
Tuberculosis	Tuberculosis	64	58	2
Total		7331	549	34

To estimate the share of personal stories for each illness, three samples were generated. Using simple random sampling without replacement, we selected 100 posts, comments and messages on boards from each group. If a group contained less than 100 items of any category, the data were included completely. The number of items in each sample is shown in Table 3.

The text was coded as containing a personal story if users told about their own health problems or about the health problems of their relatives. Keywords indicated the presence of a personal story in the text were personal pronouns (I, we, she, he) and references to relatives (child, wife, husband, mother, father, and so on). The text was coded as non-personal if it included impersonal information such as advertising, the announcements of the sale and purchase of medicines, requests to find a cure, information about new methods of treatment, talks off topic.

Table 3. Size of samples

	Posts	Comments	Boards	Total
Cancer	300	614	542	1456
Diabetes	392	1100	1100	2592
Hepatitis B, C	200	219	133	552
HIV	500	700	700	1900
Tuberculosis	200	200	200	600

The texts were classified by two coders, both independently coded 200 identical posts (Cohen's kappa = 0.93), comments (Cohen's kappa = 0.94) and messages on boards (Cohen's kappa = 0.94). In 4 of 34 groups (3 hepatitis groups and 1 oncology group) we did not find personal stories at all.

To obtain aggregated results for each illness, the shares of personal stories in posts, comments and boards were weighted according to the number of members in groups. To get summarized share of personal stories including all three types of communication, the percentages in each type were weighted according to its share in the total number of posts, comments and messages on boards. The resulted distributions are represented in Table 4.

Table 4. Distributions of personal stories in selected groups, %

	Posts	Comments	Boards	Weighted sum
Cancer	19,0	5,9	10,6	9,0
Diabetes	6,1	16,4	11,2	13,3
Hepatitis B, C	18,2	2,3	1,2	7,1
HIV	0,3	0,3	9,1	7,6
Tuberculosis	57,0	17,1	34,3	21,0

As shown in Table 4, hypothesis 1 was not confirmed. Indeed, the share of personal stories in the communities of the most stigmatized illnesses, hepatitis and HIV, is smaller, than in oncology and diabetic communities. This result is consistent with the assumption regarding the relationships between willingness to share personal stories and the extent of illness stigmatization. Simultaneously, tuberculosis, which occupies a middle position on the scale of stigmatization in accordance with the Crandall and Moriarty [9] dimensions of stigmas (see Table 1), has the largest share of personal stories what refutes the hypothesis 1. We can suppose that there is no direct relationship between illness stigmatization and users' willingness to share personal stories. Perhaps, there are other factors, which, even in the circumstances of illness stigmatization, encourage participants to exchange personal information. Clarification of these issues needs additional research.

On the social networking site Vkontakte users can publish their texts anonymously, using 4 ways: to make a publication on behalf of a group, to close the access to a profile, to create a fake profile or to provide very scant personal information that makes it impossible to identify a real person. It is rather complicated to obtain a reliable estimation of the number of fake accounts because we did not have an algorithm for this analysis and manual analysis does not always allow identifying fake accounts. Meanwhile, for comparative analysis we used a proxy indicator of the number of fake accounts. Vkontakte provides some information about deleted or banned accounts (the site marks them as "deactivated"), and this information can point out that accounts were temporally created. Temporal accounts could be created in order to disclose some details of personal condition, and then these accounts are usually deleted or abandoned.

To estimate the extent of profile openness as a possibility to identify a profile with a real person, for each account we calculated the share of filled out positions in six standard profile positions, such as first name, last name, gender, date of birth, country and city. This estimation is quite rough but sufficient for our purposes. Data were weighted by the number of members in each group.

The results of this analysis are presented in Table 5. We did not reveal any closed profiles in our sample. The share of messages on behalf of a group is no more than 1%. More than 20% of profiles are deactivated in the HIV groups and more than 10% in the tuberculosis groups. The average level of profile openness is approximately the same for all illness.

Table 5. The anonymity of "storytellers"

	Share of messages on behalf of a group	Share of deactivated profiles	Average openness of the profile
Cancer	0,4	2,9	5,11
Diabetes	0,6	8,5	5,20
Hepatitis B, C	0,0	5,1	5,12
HIV	0,0	21,7	5,28
Tuberculosis	0,2	13,9	4,96

Hypothesis 2 was partially confirmed: the most anonymity was revealed in the communities of highly stigmatized illnesses. Every fifth profile was deactivated in HIV groups and every seventh – in tuberculosis groups. Slightly bigger amount of closed profiles in diabetes groups than in hepatitis groups needs additional research for reasonable interpretation.

To examine if there is a relationship between the level of stigmatization and the extent of storyteller's anonymity we divided our communities into 3 categories: the first category contained the groups of the most stigmatized illnesses (HIV and hepatitis), the second category included illness with the middle level of stigmatization (tuberculosis) and third category contained less stigmatized illnesses (diabetes and oncology). We used Kruskal-Wallis test to compare obtained categories in relation to the differences in the shares of deactivated accounts and the average value of openness. The test revealed statistically significant differences between the first and the third categories regarding deactivated accounts. The second category did not show any significant relationships; probably because it included only two groups.

To assess the level of engagement rates (ER) we employed two indicators: the ER for posts, and the ER for boards. ER for posts was calculated as the ratio of the sum of all comments to posts and all likes to posts to the total number of views. Since on Vkontakte likes are added automatically to all repost we did not count reposts. ER for boards was calculated as the ratio of all messages on the boards to the number of members in a group. Both indicators were multiplied by 1000 for convenience of calculations and analysis.

To test the hypothesis 3 we measured using Spearmen correlation three dependencies: ER for posts and the share of personal stores in posts, ER for posts and the share of personal stores in comments, ER for boards and the share of personal stores on boards. The results are shown in Table 6.

Table 6. Dependencies between engagement rates and personal stories

	RHO	P-value	Number of cases[a]
ER_posts and personal_posts	0,66	0,0001	30
ER_posts and personal_comments	0,24	0, 2137	30
ER_boards and personal_boards	0,29	0, 4825	29

[a]The number of cases in this analysis is less than the number of groups in our sample because several groups hid the number of views or did not use boards.

The results did not confirm the dependency between engagement rate and the share of personal stories in personal comments or in personal boards. The results confirmed the dependency between the share of personal stories and engagement rate, if stories are placed in posts, what can be explained by the more visibility of posts than comments or boards.

4 Discussion

Studying online communities located on the social networking site Vkontakte and devoted to such illnesses as cancer, diabetes, hepatitis B and C, HIV, tuberculosis, we compared the willingness of users to share their personal experiences online. We hypothesized that the higher the stigmatization of the disease then the less the users tended to discuss their personal experiences online. We found that online communities devoted to different illnesses have different shares of personal stories in their content. The initial hypothesis about the relation between stigmatization of the illness and the share of personal stories in online communities devoted to this illness was not confirmed. Communities of cancer and diabetes demonstrated larger share of personal stories in the content than communities of HIV and hepatitis. At the same time the tuberculosis communities demonstrated the largest share of personal stories among all groups of illnesses. Giving that we considered tuberculosis as stigmatized illness, this result is quite surprising and needs further research by qualitative methods to understand how stigmatization influences communication in online health communities. The further research should reveal what are the differences in the topics of the personal stories in groups devoted to different illnesses.

We also hypothesized that the openness of personal profiles of users sharing personal experiences would correlate with the illness stigmatization. In our research the biggest share of deactivated profiles demonstrated communities devoted to HIV and Tuberculosis. Giving that the Kruskal-Wallis test revealed significant differences between HIV/hepatitis and diabetes/oncology groups, we can conclude that the stigmatization of illness is an important factor making users hide personal information in online health communities. The shares of messages on behalf of a group were relatively small. Perhaps these indicators were determined rather by the configuration of the internet platform than some factors related to the illness context. The average openness of profiles was relatively the same for all compared groups. It is worth noting that the measurement of profile openness did not take into account fake names, fake photos and other fake profile positions. Probably this was a reason that we could not reveal significant differences between compared groups.

The hypothesis that the greater willingness of users to share personal experience is associated with a higher ER has been confirmed. Thus, we can conclude that the share of personal stories in the online community content is the factor encouraging people with illnesses to access online peer health communities.

Our research has some limitations that worth noting. Given the small sample size, we cannot make any generalizable conclusions concerning analyzed illnesses. The results could be different both for other cultural contexts and for other internet platforms. Moreover, we did not have access to private or closed groups what make our results relevant only for public online health communities.

Although stigmatization pushes people to conceal their status and to refrain from discussing their conditions openly in offline communication, we can assume that patterns of online communication for stigmatized illnesses could be different. Perhaps there are factors that mitigate the effect of stigma and encourage people to share personal stories about their conditions in online communities. These factors may be related, for example, to specifics of the illness itself or its social context, to the policies of online community administration, to the lack of information about recent achievements in treatment etc.

Despite the fact that some of our assumptions about the influence of stigmatization on the willingness of users to share personal stories were not confirmed, we have revealed significant differences between different illness communities, both in the number of personal stories and in the shares of deactivated profiles. This suggests that the communicative structure of online communities devoted to various illnesses differ and partly determined by both the illness specificity and social context of illness. We were able to show that communication around a particular illness has its specifics. Further comparative research is needed to better understand the existing patterns of online health communication.

5 Conclusion

Peer-to-peer online health communication expands the opportunities for patient self-management, contributes to obtaining information and emotional support and promotes patient empowerment. Personal illness narratives provide a foundation for creating new meanings of certain illnesses and diversification of official medical discourse. At the same time, peer-to-peer consulting in online health communities entails certain risks, since the degree of trust in information received from other users may not correspond to the actual quality of this information. It should be borne in mind that the more actively users share their personal stories and examples, the lower the risk to overestimate the information containing in any single story. The fewer users hide their profiles, the more likely they take responsibility for the content of the posted information. In this context understanding the effects of social media for health communication in different social contexts and for various illnesses is very important and needs further research.

Studying the willingness of internet users to publicly share their personal illness experiences in online social networks with people experiencing similar conditions is important for understanding the changes in health communication that occur under the influence of the spread of information technology. Such studies can also contribute to the understanding in the comparative aspect of how particular illnesses differ in terms of their social perception and social stigmatization and how social media create new audiences – communities of people united around discussion of their personal experiences of living with illness. A comparative study of online health communication reveals both common patterns in online discussions and finds differences due to the specifics of a particular disease. A comparative analysis allows researchers to better understand what people with different illnesses expect from online health communities and tailor these communities to the needs of users by promoting more effective models of online health communication. Studying online personal narratives helps to better understand the social portrait of a particular illness, its perception in society and some of the potential opportunities and risks that can be associated with participation in online communities dedicated to certain illnesses.

Acknowledgements. This work was supported by the Russian Fund for Basic Research [grant number 18-013-00726 A]. Research was conducted at the Research park of St. Petersburg State University "Center for Sociological and Internet Research".

References

1. Andersen, K.N., Medaglia, R., Henriksen, H.Z.: Social media in public health care: impact domain propositions. Gov. Inf. Q. **29**(4), 462–469 (2012)
2. Anderson, C.L., Agarwal, R.: The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information. Inf. Syst. Res. **22**(3), 469–490 (2011)
3. Attai, D.J., Cowher, M.S., Al-Hamadani, M., Schoger, J.M., Staley, A.C., Landercasper, J.: Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J. Med. Internet Res. **17**(7) (2015). https://doi.org/10.2196/jmir.4721
4. Bender, J.L., Jimenez-Marroquin, M.C., Jadad, A.R.: Seeking support on Facebook: a content analysis of breast cancer groups. J. Med. Internet Res. **13**(1) (2011). https://doi.org/10.2196/jmir.1560
5. Blank, T.O., Schmidt, S.D., Vangsness, S.A., Monteiro, A.K., Santagata, P.V.: Differences among breast and prostate cancer online support groups. Comput. Hum. Behav. **26**(6), 1400–1404 (2010). https://doi.org/10.1016/j.chb.2010.04.016
6. Chen, A.T.: Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ. Couns. **87**, 250–257 (2012)
7. Colineau, N., Paris, C.: Talking about your health to strangers: understanding the use of online social networks by patients. New Rev. Hypermedia Multimed. **16**(1–2), 141–160 (2010)
8. Conrad, P., Bandini, J., Vasquez, A.: Illness and the internet: from private to public experience. Health **20**(1), 22–32 (2016). https://doi.org/10.1177/1363459315611941
9. Crandall, C.S., Moriarty, D.: Physical illness stigma and social rejection. Br. J. Soc. Psychol. **34**(1), 67–83 (1995)
10. Decree of the Government of the Russian Federation dated 01.12.2004 No715 (ed. 13.07.2012) On Approval of the List of Socially Significant Diseases and the List of Diseases that are Dangerous to others. http://base.garant.ru/12137881/. Accessed 22 Apr 2019
11. Flickinger, T.E.: Social support in a virtual community: analysis of a clinic-affiliated online support group for persons living with HIV/AIDS. AIDS Behav. **21**(11), 3087–3099 (2017)
12. Foundation of Public Opinion (FOM): Internet in Russia: Penetration Dynamics. Winter 2017–2018. https://fom.ru/SMI-i-internet/13999. Accessed 22 Apr 2019
13. Fox, S.: The Social Life of Health Information. Pew Research Center: Internet, Science & Tech. http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/. Accessed 22 Apr 2019
14. Gómez-Zúñiga, B., Fernandez-Luque, L., Pousada, M., Hernández-Encuentra, E., Armayones, M.: ePatients on YouTube: analysis of four experiences from the patients' perspective. Medicine 2.0 **1**(1) (2012). https://doi.org/10.2196/med2.2039
15. Google. What are Russians Looking for in the Internet about Health and Medicine. https://www.thinkwithgoogle.com/intl/ru-ru/industry/pharma/internet-and-medicine/. Accessed 22 Apr 2019

16. GoogleBlog: A Remedy for your Health-Related Questions: Health info in the Knowledge Graph. https://googleblog.blogspot.com/2015/02/health-info-knowledge-graph.html. Accessed 22 Apr 2019
17. Greene, J.A., Choudhry, N.K., Kilabuk, E., Shrank, W.H.: Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J. Gen. Intern. Med. **26**(3), 287–292 (2011)
18. Ho, Y.X., O'Connor, B.H., Mulvaney, S.A.: Features of online health communities for adolescents with type 1 diabetes. West. J. Nurs. Res. **36**(9), 1183–1198 (2014). https://doi.org/10.1177/0193945913520414
19. Horter, S., Stringer, B., Venis, S., du Cros, P.: I can also serve as an inspiration: a qualitative study of the TB&Me blogging experience and its role in MDR-TB treatment. PloS one **9**(9) (2014). https://doi.org/10.1371/journal.pone.0108591
20. Huh, J., et al.: Personas in online health communities. J. Biomed. Inform. **63**, 212–225 (2016)
21. LiveInternet, Rating, https://www.liveinternet.ru/rating/ru/. Accessed 22 Apr 2019
22. Mo, P.K., Coulson, N.S.: Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol. Behav. **11**(3), 371–374 (2008)
23. Petronio, S.: Boundaries of Privacy: Dialectics of Disclosure. SUNY Press, Albany (2002)
24. Petronio, S.: Communication boundary management: a theoretical model of managing disclosure of private information between married couples. Commun. Theory **1**(4), 311–335 (1991)
25. Pew Research Center: Internet, Science & Tech. Health Fact Sheet. http://www.pewinternet.org/fact-sheets/health-fact-sheet/. Accessed 22 Apr 2019
26. Schneider, J.W., Conrad, P.: Having Epilepsy: The Experience and Control of Illness. Temple University Press, Philadelphia (1983)
27. Smailhodzic, E., Boonstra, A., Langley, D.: Social media enabled interactions in healthcare: towards a typology. In: 23rd European Conference on Information Systems (ECIS), Münster, Germany (2015). https://doi.org/10.18151/7217481
28. Smailhodzic, E., Hooijsma, W., Boonstra, A., et al.: Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv. Res. **16** (2016). https://doi.org/10.1186/s12913-016-1691-0
29. Vkontakte: API Methods. https://vk.com/dev/groups.search. Accessed 22 Apr 2019
30. Vkontakte: Catalog, https://vk.com/catalog.php. Accessed 22 Apr 2019
31. Wai, P.P., et al.: Community-based MDR-TB care project improves treatment initiation in patients diagnosed with MDR-TB in Myanmar. PLoS ONE **13**(3) (2018). https://doi.org/10.1371/journal.pone.0194087
32. Yandex Research: Health Search. https://yandex.ru/company/researches/2016/ya_health. Accessed 22 Apr 2019
33. Zhang, Y., He, D., Sang, Y.: Facebook as a platform for health information and communication: a case study of a diabetes group. J. Med. Syst. **37**(13), 1–12 (2013)
34. Ziebland, S., Wyke. S.: Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q. **90**(2), 219–249 (2012)

An Innovative Online Tool to Self-evaluate and Compare Participatory Research Projects Labelled as Science Shops or Citizen Science

Anne-Sophie Gresle[1(✉)], Anna Cigarini[2,3], Leonardo de la Torre Avila[1], Irene Jimeno[1], Franco Bagnoli[4], Herman Dempere[5], Mireia Ribera[5], Eloi Puertas[5], Josep Perelló[2,3], María Jesús Pinazo[1], and on behalf of the InSPIRES Consortium

[1] ISGlobal, Hospital Clínic, Universitat de Barcelona, Barcelona, Spain
{Anne-sophie.gresle, leonardo.delatorrea, Irene.jimeno, mariajesus.pinazo}@isglobal.org

[2] Universitat de Barcelona Institute of Complex Systems, Barcelona, Spain
cigarini.anna@gmail.com, josep.perello@ub.edu

[3] Departament de Física de la Matèria Condensada, Universitat de Barcelona, Barcelona, Spain

[4] University of Florence, Florence, Italy

[5] Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain
herman.dempere@gmail.com, mireia.ribera@gmail.com, epuertas@ub.edu

Abstract. Participatory research keeps expanding to connect science and society through engaging projects using a multi-stakeholder strategy, including citizens. However, each participatory project follows different evaluation formats and strategies. This results in limiting evidences on best practices, hindering the scaling up of Participatory Research. Through the H2020-funded InSPIRES project, an innovative and online-based evaluation strategy was developed which is valid for Participatory Research initiatives labelled as Science Shops or Citizen Science. This strategy challenges those teams that want to undergo a self-reflection process during and after their project is active. An online-tool gathers and automatically analyses data in a harmonized way among projects. The tool delivers back a set of pieces of information through different visualizations which analyze each project's process in five dimensions, selected-constructed after a careful revision of public engagement and impact evaluation criteria proposed by different projects and researchers. The dimensions evaluated by this online instrument are: (i) Knowledge Democracy, (ii) Citizen-led Research, (iii) Participatory Dynamics, (iv) Integrity, and (v) Transformative Change. Online-based self-evaluation questionnaires were designed and personalized according to the profile of the respondents and are sent out by email in four different stages to capture the momentum of the project, as well as its short-term and mid-term impacts. The quantitative and qualitative evaluation instrument is featured within the InSPIRES Open Platform (OP) which becomes an open repository that allows comparison among participatory projects.

InSPIRES Consortium—Members listed at the end of the paper.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 59–72, 2019.
https://doi.org/10.1007/978-3-030-34770-3_5

Keywords: Impact evaluation · Digitalized evaluation · Automatic data analysis · Participatory research · Citizen Science · Science Shops

1 Introduction

Science Shops are defined as knowledge intermediary structures that jointly with civil society organizations (CSOs) co-create research questions, deploy participatory projects to respond to societal needs, include students in the work and support the translation of research results [1]. Besides, the Citizen Science term is being increasingly used to refer to a growing number of Participatory Research activities. While including a number of different practices, it broadly refers to a partnership between professional researchers and volunteers in which the volunteers are actively involved in scientific tasks that have traditionally been implemented by scientists [2, 3]. Being increasingly adopted by government institutions and international organizations, despite some challenges, Citizen Science is taking advantage of the rise of connectedness and communication technologies to quench the thirst for data and to improve the transparency and accessibility of science [4], and they often use visualization tools for easier digest of data [5]. Both Science Shops and Citizen Science could eventually refocus what parts of the natural and social worlds are subject to scientific inquiry, thereby transforming what we know about the world [6].

While Science Shops have been running in some countries for more than 40 years, the impact evaluation's studies remain limited. The few available evaluation studies reported that Science Shops are a fairly unique way (i) to put science at the service of society to address real life challenges, (ii) to provide all involved stakeholders with new skills, (iii) to offer new research ideas for established researchers, and (iv) to make research processes more transparent, democratic and inclusive by considering participation as main research driver [7, 8]. Albeit having a shorter tradition, very similar vision and approach is shared by Citizen Science practices.

The main techniques and tools that are considered for performing impact evaluation studies are interviews and paper-format based questionnaires, which provide outputs that makes comparison among projects very difficult. Furthermore, besides the fact that the evaluation's conclusions are rarely published and shared, each project follow their own format and strategy. This results in limiting evidence on best practices, hindering the scaling up of the more open perspective that these two classes of Participatory Research is providing [9, 10].

Inspired by the Citizen Science approach where digital tools and internet have been widely developed for engaging non-academic participants, and through the InSPIRES H2020 funded project, an innovative and online-based evaluation strategy was developed which can be used for many Participatory Research initiatives that want to be evaluated or want to undergo a self-reflection process during and after the project is active. The proposed online-tool gathers and automatically analyses data in a harmonized way among projects. The data gathered delivers back a set of pieces of information through different visualizations that are covering several aspects related to most of Participatory Research projects. Online-based self-evaluation questionnaires that were designed and personalized according to the profile of the respondents (i.e. civil society members, researchers, students, project managers), are sent out by email and responded online, in four different stages to capture the momentum of a project, as well as its short-term and midterm impacts. The dimensions evaluated by this online instrument are: (i) Knowledge Democracy (i.e. transdisciplinarity and relevance of topics), (ii) Citizen-led Research (alignment of project goals to the community demands and efficacy of engagement techniques), (iii) Participatory Dynamics (degree and quality of engagement), (iv) Integrity (ethics, transparency of data management, gender, etc.), and (v) Transformative changes (individual learning, personal growth, sustainability, impact on policies, etc.).

This online tool is featured within the InSPIRES Open Platform (OP), accessible via this URL https://app.inspiresproject.com, which is an open repository for Participatory Research projects and their promoting structures, such as at the moment Science Shops or Citizen Science research groups. This is a quite unique practical approach to carry out impact evaluation studies and to share them. The OP is completely free, open and its goal is to act as a hub for comparative evaluation of Science Shops and Citizen Science projects. Individuals contributing to the platform and participating in a particular project keep ownership and privacy of their own data and can download the data of their projects in order to perform more in-depth project-based analysis.

After a revision of the state of the art regarding impact evaluation tools for Participatory Research projects, the methods section describes the OP, the evaluation and communication strategy, and the collective effort implemented to produce such tool which has been followed by a multidisciplinary team composed by a mathematician, a statistician, an UX designer, and researchers from different fields that include experts in Science Shops, Citizen Science and other Participatory Research practices. The results section shows the platform and the conclusion section includes a reflection on the takeaways of the project and future work to be done.

2 State of the Art

More and more funding agencies at national and European levels demand to take into consideration the social impacts of funded projects on their evaluation [11]. The European Commission, for example, has put at the core of its research and innovation policies two concepts that seek for generating and evaluating social impacts: the Responsible Research and Innovation (RRI) since 2004 and the Open Science and Open Innovation since 2016. These two paradigms advocate the promotion of more ethical, open, inclusive, reflexive and participatory science [12, 13]. Participatory Research projects, such as Science Shops and Citizen Science, have the potential to integrate the RRI and Open Science characteristics and contribute to social impacts.

Going beyond the current international evaluation framework has become an important task in order to understand impacts of Participatory Research projects in dimensions that overcome classical bibliometric indicators in academia. In Participatory Research, evaluation can be performed at two different levels, first in regards with co-learning processes generating among all stakeholders during a project and secondly in regards to the quality and use of results being collectively produced. Elements from these two dimensions are not captured in traditional research evaluation methods in academic contexts and/or by public funding agencies.

While many frameworks and guidebooks for evaluation are available, a majority of published research with a stakeholder partner engagement dimension does not include an evaluation component [9]. According to this study, most evaluations are qualitative and use self-report through focus group, one-on-one semi structured interviews, informal observations and/or written surveys with open-ended text responses, and the level of details regarding evaluation design, strategy and results are very limited. In a review of research assessment models and methods, Milat et al. 2015 [14] discussed four different theoretical frameworks. They find that they all differ in terminology and approaches and that the typical methods reported to perform the evaluation are desk analysis, bibliometrics, panel assessments, interviews and case studies, relying mostly on principal investigators interviews and/or peer review and they rarely include other stakeholders such as policy-makers and other important end-users of research creating a clear bias in evaluation studies. They report that multidimensional models are more valuable as they capture broader range of potential benefits including capacity building, policy and product and service development, as well as societal and economic impacts. Their review also suggests that the research sector should use broader indicators to better capture policy and practice outcomes, as bibliometric indices do not say much about the real-world benefits of research.

For Science Shops projects, an evaluation strategy was proposed by a former EU funded project, titled PERARES [8]. The evaluation consisted of a series of quantitative paper-based questionnaires to be distributed to all stakeholders at different times of the project. After having performed interviews to experts participating in the PERARES project, findings from InSPIRES suggest that the evaluation task is not routinely integrated within Science Shops projects because of a lack of time and resources. As for Citizen Science, there are currently no commonly established indicators to evaluate these types of projects [10]. Kieslinger et al. recommend to develop a framework and transform it into a practical assessment tool for projects and initiatives, through a mix of quantitative and qualitative methods such as for example online surveys, statistics, in-depth interviews and focus groups.

The InSPIRES Open Platform aims to bring together civil society, practitioners and other stakeholders from across and beyond Europe to roll out innovative models for Science Shops and Citizen Science which systematically include an impact evaluation study. Science Shops, as mission-oriented intermediary units between the scientific sphere and civil society organisations, facilitates citizen-driven Open Science projects to respond to the needs of civil society organisations and most of the time including students in the work process. Like in Citizen Science, the aim is to co-create and prioritize research questions together with civil society members as well as develop projects in a most participatory manner. As many other EU funded Participatory Research project, InSPIRES had as a core objective to develop an improved impact evaluation strategy and systematically introduce it within Participatory Research projects to capture process and results outcomes. After months, of discussions among partners and external advisors, it was decided to develop an online based impact evaluation tool, using quantitative and qualitative indicators to capture the produced impacts. This approach seemed to be the right equilibrium between a broad evaluation framework and a viable application. Constructing an evaluation tool requires obviously a simplification of the object of study [15], and the limitation of time and resources constrains the scope of the evaluation. But still, it is the first of its kind to propose a harmonized evaluation data collection and analysis, allowing for comparison among participatory research projects.

The InSPIRES Open Platform (OP) was developed and is on the one hand an open repository for Participatory Research projects and promoting structures, such as Science Shops and Citizen Science research groups, and on the other hand, a quantitative and qualitative evaluation instrument capturing data in five different dimensions. The uniqueness of the approach is that the impact of the evaluation tool is both an online questionnaire and a platform for automatic data analysis visualization. Thereby, offering real-time project evaluation reports comparing performance of one project in relation to the other ones registered on the platform. To our knowledge, this is the first time that a platform featuring such characteristics is developed and proposed to the Participatory Research community.

3 Methods

3.1 The Choice of Dimensions, Indicators and Questions

While providing a set of indicators that allows for within and between project monitoring and impact assessment, the online-based evaluation tool is primarily aimed at enhancing awareness and self-reflection for project members by giving them the opportunity to reflect upon the purpose and design of the research, both during and by the end of the project. In line with the principles of Action Research [16] and experiential learning [17], the online-based evaluation tool was therefore designed in the hope that the results feed back to the actors involved would change practice through critical reflection and facilitate the direct application of research findings in a practical context.

Several available tools have set the basis for the definition and selection of relevant indicators. These include the measures adopted in the assessment of public engagement in research as for the PERARES EU Project [8], the metrics developed within the MoRRi EU Project for RRI monitoring [18], and the scientific, individual and socio-ecological criteria as proposed by Kieslinger and colleagues [10] for the evaluation of Citizen Science projects. After a careful revision of criteria proposed by these EU-funded projects and in tune with the criteria proposed by Kieslinger and colleagues [10] for the evaluation of Citizen Science projects, the final classificatory scheme of the online-based evaluation tool covers the following five core dimensions: (i) Knowledge Democracy, (ii) Citizen-led Research, (iii) Participatory Dynamics, (iv) Integrity, and (v) Transformative Change. Each core dimension encompasses different indicators (see Table 1 for further details), and includes from 12 to 23 items each, positively worded on a 0 to 7 scale, plus an open-ended question to capture qualitative feedback.

A defining attribute of Participatory Research projects is that they should value and weight equally the contribution of the participants involved [19]. Evaluation strategies of Participatory Research projects should thus, by their very nature, assess the project value and impact for the different actors involved through trust-related mechanisms and a continuing commitment to power-sharing [20]. In an attempt to include and accommodate multiple viewpoints and needs into the design and conduct of the evaluation process, the online-based evaluation items were thus designed and personalized according to project members profiles which include: project manager(s), professional scientist(s), student(s), and involved members of civil society.

Further, just as Participatory Research projects should be improved over time by the growing experience and reflection from involved project members, the same adaptive capacity and openness is requested from project evaluation [20]. Ideally, indeed, Participatory Research projects should be continuously reviewed and amended by feedback from the community [21]. This to say that evaluation should be all-inclusive but not tight: rather, in the course of a participatory project the evaluation should allow for reflecting developments and contextual changes in the projects [22]. Therefore, with the purpose of reconciling varied perspectives through interactive processes, the online-based evaluation tool is structured along four project stages to capture the momentum of the project, as well as its short-term and mid-term impacts: (a) early stage evaluation; (b) mid-point evaluation; (c) end-point evaluation; and (d) post-project evaluation, to be carried out approximately six months after the project is completed.

Yet, the most important reason for developing a digital version of the evaluation tool is that it supplies open, shareable and, primarily, comparable data. Being the idea of sharing knowledge one of the most powerful aspects in collaborative and participatory practices such as Science Shops and Citizen Science, the platform is thus designed accordingly. The data of each project is visualized allowing comparisons with

Table 1. Core dimensions and relative indicators of the online-based evaluation tool.

Dimension	Indicators	Number of items
Knowledge Democracy	● Scientific relevance ● Openness ● Transdisciplinarity	12
Citizen-led Research	● Community alignment ● Responsiveness to community alignment	20
Participatory Dynamics	● Motivation ● Degree of engagement ● Satisfaction with the participatory dynamics	23
Integrity	● Expectation alignment ● Gender perspective ● Inclusivity ● Transparency ● Reflexivity ● Resource availability	20
Transformative Change	● Knowledge and skills ● Self-improvement ● Collective capacity ● Policy impact	19

those projects that have already gone through their own evaluation process and uploaded their data. While keeping privacy and anonymization through data aggregation, the visualization wants to trigger self-reflection using other projects as a reference point (Table 2).

Table 2. Number of items per project member profile across the core dimensions.

Dimension	Project manager	Scientist	Student	Civil society
Knowledge Democracy	4	4	2	2
Citizen-led Research	4	4	4	8
Participatory Dynamics	5	6	5	7
Integrity	12	3	2	3
Transformative Change	4	3	5	7

3.2 Agile and Sprint

AGILE methodologies were adopted by the entire team during the OP development, in order to guarantee the alignment of all the stakeholders working jointly on the project, and to ensure that a minimum viable product would be shipped in the very time-constrained deadline of 3 months. AGILE and other adjoint techniques are commonly used in software development [23] to manage expectations and focus the attention of participants.

SCRUM [24] was the AGILE methodology used in the OP development. After initial discovery of the requirements of the objectives described, the team planned very regular meetings, at the end of each 2-week interval, being this interval frequency known as a "Sprint". Each meeting consisted of an open debate on the current state of the development and a product demo with a fully working and deployed version of the OP. This demo kept the entire team engaged and promoted participation as the product becomes, since the first meeting, a real and tangible entity that can be commented on. Meetings provided the opportunity to raise concerns, revaluate past decisions and decide on top priority items for the next Sprint.

Going into more detail, the Sprint plans for this project were as follows:

1. Sprints 1 and 2 were focused on building up the database and application infrastructure, the setting up of user registration and login flows, and the design of the necessary data model to support the objectives proposed during the first meeting. Registering Structures and Projects was completed at the end of this period.
2. Sprint 3 was the mid-way OP development milestone to finish up the features started in sprints 1 and 2, and to prepare the system to start with the "Project Evaluation" set of objectives.

3. Sprints 4 and 5 were solely focused on developing the entry forms and visualization technologies for Evaluation. The data model was expanded with the, until that moment, unfinished requirements for the "Project Evaluation" features.
4. The last sprint, completing the 3 months of work, was focused on refining the look and feel of the platform, adding quality-of-life features such as forgotten password resolution and email alerts, and completing legal requirements such as data privacy disclaimers.

On this particular project, not all of the requirements were fully signed off from the start. Instead, as things became final and available, they were brought to discussion to the bi-weekly sprint meetings and incorporated on the rolling development cycle. We believe this would not have been possible if we had chosen to follow a more traditional software development methodology, such as "Waterfall Development" in which requirements are set from the start and can't be changed moving forward.

3.3 The Technological and Design Pipeline

Our platform was deployed as a cloud application, this means that the platform executes in a remote machine (server) and is ready for connections from any kind of client device (i.e. mobile, tablet, laptop,…). Usually, these applications have two important parts to be developed: Frontend (display and interaction with data) and Backend (the provider of data and program logic).

Visualizations were developed on Altair, a Python wrapper of the Vega-Lite Library. The visualizations of evaluation results were organized in three levels depending on information interests and disclosure policy:

(a) Public level (Fig. 1), prioritizing aesthetics and simplicity. This level is based on the project logo and associates each part of the hummingbird with an evaluation dimension. Its wing is associated with Transformative Change, because it is the most active part of the Hummingbird, and the one that creates movement. The head is associated with Knowledge Democracy, following the natural relation of head, mind and knowledge. The heart (or lower body) is associated with Citizen-led Research, because citizenship is the heartbeat of Open Science. The tail is associated with Participatory Dynamics, because it acts as a helm. And, finally, the upper body is associated with Integrity, because without it all other aspects would fall apart. This visualization offers maximum aggregation and simply gives the relative position of each project within the five evaluation criteria, through colours and arrows. It fulfils two needs: to have a quick overview of project excellence and to rapidly compare between projects. This visualization is open to everyone to comply with open data requirements as well as RRI ones such as transparency and openness.

(b) Participants level (Figs. 2 and 3): prioritizing usefulness but preserving privacy. This visualization gives far more detail than the public one and complements the dimension' evaluation with indicators. For each indicator a bullet chart is shown with the evaluation of the project related to the other projects in the platform. Additionally, the project overall position is shown as well as the project evolution through the different data collection phases. This visualization is open to all participants in the project.

(c) Project manager level (Fig. 4): prioritizing the view of the different roles in the project. The structure and presentation are very similar to the participants' visualization but for every dimension the aggregation is done by role. It fulfils the need of the project managers to review the perceptions of the research team and to investigate or compensate deviations. Project managers have also access to public and participant level visualizations, being able to review the project from different perspectives.

The selection of charts followed task and use-case scenarios design: the public one is aiming for the wow effect; the bullet charts are a substitute for gauges and provide "a rich display of data in a small space" [25]; Line charts are the most suitable charts for viewing time evolution; Scatterplots relate two variables, and in the implementation act as a quadrant chart.

The OP's backend was developed on top of Django, a powerful Python framework to act as a controller, using a PostgreSQL database to store the data model. Django allows a rapid implementation based on an application skeleton which can be highly customized, being a perfect platform for agile projects. To deal with the calculations for the evaluation visualizations, the Backend leverages Pandas (a Python library) to execute most of the number crunching. Projects data can be exported in CSV (comma separated values) which allow easy takeout of data to process further in other tools.

4 Results

The platform was successfully deployed to a cloud provider and is currently online. First users are testing the software with about 25 on-going Participatory Research projects, mostly Science Shops and Citizen Science ones, and reporting any issues they find.

The figures below show the three levels of visualization of evaluation results: (Fig. 1) the public visualisation, available to everybody, (Figs. 2 and 3) two graphics from the second level of visualization available only to participants of the project, and (Fig. 4) level accessible only to project managers.

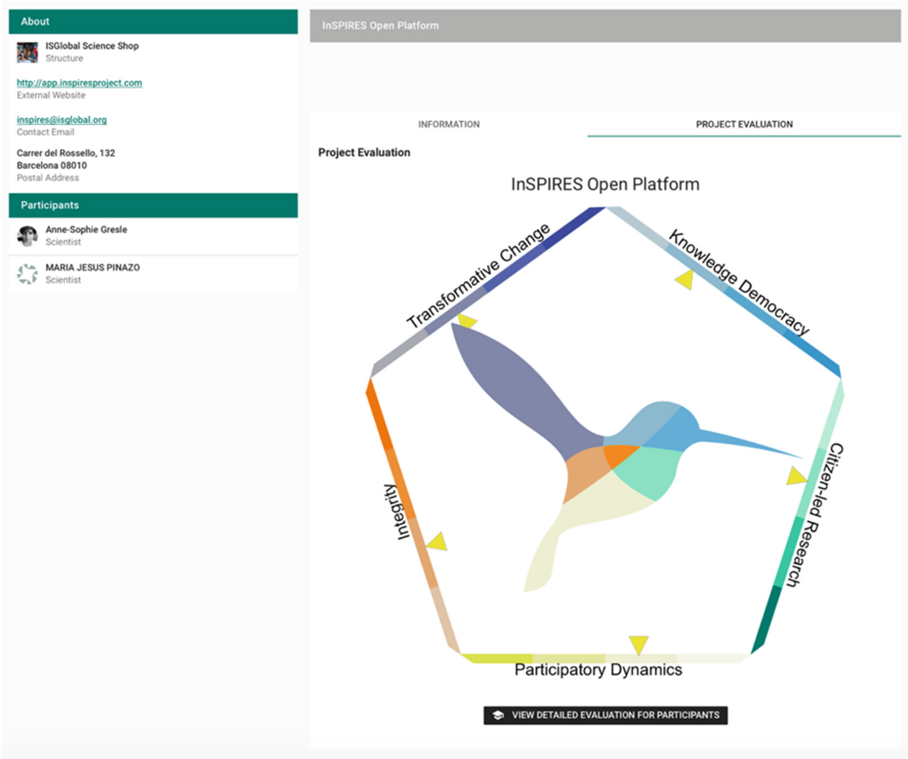


Fig. 1. Public visualization

The *General Evaluation* (Fig. 1) is the public visualization. It displays the five principles and reflects the overall evaluation of the project. Each principle splits the evaluation into four quartiles. In the example above the yellow marker shows the project quartile for each dimension.

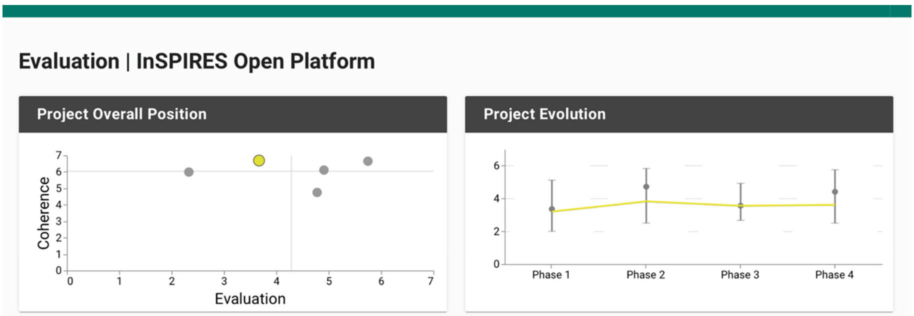


Fig. 2. Project overall visualization for participants and project managers

The *Project Overall Position chart* (Fig. 2) condensates the evaluation scores of all projects into one indicator (evaluation) and calculates, for all projects, the difference of the evaluation scores between the five principles (coherence). The current project is represented by the yellow dot.

The *Project Evolution* chart (Fig. 2) shows, for the current project, the aggregated score for each one of the four phases of the evaluation process represented by the yellow line. The median and quartiles Q1 and Q3 of the evaluation of all projects are displayed for each phase.

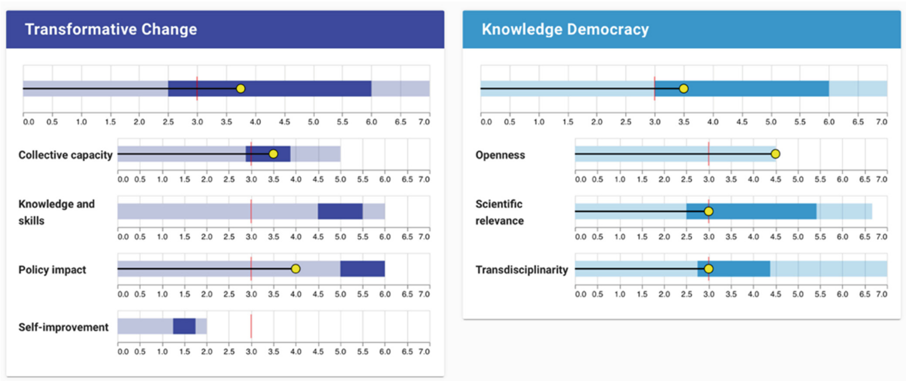


Fig. 3. Principles and subthemes visualization for participants

A meter (bullet chart) (Figs. 3 and 4) is shown for every principle and for the different dimensions that constitute each principle. The yellow dot represents the evaluation of the current project. The shadows represent other projects evaluation: a lighter shadow represents the projects with the lowest and the highest evaluation scores, whereas an opaquer shadow represents the projects within the first and the third quartile of the evaluation scores.

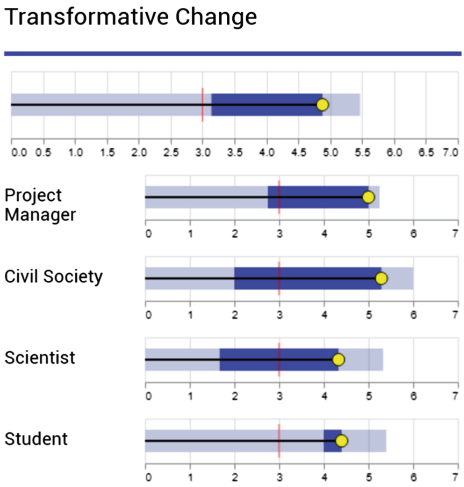


Fig. 4. Roles visualization for project managers

The Project Manager visualization gives a fine-grained detail of the self-reflection process while providing an internal view for management purposes.

5 Conclusions and Future Work

The Inspires Open Platform represents a step forward in Open Science evaluation and in the use of technology and visualization techniques that may impact on future projects on the area. We freely offer it to open science structures and Participatory Research projects to increase their visibility and share good practices.

The platform is currently being tested by 25 projects which are introducing actual data and are going through all the phases of the evaluation processes, this will verify the correct user experience of the participants with all the interaction and also the adequacy of the questions to real and specific projects.

As a future work the authors will do more research on the users understanding of the visualizations and on the data gathered by the platform to make some complex cross-comparison and analysis among the projects in the Inspires Open Platform.

InSPIRES Consortium Members

Barcelona Institute for Global Health: David Rojas Rueda, Valeria Santoro. Environmental Social Science Research Group: Bálint Balázs, Janka Horvath. Fundació Privada Institute de Reserva de la Sida-Caixa: Rosina Malagrida, Marina Pino. Université de Lyon: Shailaja Baichoo, Florence Belaen. VU Institute for Research on Innovation and Communication in the Health and Life Sciences: Marjolein Zweekhorst, Eduardo Muniz Pereira Urias. Universita deglo Studi di Firenze: Giovanna Pacini. Institute Pasteur de Tunis: Hichem Ben Hassine, Sonia Maatoug. Ciencia y Estudios Aplicados para el Desarrollo en Salud y Medio Ambiente: Daniel Lozano, Claire Billot, Faustino Torrico.

Acknowledgements. We would like to thank all the members of the InSPIRES Consortium, and especially the VU and IRSICaixa Team, and all the external partners that have participated in the several iterations, whose suggestions helped in the creation and development of the tool. The InSPIRES consortium also acknowledges the support of the European Union grant number 741677.

Declaration of Interests. We declare no competing interests.

References

1. Fisher, C., Leydesdorff, L., Schophaus, M.: Science shops in Europe: the public as stakeholder. Sci. Public Policy **31**, 199–211 (2004). https://doi.org/10.1038/35108157
2. Heigl, F., Kieslinger, B., Paul, K.T., Uhlik, J., Dörler, D.: Opinion: toward an international definition of citizen science. Proc. Natl. Acad. Sci. **116**(17), 8089–8092 (2019). https://doi.org/10.1073/pnas.1903393116
3. Auerbach, J., et al.: The problem with delineating narrow criteria for citizen science. Proc. Natl. Acad. Sci. **116**(31), 15336–15337 (2019). https://doi.org/10.1073/pnas.1909278116
4. Irwin, A.: No PhDs needed: how citizen science is transforming research. Nature **562**, 480–482 (2018)
5. Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., Crowston, K.: The future of citizen science: emerging technologies and shifting paradigms. Front. Ecol. Environ. **10**(6), 298–304 (2012). https://doi.org/10.1890/110294
6. Strasser, B.J., Baudry, J., Mahr, D., Sanchez, G., Tancoigne, E.: "Citizen Science"? Rethinking science and public participation. Sci. Technol. Stud. **32**, 52–76 (2019). https://doi.org/10.23987/sts.60425
7. Zaal, R., Leydesdorff, L.: Amsterdam science shop and its influence on university research: the effects of ten year of dealing with non-academic questions. Sci. Public Policy **14**(6), 310–316 (1946). https://doi.org/10.1093/spp/14.6.310
8. PERARES Final report D9.2. Evaluating Projects of Public Engagement with Research and Research Engagement with Society (2014). https://www.livingknowledge.org/fileadmin/Dateien-Living-Knowledge/Library/Project_reports/PERARES_Evaluating_Projects_of_PER_Final_report__WP9_Monitoring_and_Evaluation_2014.pdf. Accessed 23 July 2019
9. Esmail, L., Moore, E., Rein, A.: Evaluating patient and stakeholder engagement in research: moving from theory to practice. J. Comp. Eff. Res. **4**(2), 133–145 (2015). https://doi.org/10.2217/cer.14.79
10. Kieslinger, B., Schäfer, T., Heigl, F., Dörler, F., Richter, A., Bonn, A.: The Challenge of Evaluation: An Open Framework for Evaluating Citizen Science Activities (2018). https://doi.org/10.17605/OSF.IO/ENZC9
11. Saris 2020: "Avaluació Responsable Avaluació per Millorar." Agència de Qualitat i Avaluació Sanitàries de Catalunya. Departament de Salut. Generalitat de Catalunya (2018)
12. Von Schomberg, R.: A vision of responsible innovation. In: Owen, R., Heintz, M., Bessant, J. (eds.) Responsible Innovation, pp. 1–35 (2013). https://doi.org/10.1002/9781118551424.ch3
13. Fecher, B., Friesike, S.: Open science: one term, five schools of thought. In: Bartling, S., Friesike, S. (eds.) Opening Science, pp. 17–47. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-00026-8_2
14. Milat, et al.: Health Res. Policy Syst. **13**, 18 (2015). https://doi.org/10.1186/s12961-015-0003-1

15. Espeland, W., Sauder, M.: Rankings and reactivity: how public measures recreate social worlds. Am. J. Sociol. **113**(1), 1–40 (2007). https://doi.org/10.1086/517897
16. Lewin, K.: Action research and minority problems. J. Soc. Issues **2**(4), 34–46 (1946). https://doi.org/10.1111/j.1540-4560.1946.tb02295.x
17. Kolb, D.: Experiential Learning: Experience at the Source of Learning and Development. Kogan Page, London (1984)
18. MORRI Progress report D3.2. Metrics and indicators of Responsible Research and Innovation. Monitoring the Evolution and Benefits of Responsible Research and Innovation (2015). https://www.rri-tools.eu/documents/10184/47609/MORRI-D3.2/aa871252-6b2c-42ae-a8d8-a8c442d1d557. Accessed 23 July 2019
19. Whyte, W.F.: Participatory Action Research. Sage, Newbury Park (1991). https://dx.doi.org/10.4135/9781412985383
20. Chevalier, J.M., Buckles, D.J.: Participatory Action Research: Theory and Methods for Engaged Inquiry. Routledge, London (2019). https://doi.org/10.4324/9780203107386
21. Feuerstein, M.T.: Partners in Evaluation: Evaluating Development and Community Programmes with Participants. Macmillan Publishers, London (1986)
22. McAllister, K.: Understanding participation: monitoring and evaluating process, outputs and outcomes in rural poverty and environment. Working paper series, 2. International Development Research Centre, Ottawa (1999)
23. Cockburn, A.: Agile Software Development, Addison Wesley, Boston (2002)
24. Kenneth, R.: Essential Scrum: A Practical Guide to the Most Popular Agile Process (2012). ISBN 978-0137043293
25. Few, S.: Bullet Chart Design Specification (2013). https://www.perceptualedge.com/articles/misc/Bullet_Graph_Design_Spec.pdf

Cognitive Aspects and Internet-Mediated Communication

Social Norm Spreading in Real and Virtual Environments: Pro-social Versus Pro-self Norm

Andrea Guazzini[1(✉)], Sara Panerati[3], Viola Filindassi[2], Stefania Collodi[3], and Zoran Levnajic[4]

[1] Centre for the Study of Complex Dynamics, CSDC, and Department of Education, Literatures, Interculture, Languages, and Psychology, University of Florence, Florence, Italy
andrea.guazzini@unifi.it
[2] Department of Education, Literatures, Interculture, Languages, and Psychology, University of Florence, Florence, Italy
[3] University of Groningen, Groningen, The Netherlands
[4] Complex Systems and Data Science Lab, Faculty of Information Studies in Novo Mesto, Novo Mesto, Slovenia

Abstract. It is a fact that technologies do not have fixed effects on people. Some researches show how phenomena such as social influence and conformity appear increasingly multiform and complex, in particular, because people engage more and more in web interactions. The study of conformity in the online environment has highlighted how, to model these dynamics, it is necessary to consider the peculiarities of such environment, since it presents characteristics that differentiate its interactions from the face-to-face ones. Therefore, this research aims at investigating whether and how the type of environment influences the level of conformity to different types of local norms shown to the experimental subjects during a bargaining web-based game. The evidence of the research, conducted on 484 participants, have suggested that there are different psychological processes involved in the conformity phenomenon depending on these factors. The results are discussed in the light of Self-categorization theory, as well as the SIDE model, and illustrate the relevance of considering these processes and their characteristics to promote the implementation of more efficient (and effective) online environments.

Keywords: Social dynamics online · Conformity · Local norms spreading · Ultimatum Game · SIDE model

Highlights

– Conformity to social norms in small group dynamics varies significantly from face-to-face to computer-mediated communication.
– Different spreading dynamics and levels of conformity characterize systems based on Pro-self and Pro-social norms.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 75–90, 2019.
https://doi.org/10.1007/978-3-030-34770-3_6

– Pro-social systems are more contagious in FtF than in CMC, while no differences are revealed for the Pro-Self norm.

1 Introduction

Interest in the new Information and Communication Technologies (ICTs) has grown exponentially in recent years. ICTs have proven to be not only useful communication tools but instruments with great potential also in the improvement of healthcare assistance, as shown by thousands of web applications developed for mental health, support, and training. For what concerns the Psychology field, online environments appear to represent a common context in which services and products are provided to individuals, even therapies are conveyed by ICTs, as represented by Cybertherapy. Furthermore, although online interactions usually take place among isolated individuals, the communication patterns mediated by the technological devices exert influence on the interactors, determining peculiar (social) dynamics in this context. Therefore, it appears particularly relevant understanding what characterizes the online environment and computer-mediated communication, focusing on those dynamics that make this context unique and that differentiate it from face-to-face interactions. In their work, Daft and Lengel [11] identify two main differences between offline and online environments. In the first place, the online environment is considered "poorer" than the offline one. In particular, computer-mediated interactions differ from the face-to-face ones for all the information conveyed by non-verbal communication, such as body language, as well as for social and visual cues, such as age, gender, ethnicity and status. A communication medium can be considered "rich" when it provides immediate feedback (synchronicity) to the user and contains multiple cues (i.e., the possibility to see or hear someone else). All these features appear to be missing in the majority of the media available in online environments. Computer-mediated interactions rarely have the capability to convey non-verbal and social feedback, making it hard to extend to this type of environment complex tasks, characterized by a certain level of ambiguity and uncertainty. In the second place, the other main difference between face-to-face and computer-mediated interactions concerns with the increased anonymity that characterize the online environments. From many literature contributions, emerge that anonymity can lead to a reduction in the individuals' awareness of their behaviours, inhibitory and regulatory conducts, as well as in the others' reactions. The Deindividuation Theory of Zimbardo [48], postulates that reduced levels of interest in self-observation and social evaluation, lead anonymous individuals immersed in the group to engage in aggressive and antisocial behaviours, or, more in general, in those conducts normally inhibited because in conflict with the social norms. According to this theory, online environments should be affected by many negative behaviours, because characterized by more chances to be and to act anonymously. These assumptions do not find a complete confirmation in the following studies, which proved that increased anonymity is not only linked to aggressive or antisocial behavior but instead, it appears to lead also to pro-social behaviours and conformity phenomena [22]. Conformity dynamics characterize individuals interacting in a group context, and when one finds in a small group, the phenomenon of conformity primarily refers to the adoption of rules and norms strictly tied to such context. Various studies have shown their interest in the conformity dynamics [8], and some of them focus on the differences between face-to-face and computer-mediated interactions. In particular, Perfumi, Cardelli, Bagnoli, and Guazzini [32] investigate the phenomenon of conformity online with an adapted version of Ash's experiment. In their study, they found that online conformity increased at the increase of the ambiguity and the difficulty of the task, coherently with what was discovered offline. Moreover, the adoption of the rules and the norms conveyed by the context was lower when the interaction was computer-mediated, rather than face-to-face. This difference was justified with the lack of non-verbal cues that characterized the online environment. More social cues characterize a specific context, then it is more likely that an individual will conform to a norm [25].

On the basis of literature evidence, it has been hypothesized (H1) that social norms adoption in a small group context is influenced by the way in which people interact. In particular, one should expect that lower levels of conformity will be registered in the CMC condition compared to the FtF one.

To study the social dynamics in small groups grew stronger the use of economic games in the last three decades, stimulating the development of different software to fit multifaceted contexts and social phenomena [37]. According to literature, the Ultimatum Game (UG) of Guth, Schmittberger, and Schwarze [20] is the most studied game in experimental economics and it fulfills a paramount role also in Social Psychology research, given its large employment in the investigation of human cultures' prototypical features and social norms [21]. The UG is a game in which a sum of money is shared between two players, the proposer and the responder. The first player is called to split the money with his interactor, who can decide to accept or refuse the allocation. If the responder decides to accept it, then the amount of money is shared as the proposer wished. On the contrary, if he decides to refuse the allocation, then nobody earns anything. Even though the Rational Choice and the Expected Utility Theory postulate that a rational proposer should offer the least amount allowed, and a rational responder should accept all the received allocations, people seem to behave differently. Usually, proposers split the stakes fairly, offering between 40–50% of the surplus, while responders typically reject "unfair" offers, namely allocations between 20–40% of the surplus [18]. Therefore, with the idea of the *"Homo economicus"* on the rocks, social preferences, norms, and emotions appear to be reliable landmarks in the prediction and interpretation of humans' behaviour, also in economic games. Various intercultural studies have been conducted to investigate how the construct of fairness, inequity aversion, an strategic behaviour that characterize the UG can change depending on the social contexts in which the game is played. In particular, from those studies has emerged that the deep variations, in terms of preferences and expectations, that can be observed between groups are tied to norms [21]. According to Bicchieri and Chavez [2], three necessary conditions should be satisfied so that a social norm can form and spread. The first one regards the awareness of people about the existence of a specific norm and the context in which it should be applied. The authors argue that it is necessary that individuals know the norms that characterize their environment and how to behave accordingly with the situation. In fact, as norm unawareness can probably lead to non-conformity, norm salience can probably lead to increased conformity [7]. The second point refers to the importance of the conditional preference for a norm that leads people to follow it. In particular, individuals are more willing to respect a norm that they believe others will follow (empirical expectations) and that think others want them to follow, with punishment as a consequence of transgression (normative expectations). Finally, the third condition that should be satisfied to favour the norm formation and spreading refers to the existence of expectations. In fact, even in the case in which individuals know which norm follow in a specific context, it is necessary that they also have the right kind of expectations about it. Applying these considerations to the Ultimatum Game, it appears that this context can favour the formation of empirical and normative expectations, influencing the players' behaviour. According to the literature, the offers in the Ultimatum Game are driven by a combination of inequity aversion and strategic self-interest [5,21], constructs that are analyzed together with social norms by Fershtman, Gneezy, and List [14]. In their work on Trust and Dictator game they argue that social norms deeply influence players' behaviour, therefore when the "inequity aversion" norm is salient it prevails, but when players can behave selfishly because protected by the existence of a norm that justifies their behaviour, they follow their self-interest. Therefore, individuals appear to be particularly influenced by those social norms that are consistent with their priorities, usually choosing and following "a relevant social norm that justifies a behaviour or an action that maximizes their preferences over outcomes" [14]. In experimental economic games, players' behaviour appears to be deeply influenced not only by the peculiar features of the context but also from the norms that characterize the social aggregates they belong. From literature emerges that different norms have different effects on people's behaviour. In particular, Zhou, Liu, and Ho [47] show that selfish norms appear to be "longer-lasting" and "more impactful" than cooperative norms. Consequently, with the aim of understanding and predicting people's behaviour, context's features and norms, as well as individual's identification with the group he belongs cannot be overlooked [4].

In the light of literature evidence, has been hypothesized (H2) that selfish (or pro-self) and cooperative (or pro-social) norms have a peculiar impact on people's behaviour. In particular, priming the individuals with a pro-self and a pro-social norm will probably have a different effect on their subsequent behaviour. In fact, one can expect that people will conform more to the pro-self norm rather than the pro-social one since it will provide them with a justification for their selfish conducts. Therefore, it can be assumed that pro-self norms will spread more and more easily than pro-social ones, especially in cases in which selfish behaviours can find a justification in the context.

According to the Social Identity Model of Deindividuation Effects (SIDE), one of the key factors in the prediction and the interpretation of people's behaviour is represented by the level of identification that a specific individual has with the group he belongs. Another paramount role is played by the anonymity that, added to identification, foster people's tendency to conform to the norms of such group [38]. Starting from the construct of deindividuation and flipping the perspective of the classical theories about it [12,15,36,48], the SIDE model provides an explanation of the phenomenon of conformity, considering the importance of the norms that characterize every particular social aggregate, namely the local norms. More specifically, the SIDE model conceptualizes deindividuation incorporating two different theories: the Social Identity theory [41] and the self-categorization theory [43]. The first one provides evidence on how people infer some aspect of their identities from the groups they belong, and since they are part of many groups, playing different roles and with different behavioural standards, all these aspects contribute to their self-conceptualization [44]. In everyday life, it appears quite easy to favour the emergence of group identity (generally defined as an individual perception derived from the affiliation with a social group) as suggested by the work of Turner, Hogg, Oakes, Reicher, and Wetherell [43], that provide evidence on the fact that it's enough that people conceptualize themselves as members of the same social category to make group identity develop. Considering the second theory that the SIDE model incorporates, namely the Self-categorization theory, it broadens the previous perspective and provides evidence of the existence of two different types of categorization that people make: personal and social. While in the case of personal categorization, people define themselves through an individual identity that embodies all the attributes that define a person as a unique human being; in the case of social categorization, people refer to the groups they belong to define themselves and their social identity. People experience different social identifications at different time points and in different contexts, therefore is common that numerous categories are used to define themselves (e.g., we vs them, males vs females). Following this consideration, it emerges that when the social identity becomes salient people categorize themselves more as representative members of a particular social category and group than as a unique individual. This categorization process and his consequences (in particular the adherence to the local norms) are automatic, and represent the "cognitive" side of SIDE. According to the model, when the social identity is salient people conform more to the group's norms and tend to behave more following its behavioural standards. Furthermore, there are other factors that come into play in this scenario, such as the anonymity, that is capable of fostering conformity to the local norms when the group member is depersonalized and with a salient social identity [39]. The SIDE model has been often adopted to explain the research results on depersonalization both offline [26,40] and online [34,35]. In particular, the widespread diffusion of the new technologies enabled not only the investigation of phenomenon like the conformity to norms with new methods, moving experiments from laboratory settings to online environments, but also raised questions about the differences that can be observed between face-to-face and the computer-mediated interactions. In those contexts in which information is relatively poor (e.g. CMC), people appear to be more sensitive to salient social norms; furthermore, these contexts seem also capable of setting off a deindividuation process simply on the basis of their features, making group members more sensitive to the influence of rules, stereotypes, social attraction, as well as favouritism towards the ingroup [34]. McKenna and Bargh [29] sustain that offline and online environments share some characteristics, as well they distinguish for others. In particular, in online interactions people can be more easily anonymous and physically isolated, as well as communication can be characterized by a reduction of physical appearance and visual cues. These characteristics have a significant effect on people's behaviour, on the social influence that they exert, and on the one from which are affected in computer-mediated interactions. According to Laporte et al. [25], the differences in conformity among individuals are associated with the number of social cues available in a specific interaction and the communication medium. In their work, the experiment participants were asked to answer to questions (whose ambiguity degree could vary) as members of a group, with the chance to see the other participants through a simple photo or a live video (webcam). Manipulating the degree to which social cues were visible to the experiment participants, the researchers found that higher levels of conformity could be observed in the condition in which there were more social cues, namely in the live video one. Such results appear to be in line with the literature on the construct of visibility, that is conceptualized as a factor capable of setting off categorization processes, that can also occur unintentionally [3], as well as a salient group identity [13,45]. People appear to be highly sensitive to the social context, and this shines through their behaviour, being capable of tuning their conduct with the others' basing on both, explicit and implicit cues, such as nonverbal behaviour [17,24] or conversational pauses [23]. Consequently, pro-social behaviours can derive from a mere "silent identification" that is not only capable of exerting higher levels of conformity but also to reduce the social distance between individuals [4], the fact that has an impact on people's selfish and cooperative tendencies [6].

Starting from the literature evidence can be hypothesized (H3) that an interaction effect can occur between how people interact (face-to-face or computer-mediated) and the type of local norm with which experiment participants were primed (pro-self or pro-social).

The paper is organized as follows. Firstly will be presented the method of the study, divided into three subsections, namely participants and design, procedure, and manipulations. Then, it will be shown the results of the study, that will be discussed in the ad hoc section. Finally, some general considerations about the impact of the present research will be provided in the conclusions of the paper. The main objective of this contribution is to shed light on the differences between face-to-face and computer-mediated interactions in terms of conformity to local norms. These results appear to be useful to deepen the knowledge on the online environments, as well as to design contexts tailored to the norms that characterize it and the consequent behaviour elicited in human beings.

2 Method

2.1 Participants and Design

The sample of the present research was formed by 484 psychology students of the University of Florence (99 males and 385 females) with an average age of 21.99 (SD = 3.70). All the participants were recruited through a voluntary census, and their data have been managed by the guidelines for the ethical treatment of human participants of the Italian Psychological Association (AIP). The experimental design was 2 (environment: CMC and FtF) $\times$ 2 (local norm: pro-self and pro-social), and careful management of the confounding variables was performed. In particular, following the literature evidence, anonymity, social identity salience, individual and group setting emerged as the main influential ones, so their effects have been balanced.

2.2 Procedure

To test our experimental hypothesis, an ad hoc software was developed, and the participants played a modified version of the Ultimatum Game online, acting before as respondents and then as proposers. Both phases of the game were repeated fifteen times, and the participants played a total amount of thirty rounds. Each participant had a personal tablet with which playing, and although he thought to interact with other human players, he played with a system of bots. Such a system was composed of 5 playing bots, and since the rounds were 15 for each game's phase, the experimental participant could randomly encounter the same player more than one time.

2.3 Manipulations

Two independent variables were taken into account in this study: the type of environment and the type of local norm. Although the participants always played the game through the tablet, manipulation of the type of environment in which they were playing was carried out. In particular, they could be randomly assigned to two different conditions: the FtF and the CMC. In the FtF condition participants could see each other during the game, while in the CMC condition they could not. Therefore, while in the first condition they could also exploit visual cues to interact with the other players, in the second one, their interaction was limited to the tablet. The second independent variable manipulated in this study was the type of local norm, that was generated using the average donation (4.11 resources) observed within the population in the Ultimatum Game context as a threshold (see [42] for a meta-analysis). The below-threshold norm was renamed "pro-self" and referred to an amount of 3.5 resources, while the above-threshold norm was renamed "pro-social" and referred to an amount of 6.5 resources. To prime a specific norm, the participants as first thing played as respondents, receiving 15 allocations of different sizes, whose average could be 3.5 euros or 6.5 euros based on the condition to which they were randomly assigned. Participants were explicitly instructed to maximise their gain, considering the operating rules of the classic Ultimatum Game, and having the chance to accept or refuse the received allocation. Once the receiving phase was over, the participants played as proposers, allocating at will a sum of money between 0 and 10 euros for 15 times to the other players, always considering the operating rules of the Ultimatum Game.

3 Results

To verify the preconditions necessary for the inferential analyses on the experimental data, the normality of the distribution of the continuous variables was assessed through the analysis of asymmetry and kurtosis values. Consequently, to compare results per conditions 2 (environment: CMC and FtF) $\times$ 2 (local norm: pro-self and pro-social) an inferential analyses using a general linear mixed model (GLMM) approach [28] has been used, due to the structure of the repeated measures of the experimental data. In the study the "adherence to the local norm" was considered as dependent variable and it was operationalized as the gap between the norm given in the receiving phase (i.e., 3.5 or 6.5 euros) and the average allocation of the experimental participant during the donation phase. Therefore, higher was the distance between the provided norm and the average amount of money donated by the participant, lower was the adherence to the local norm. The deviation from the provided local norm has been considered as a measure of anti-conformity behavior. The results of the study will be presented following the experimental hypothesis, therefore three subsections have been developed.

3.1 Effects of Face-to-Face and Computer-Mediated Communication on Norms Spreading in Small Groups

To understand the spreading dynamic of local norms within small groups of interacting John Does, the GLMM technique has been used to explain the relationship between the type of environment in which the interactions take place and the levels of conformity to the local norms. In particular, in this subsection are presented the discrepancies in local norms adherence associated with the face-to-face and the computer-mediated communication. As shown in Table 1, a significant difference on the levels of conformity to the local norm has been observed between the two environments ($t = -4.29$, $p < .001$), with the FtF characterized by a higher adherence to the norm compared with the CMC one. In other words, the participants were more anti-conformist when communication was exclusively tied to the provided technological device, and no other visual cues were involved. Focusing on norms, in the next subsection will be shown how to different types of norms correspond to different levels of conformity (or anti-conformity).

Table 1. Generalised linear mixed models 1. Effects of the type of environment on local norms conformity

Type of environment and local norm conformity				
Best model	*Akaike*		**F**	**Df-1(2)**
	21970.61		18.38***	1(7258)
Fixed effects				
Factor	**F**	**Df-1(2)**	**Coefficient (β)**	**Student t**
CMC vs FtF	18.38***	1(7258)	−0.11	−4.288***

$^{***} = p < 0.001$; $^{*} = p < 0.05$.

3.2 Pro-self and Pro-social Norms Spreading in Small Groups

Starting from the hypothesis that not all the types of norms spread in the same way in small interconnected groups, the levels of conformity to pro-self and pro-social norms have been compared through a GLMM model. Table 2 shows how a significant difference between the adhesion to the norms has emerged in the small group context (t = −18.13, p < .001). In particular, the pro-self norm appeared to be characterized by higher levels of conformity and diffusion. Finally, to shed light on the interaction between the effects of the face-to-face and computer-mediated communication on the different types of norms spreading, a model of this complex relationship is provided in the next subsection.

Table 2. Generalised linear mixed models 2. Effects of the type of norm on local norms conformity

Type of norm and local norm conformity				
Best model	*Akaike*		**F**	**Df-1(2)**
	21667.53		328.67***	1(7258)
Fixed effects				
Factor	**F**	**Df-1(2)**	**Coefficient (β)**	**Student t**
Pro-self vs Pro-social	328.67***	1(7258)	−0.46	−18.129***

$^{***} = p < 0.001$; $^{*} = p < 0.05$.

3.3 Interaction Effects on Norms Spreading in Small Groups

With the aim of testing the third hypothesis of the study, namely that an interaction between the type of environment and the type of norm can be detected on the conformity levels to the local norms, a GLMM approach has been used. From the previous results has emerged that higher levels of conformity to the norm can be observed when the interaction within the small group happened face-to-face (even if a technological device mediated it), as well as when the norm was pro-self. Combining the effects of the face-to-face and the computer-mediated communication with the pro-self and the pro-social norms has emerged a significant interaction between these variables on the levels of conformity to the local norm (t = −4.29, p < .001). As shown in Table 3, participants appeared to adhere differently to the experimental local norm depending on the type of the environment. In fact, the highest levels of conformity were detected in the condition in which the small group members were interacting face-to-face, and were primed norm was the pro-social one.

Table 3. Generalised linear mixed models 3. Combined effects of type of environment × Type of norm on local norms conformity

Environment × Norm and local norm conformity				
Best model	*Akaike*		**F**	**Df-1(2)**
	21654.95		117.29***	3(7256)
Fixed effects				
Factor	**F**	**Df-1(2)**	**Coefficient (β)**	**Student t**
Pro-self vs Pro-social	328.01***	1(7256)	−0.51	−14.352***
CMC Vs FtF	17.14***	1(7256)	−0.16	−4.523***
Pro-social + FtF	4.97*	1(7256)	0.11	2.230*

*** $= p < 0.001$; * $= p < 0.05$.

4 Discussion

The present study contributes to an interdisciplinary line of research that involves knowledge coming from different fields, especially from social sciences, economic sciences, and information engineering. In particular, the purpose of this contribution is to increase the understanding of what one should expect in terms of conformity to norms, when different types of norms are involved, as well as when such phenomenon is investigated through face-to-face or computer-mediated interactions in the small group context. To do so, a customized economic game has been administered to the experimental sample. In particular, the participants played a switched version of the Ultimatum Game through face-to-face or computer-mediated interactions (repeatedly playing before as responders and then as proposers), as well as they were primed with a salient norm (selfish or cooperative) in the first phase of the game. Taking inspiration from the large body of literature concerning the differences between face-to-face and computer-mediated interactions, and by controlling the most important confounding variables identified by literature, has been hypothesized that conformity to the local norms would be lower in the Ultimatum Game played through computer-mediated interaction than in face-to-face (H1). Previous researches [8,32] have argued that although similar dynamics characterize offline and online interplays, and that a certain level of conformity can be observed online, it's important to underline that those tasks that were performed through computer-mediated communication registered lower levels of rules' adhesion with respect to those carried out through face-to-face interaction. In line with previous evidence, in this study has been found that participants conformed more to the local norm when playing the UG face-to-face, confirming the first hypothesis of the research. In particular, the experimental sample showed increased levels of anti-conformist behaviour when interacting through a technological device without any available visual or social cue about the interactor (CMC condition), as well as increased levels of conformity occurred when people played, always by means of a tablet, with many accessible cues (FtF condition). Therefore, results interpretation appear to revolve around the availability of cues in a certain context, whether they are social or visual, that seem to contribute to the phenomenon of conformity. In fact, as the number of cues increases, also the information that one has about the interactor increases, fostering the individual perception of the others, as well as the intention to conform to the norms of the group. Moreover, according to the literature on the Ultimatum Game, people's behaviour is made of a combination between inequity aversion and strategic self-interest, factors that have brought scholars to the wording of *"Homo economicus"*, currently in contrast with the construct of *"Homo sociologicus"*. While the first one belongs to the economic field and tries to model humans' conduct, the second belongs to sociology and proposes a vision of the individual behaviour as almost entirely dependent from the context, and in particular, from norms. Recent literature on economic games argues that people acts combining personal interests and norms' adherence, showing a preference for those norms that provide them with an excuse for selfish behaviours. Consequently, as stated by Fershtman, Gneezy, and List [14] "when socially acceptable actions provide one person with a greater portion of the rents, people will exert extra effort to secure those rents, to the detriment of the other player". To compare the participants' conformity levels to the pro-self and the pro-social norm in the small group context (H2), independently from the way in which interactions took place, participants' allocations have been analyzed with the norm primed in the first phase of the game. In accordance with literature evidence, the results show that higher levels of conformity were registered for the pro-self norm, namely participants appeared to be more influenced by the pro-self norm rather than the pro-social one during their allocations. Therefore, it can be assumed that people found justification for their self-interested behaviour in the pro-self norm made salient at the beginning of the game. This result appears to be supported by the economic and the psychological literature which shows that, compared to cooperative norms, the selfish ones seem to be particularly impactful and more easily absorbed by individuals [16]. Furthermore, also Zhou and Ho [47] provide evidence consistent with our results for a wider context and its norms, namely the cultural norms. In their study, the authors argue that selfish norms are more influential than cooperative ones and that it is enough to expose an individual characterized by a cooperative background to a selfish environment to elicit a pro-self playing behaviour in an economic game. At the same time, the opposite is not true. In fact, just one exposition to a cooperative environment is not sufficient to elicit a pro-social behaviour in a player with a selfish background, determining any change in his behaviour. If one of the objectives of the literature is to model humans' behaviour through the employment of economic games, is necessary that this branch of knowledge takes into consideration all the factors that play a role in determining people's conducts. Norms represent one of these factors. Humans' behaviour in economic games appears to be deeply influenced by them, whether they are strictly tied to the context, like the local norms, or they belong to broaden scenarios, such as social and cultural ones. Social norms arise, develop, and spread through the interaction among individuals, who not only contribute to the creation of those norms but are also deeply influenced by them. Consequently, norms and group standards appear to be particularly influenced also by the context in which they occur. As shown in the previous section, this study reveals the existence of an interaction between the type of norm to which an individual has been exposed and the way in which people interact, confirming the study's hypothesis (H3) and determining two different patterns of conformity elicitation. In particular, the degree of adherence to the pro-self standard does not present significant differences if considered in relation to the face-to-face or the computer-mediated interaction. This reasonably indicates that, regardless of the type of environment, the selfish norm usually prevails. On the other hand, considering the interaction between the type of norm and how people interact, one can detect a certain deviation in the degree of adhesion to the norm. In fact, those who played the game in the condition in which they could see each other (i.e., face-to-face), conformed significantly more to the pro-social norm than those who instead were forced to interact without any visual cue. In this case, the availability of visual cues about the other members of one's group, have probably played a paramount role in determining a categorization process of the participants as members of such group [25], making the social identity salient, that represents a fundamental factor capable of predicting higher levels of conformity to the group's norms. As shown above, norms spreading was greater when participants interacted in face-to-face, and the deviation that occurred in the adherence to the pro-social standard may have been affected by this mechanism. At the same time, no differences were likely observed for the pro-self norm dependently from how players interacted, since individuals held such norm as a personal one, acting following it also when the social identity was not salient. On the other hand, for what concerns the pro-social norm, the issue is more complex. In fact, one can hypothesize that concerning the face-to-face environment, the presence of a salient social identity has brought the participants to adhere to the group's standards (i.e., the type of donation considered legitimate in that context). On the contrary, in CMC and given the absence of these psychological mechanisms, individuals have played following their standards (that is, being able to maximize their gain), thus being less influenced by the experimental norm. These considerations are endorsed by many literature contributions that have shown how, in CMC, the absence of psychological processes, such as the salience of social identity [31], or more simply the lack of communicative exchanges [33] and visual cues, is able to influence the process of adherence to the norms, in particular reducing it. Cinnerella and Green [8] have suggested that these characteristics can influence the behaviour of the participants, since in such conditions they may have difficulty in forming a sense of belonging to the group and therefore a common and salient social identity that requires the presence of additional contextual factors to explain themselves. These results would therefore be in line with the SIDE model [39], according to which the development of a social identity linked to belonging to one's group and the maintenance of the latter as salient and meaningful, would be able to produce a higher level of conformity, even in an online environment. In fact, perceiving their social identity as salient would lead individuals to become more sensitive to the social stimuli present in the environment, which in turn will make them more inclined to adapt themselves to the local regulations that characterize such group [22]. If this process of depersonalization does not occur, individuals will tend to act according to their personal standards. Since this process strongly depends on the context, it is not surprising that the participants deviated from the experimental standard, thus conforming less to the local pro-social norm respect to those who interacted in face-to-face.

5 Conclusion

From our perspective, it is essential to stress the importance of research on social dynamics in the online environment, both from a strictly scientific point of view, as well as for the many practical applications that this knowledge allows to reach. In particular, thanks to this research vein, many projects of psychological interest have been promoted and implemented, exploiting the potentiality that the online environment owns. To name a few, the use of new technologies combined with this scientific knowledge has led to the development of prevention projects in virtual environments against cyberbullying [10,46], phobias [19], and for the treatment of psychological diseases - e.g., depression [30]. Furthermore, research related to conformity and social influence in online environments has proved its usefulness in many areas and has paved the way for many different applications, including health in the workplace (e.g., [1]), where the use of avatars and game-based activities are always more common in many companies (e.g., for the personnel selection). Also, the construct of depersonalization has shown a full versatility in many other fields besides those just mentioned; in fact it has been applied to learning processes in groups, for example in schools as shown by the research of Le Hénaff, Michinov, Le Bohec and Delaval [27], as well as in the field of advertising and persuasive messages [9]. Those just mentioned are just some of the points where research on conformity and social influence in the virtual environment has contributed, also in a practical sense (e.g., helping to design the ergonomics and functioning of many software and online platforms). Although not exhaustive, these examples allow us to focus on the importance of modelling people's behaviour in online environments; in fact, knowing which variables are involved in defining the social dynamics and which interactions are significant to be able to amplify or decrease their effects, will lead to the implementation of projects and services that can exploit this knowledge to develop tailored environments for their users. All these evidence contribute to bring out the importance of this branch of research and, in particular, provide a well-established framework in which our study places.

Acknowledgments. We wish to thank Dr. Silvia Grasso for her invaluable assistance in the management of the experiment.

Funding. This research received no external funding.

Conflict of Interest. The authors declare there is no conflict of interest.

References

1. Armstrong, M.B., Landers, R.N., Collmus, A.B.: Gamifying recruitment, selection, training, and performance management: game-thinking in human resource management. In: Emerging Research and Trends in Gamification, pp. 140–165. IGI Global (2016)
2. Bicchieri, C., Chavez, A.: Behaving as expected: public information and fairness norms. J. Behav. Decis. Mak. **23**(2), 161–178 (2010)
3. Biernat, M., Vescio, T.K.: Categorization and stereotyping: effects of group context on memory and social judgment. J. Exp. Soc. Psychol. **29**(2), 166–202 (1993)
4. Bohnet, I., Frey, B.S.: Social distance and other-regarding behavior in dictator games: comment. Am. Econ. Rev. **89**(1), 335–339 (1999)
5. Camerer, C.F.: Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton (2011)
6. Charness, G., Gneezy, U.: What's in a name? Anonymity and social distance in dictator and ultimatum games. J. Econ. Behav. Organ. **68**(1), 29–35 (2008)
7. Cialdini, R.B., Kallgren, C.A., Reno, R.R.: A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. In: Advances in Experimental Social Psychology, vol. 24, pp. 201–234. Elsevier (1991)
8. Cinnirella, M., Green, B.: Does 'cyber-conformity' vary cross-culturally? Exploring the effect of culture and communication medium on social conformity. Comput. Hum. Behav. **23**(4), 2011–2025 (2007)
9. Coggrave, S., Oldmeadow, J.A.: Social inhibition of message processing: effects of presence of others on processing persuasive messages
10. Cross, E., Richardson, B., Douglas, T., et al.: Virtual Violence: Protecting Children from Cyberbullying. Beatbullying, London (2009)
11. Daft, R.L., Lengel, R.H.: Organizational information requirements, media richness and structural design. Manag. Sci. **32**(5), 554–571 (1986)
12. Diener, E.: Deindividuation: the absence of self-awareness and self-regulation in group members. The psychology of group influence 209242 (1980)
13. Elsbach, K.D., Glynn, M.A.: Believing your own"PR": embedding identification in strategic reputation. Adv. Strat. Manag. **13**, 65–90 (1996)
14. Fershtman, C., Gneezy, U., List, J.A.: Equity aversion: social norms and the desire to be ahead. Am. Econ. J. Microecon. **4**(4), 131–144 (2012)
15. Festinger, L., Pepitone, A., Newcomb, T.: Some consequences of de-individuation in a group. J. Abnorm. Soc. Psychol. **47**(2S), 382 (1952)
16. Fisman, R., Miguel, E.: Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets. J. Polit. Econ. **115**(6), 1020–1048 (2007)
17. Forgas, J.P., Jones, R.: Interpersonal Behaviour: The Psychology of Social Interaction. Pergamon Press, Sydney (1985)

18. Gabay, A.S., Radua, J., Kempton, M.J., Mehta, M.A.: The ultimatum game and the brain: a meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. **47**, 549–558 (2014)
19. Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness Iii, T., Botella, C.: Virtual reality in the treatment of spider phobia: a controlled study. Behav. Res. Ther. **40**(9), 983–993 (2002)
20. Güth, W., Schmittberger, R., Schwarze, B.: An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. **3**(4), 367–388 (1982)
21. Henrich, J., et al.: In search of homo economicus: behavioral experiments in 15 small-scale societies. Am. Econ. Rev. **91**(2), 73–78 (2001)
22. Johnson, R.D., Downing, L.L.: Deindividuation and valence of cues: effects on prosocial and antisocial behavior. J. Pers. Soc. Psychol. **37**(9), 1532 (1979)
23. Koudenburg, N., Postmes, T., Gordijn, E.H.: Resounding silences: subtle norm regulation in everyday interactions. Soc. Psychol. Q. **76**(3), 224–241 (2013)
24. Kulyk, O., Wang, J., Terken, J.: Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings. In: Renals, S., Bengio, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 150–161. Springer, Heidelberg (2006). https://doi.org/10.1007/11677482_13
25. Laporte, L., van Nimwegen, C., Uyttendaele, A.J.: Do people say what they think: social conformity behavior in varying degrees of online social presence. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 305–314. ACM (2010)
26. Le Hénaff, B., Michinov, N., Le Bohec, O.: Applying the side model to brainwriting: the impact of intergroup comparison and anonymity on creative performance. J. Appl. Soc. Psychol. **48**(7), 351–359 (2018)
27. Le Hénaff, B., Michinov, N., Le Bohec, O., Delaval, M.: Social gaming is inside: impact of anonymity and group identity on performance in a team game-based learning environment. Comput. Educ. **82**, 84–95 (2015)
28. McCulloch, C., Searle, S., Neuhaus, J.: Generalized, Linear, and Mixed Models. Wiley, New York (2001)
29. McKenna, K.Y., Bargh, J.A.: Plan 9 from cyberspace: the implications of the internet for personality and social psychology. Pers. Soc. Psychol. Rev. **4**(1), 57–75 (2000)
30. Pagliari, C., et al.: Psychosocial implications of avatar use in supporting therapy for depression. Annu. Rev. Cybertherapy Telemed. **181**, 329–333 (2012)
31. Perfumi, S.C., Bagnoli, F., Caudek, C., Guazzini, A.: Deindividuation effects on normative and informational social influence within computer-mediated-communication. Comput. Hum. Behav. **92**, 230–237 (2019)
32. Coppolino Perfumi, S., Cardelli, C., Bagnoli, F., Guazzini, A.: Conformity in virtual environments: a hybrid neurophysiological and psychosocial approach. In: Bagnoli, F., et al. (eds.) INSCI 2016. LNCS, vol. 9934, pp. 148–157. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45982-0_14
33. Postmes, T., Lea, M.: Social processes and group decision making: anonymity in group decision support systems. Ergonomics **43**(8), 1252–1274 (2000)
34. Postmes, T., Spears, R., Lea, M.: Breaching or building social boundaries? Side-effects of computer-mediated communication. Commun. Res. **25**(6), 689–715 (1998)
35. Postmes, T., Spears, R., Lea, M.: Intergroup differentiation in computer-mediated communication: effects of depersonalization. Group Dyn. Theory Res. Pract. **6**(1), 3 (2002)

36. Prentice-Dunn, S., Rogers, R.W.: Effects of public and private self-awareness on deindividuation and aggression. J. Pers. Soc. Psychol. **43**(3), 503 (1982)
37. Rabin, M.: Incorporating fairness into game theory and economics. Am. Econ. Rev. **83**, 1281–1302 (1993)
38. Reicher, S.D.: Social influence in the crowd: attitudinal and behavioural effects of de-individuation in conditions of high and low group salience. Br. J. Soc. Psychol. **23**(4), 341–350 (1984)
39. Reicher, S.D., Spears, R., Postmes, T.: A social identity model of deindividuation phenomena. Eur. Rev. Soc. Psychol. **6**(1), 161–198 (1995)
40. Reicher, S., Levine, R.M., Gordijn, E.: More on deindividuation, power relations between groups and the expression of social identity: three studies on the effects of visibility to the in-group. Br. J. Soc. Psychol. **37**(1), 15–40 (1998)
41. Tajfel, H., Turner, J.C., Austin, W.G., Worchel, S.: An integrative theory of intergroup conflict. Organizational identity: a reader, pp. 56–65 (1979)
42. Tisserand, J.C.: Ultimatum game: a meta-analysis of the past three decades of experimental research. In: Proceedings of International Academic Conferences International Institute of Social and Economic Sciences (2014). No. 0802032 in Antibes ISBN 978-80-87927-05-2
43. Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S.: Rediscovering the Social Group: A Self-categorization Theory. Basil Blackwell, Cambridge (1987)
44. Vilanova, F., Beria, F.M., Costa, Â.B., Koller, S.H.: Deindividuation: from Le Bon to the social identity model of deindividuation effects. Cogent Psychol. **4**(1), 1308104 (2017)
45. Waldzus, S., Schubert, T.: Group norm and category norm in anonymous situations: two sources of social influence. na (2000)
46. Wolke, D., Sapouna, M.: Fear Not: An Innovative Interdisciplinary Virtual Intervention to Reduce Bullying and Victimisation. The Impact of Technology on Relationships in Educational Settings. Routledge, Abington (2012)
47. Zhou, X., Liu, Y., Ho, B.: The cultural transmission of cooperative norms. Front. Psychol. **6**, 1554 (2015)
48. Zimbardo, P.G.: The human choice: individuation, reason, and order versus deindividuation, impulse, and chaos. In: Nebraska Symposium on Motivation. University of Nebraska Press (1969)

On Internet Resilience in Small Island States

Patrick Hosein(✉), Shiva Ramoudith, and Kim Mallalieu

The University of the West Indies, St. Augustine, Trinidad and Tobago
shiva@lab.tt, {patrick.hosein,kim.mallalieu}@sta.uwi.edu

Abstract. Internet service disruption is at best an inconvenience and at worst devastating for many businesses, governments and individuals. Small Island Developing States (SIDS), which are particularly vulnerable to natural disasters, often enjoy the least resilient network architectures and suffer the longest periods of disruption when service is most needed. As authoritative Domain Name System (DNS) servers generally lie off-island, the failure of international access links cripples local service availability, even when domestic links are physically operational. This paper proposes a simple monitoring strategy to reduce the impact of infrastructure losses in such circumstances. It examines the probability that local hosts are unreachable due to unavailable DNS mapping data gathered over a monitoring period and reports on a complementary strategy, and associated platform, to manually capture missing mappings. Together these mechanisms generate a set of local DNS mappings that enable SIDS' Internet Exchange Points (IXPs) to perform intra-island routing independent of external DNS servers.

Keywords: Domain name service · DNS · Network availability · Internet resilience

1 Introduction

Small Island Developing States (SIDS) [5] face specific social, economic and environmental vulnerabilities [1]. Some important characteristics of SIDS include their high dependence on external and remote markets, long distances from export markets and import resources, and low resilience to natural disasters such as storm surges, landslides, droughts and weather systems [4].

Natural disasters often result in a loss of Internet connectivity in SIDS. As for the rest of the world, Internet connectivity is increasingly important for these states during all stages of the disaster risk management (DRM) cycle: mitigation, preparation, response and recovery. Many DRM systems rely on it for the quick dissemination and receipt of critical information both within the affected jurisdictions and with the outside world.

Many regulatory frameworks include broadband provisions in universal service obligations of operators. Such regulations and associated national policies,

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 91–96, 2019.
https://doi.org/10.1007/978-3-030-34770-3_7

consider access to the Internet as a right of all citizens, alongside that of water and electricity. There is also increasing interest in obligations and other provisions relating to network resilience and service availability during and following natural disasters. Simple, low-cost provisions are the most practical and sustainable for SIDS which are prone to high failure rates in connectivity at the backbone and local levels. Several such solutions together represent an effective means of increasing the overall resilience of SIDS.

2 Problem Statement

In order to access a website a user types the URL, say www.google.com, in the address bar of a browser. Among other things, this URL specifies a specific computer which hosts the website on the Internet. While the URL is a convenient way for human users to remember Internet addresses, computers recognize each other by their Internet Protocol (IP) addresses, not their URLs. An example IP address for Google is `172.217.11.14`. A Google search may be performed by typing `172.217.11.14` into a browser. In a similar manner, specific handsets in the telecommunications network are identified, and may be accessed, by their telephone numbers.

In order to retrieve a web page a computer must map the URL to its corresponding IP address. This is equivalent to looking up a phone number in the Telephone Whitepages. The Domain Name System, DNS, performs this mapping. The DNS uses a hierarchical structure to provide the lookup function. For example, if we consider accessing google.com, the computer first queries a root server (called a nameserver) that contains the locations (IP addresses) of servers that contain information about .com addresses. It then queries those servers to obtain locations of servers that can resolve google.com addresses and finally it queries the google.com nameservers to obtain the IP address associated with google.com [3].

As nameservers are distributed around the Internet, the DNS function cannot be fully performed when international links are down. This impacts both access to sites within the affected country and abroad. That is to say, for example, that if persons in an affected area wish to access the website of the local disaster management agency, they may not be able to do so even if the physical infrastructure were unaffected. Local connectivity is not only important for disaster mitigation and preparation, but it is invaluable for disaster response and recovery. Access provisions during the critical hours of the disaster and prior to the restoration of regular service following the disaster, is for some the difference between life and death.

3 Proposed Approach

In principle, the impact of a loss of access to external nameservers may be mitigated by maintaining a copy of all possible DNS mappings on the Internet, in-country. While this is impractical, it is reasonable for SIDS to maintain DNS mappings for all hosts located in their particular country. These hosts are the computers capable of local connections. We recommend this approach to ensure that all hosts in a country remain locally accessible through their regular URLs when physical connectivity is available locally but international connectivity is lost.

We propose that each ISP provides DNS service for its customers and that the log files of all DNS requests are scanned daily. All mappings that are performed with destination IPs located on the island should be saved. Over time a list of all hostnames associated with local IP addresses will be generated. When access to the global Internet is lost the local IXP can use this saved list of mappings to perform the DNS function. Customers who have manually changed their DNS settings will not be accommodated.

ISPs are expected to apply rules of retention which specify the monitoring period and flush mapping entries outside of it. The proposed strategy will generate the mappings of all local hosts accessed within the monitoring period but will not capture those for hosts not accessed over the monitoring period. The authors have developed a simple web platform to capture the mappings of sites which are important during disasters but which may not have been accessed within the relevant monitoring period. Domain names of sites which have never been accessed, or those that have not been accessed recently, are manually entered on the platform's web-based interface, and a script is run to extract all associated IP addresses. This includes websites, mail servers, etc. All mappings to local IP addresses are stored and can be added to the mappings obtained by the ISPs to be used in times of disaster. The platform, may be publicly accessed at https://dnscache.lab.tt.

4 Analysis of the Solution

To examine the likelihood of missing mappings over a monitoring period, we consider that there are K active services (web sites, email servers, etc.) with locally assigned IP addresses on the island. Multiple services may map to a single IP, as for example a single web server with virtual hosting support; and a single service may map to multiple IPs, for example for fault tolerance. All services must be mapped to their IP/s once the IPs are local.

Access to services is not a stationary process. Though the rate of access is likely to vary by time of day, day of week, month, etc., this basic analysis assumes only variations by time of day. Each day is segmented into small periods over which the request rate for a particular service is assumed constant. We further assume that the request rate for a given service, k, is the same for the same period each day. We denote this request rate λ_k. Since the number of potential clients that access the service is large and their requests are independent, we assume that requests may be modelled using a Poisson process.

We normalize the duration of a period under consideration to unit length and assume that during the last instance of this period an outage occurs at which time the regular DNS lookup cannot be performed. We assume that data had been collected for this site over the previous N periods, that is to say over N days prior. We next compute a bound on the miss probability, that is to say the probability that an attempt to access a service failed as it had not been accessed over the monitoring period, and a mapping had therefore not been stored.

Let us consider a single service k. The probability that a request for access is made during the outage period is simply $1-e^{-\lambda_k}$ since the period is one unit. The probability that the site was not accessed during the monitoring period is the probability that it was not accessed during the N prior periods times the probability that it was not accessed during all remaining time that it was monitored, that is to say those times that do not fall during the concerned periods. We denote the latter probability by β. We can now write the miss probability, P_k, for this site as the product of the probability that an attempt was made to access it during the outage period times the probability that it was not accessed during the monitoring period, hence a mapping was not recorded.

$$P_k = (1-e^{-\lambda_k})e^{-\lambda_k N}\beta \leq (1-e^{-\lambda_k})e^{-\lambda_k N} \tag{1}$$

Now consider the function

$$F(\lambda) \equiv (1-e^{-\lambda})e^{-\lambda N} \tag{2}$$

Note that this function is zero when $\lambda = 0$ as well as when λ tends to infinity. Taking the derivative of F with respect to λ and setting to zero we get

$$-Ne^{-\lambda N} + (N+1)e^{-\lambda(N+1)} = 0 \tag{3}$$

and solving we obtain $e^{-\lambda} = N/(N+1)$. Note that the second derivative of F with respect to λ is given by

$$F''(\lambda) = N^2 e^{-\lambda N} - (N+1)^2 e^{-\lambda(N+1)} \tag{4}$$

and evaluating at the inflection point obtained above we get

$$F''(-\ln(N/(N+1))) = -Ne^{-\lambda N} < 0 \tag{5}$$

and so the inflection point is a maximum. Evaluating F at this maximum we get

$$F(-\ln(N/(N+1))) = \frac{1}{N+1}\left(\frac{N}{N+1}\right)^N \tag{6}$$

Let us denote by $F^*(N)$ the maximum value of $F(\lambda)$ over λ for a given value of N. We plot $F^*(N)$ as a function of N in Fig. 1. We can now use this upper bound on $F(\lambda)$ in Eq. 1 to obtain

$$P_k \leq F^*(N) \tag{7}$$

and finally we can compute the average probability of miss over all sites as

$$P_M = \frac{1}{K}\sum_{k=1}^{K} P_k < F^*(N) \tag{8}$$

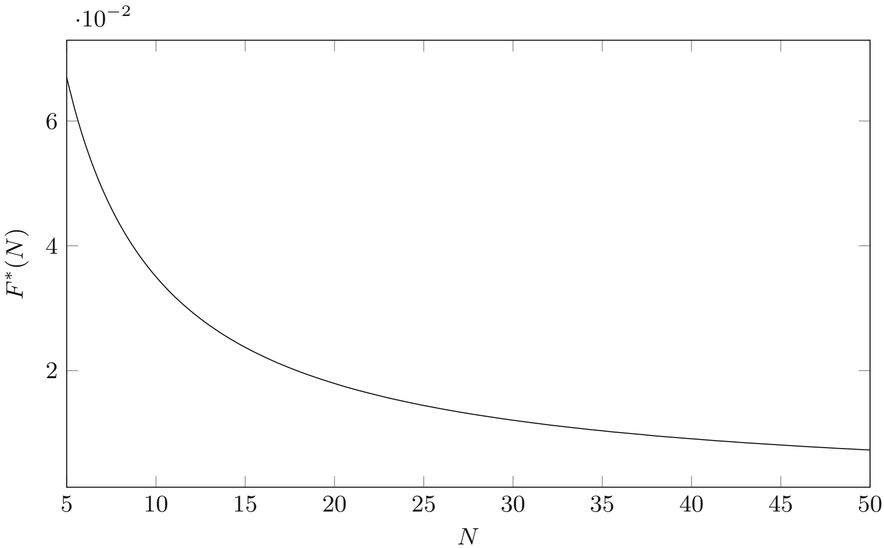


Fig. 1. The function $F^*(N)$

Note that this bound on the probability of miss was computed for a period of time over which the request rate was constant. We can break up all of time into such periods but the above bound holds for each of them. Therefore we conclude that the miss probability at all times is bounded by $F^*(N)$ where N is the ratio of the monitoring period and the outage period.

Consider, for example, the case of $N = 42$ corresponding to six weeks of monitoring. We have $F^*(42) = 0.009$ which means that if we monitor requests for a 6 week period and have an outage of 1 day then the average miss probability will be less than 1%. This low value can be explained by the following observation. If a site is very active then the probability that it is accessed at least once over the six week period will be close to 1 and hence the probability of miss is close to zero. If on the other hand a site is very rarely accessed then the probability that it is accessed during the outage period will be close to zero and hence the miss probability will also be close to zero.

In practice the Poisson arrival assumption may not always hold. For example, there may be sites that are rarely accessed under normal conditions but once a disaster takes place they are frequently accessed. This is the case, for example, for sites that are used to manage disasters. To capture these missed sites, a DNS information collection platform has been built. The platform accepts manual entries for sites, at risk of not being captured over a monitoring period, in advance so that they are accessible during the response and recovery phases of the DRM cycle. Between these and the local sites captured through DNS log files at the caching nameserver, all sites likely to be accessed during these critical phases will be accessible.

5 DNS Information Collection Platform

The DNS information collection platform is a web application built on the Django framework [2]. The website presents a simple interface through which users can submit a domain name that they would like to be cached. A Captcha ensures that entries are made by humans. A script is run to determine which, if any, DNS records of the domain map to IP addresses located within the country. These mappings are then stored. Since these mappings are subject to change, another script periodically scans the cached data and checks for any updates. If the island becomes totally disconnected from the Internet, this cached data can be added to the data collected from DNS log files and used to provide the DNS function locally.

6 Conclusions and Future Work

The increasing reliance on Internet resources and services, alongside the increasing risks of natural disasters, exacerbate existing vulnerabilities of small island developing states. At the same time, the building of resilience in such countries must take keen account of local constraints. It is for this reason that a variety of simple, low-cost measures that together respond to the most critical needs, is the most likely strategy to gain traction and yield results. This paper has proposed, and performed preliminary examinations on, a complementary strategy to retain local routing capability in the face of loss of access to DNS servers off-island.

Implementation of the strategy calls for its adoption by local ISPs, voluntary submissions of URLs by agents involved in the disaster risk management ecosystem, and the maintenance of a very simple portal which is available for replication. It is recommended that local regulators and policy makers opportunistically promote the system amongst ISPs and IXPs.

References

1. Briony, J.G.: Building resilience in small island developing states: social media during the 2017 atlantic hurricane season. In: International Conference of Information Systems for Crisis Response and Management Pan Asia 2018 (08/11/18) (November 2018). https://eprints.soton.ac.uk/425191/
2. Holovaty, A., Kaplan-Moss, J.: The definitive guide to Django: web development done right. Apress (2009)
3. Liu, C., Albitz, P.: DNS and BIND, 5th edn. O'Reilly Media Inc., Newton (2006)
4. Trivedi, K.S., Kim, D.S., Ghosh, R.: Resilience in computer systems and networks. In: Proceedings of the 2009 International Conference on Computer-Aided Design, pp. 74–77. ICCAD 2009, ACM, New York, NY, USA (2009). https://doi.org/10.1145/1687399.1687415,http://doi.acm.org/10.1145/1687399.1687415
5. Wong, P.P.: Small island developing states. Wiley Interdisc. Rev.: Clim. Change **2**(1), 1–6 (2011). https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.84

Science Cafés and Science Shops for Public Engagement

Giovanna Pacini[1,3] and Franco Bagnoli[1,2,3](✉)

[1] Department of Physics and Astronomy and Center for the Study of Complex Dynamics, University of Florence,
Via G. Sansone 1, Sesto Fiorentino, Florence, Italy
{giovanna.pacini,franco.bagnoli}@unifi.it
[2] INFN, sez. Firenze, Florence, Italy
[3] Associazione Caffè-Scienza Firenze-Prato, Florence, Italy

Abstract. We illustrate a method for enhancing the participation of citizens in University research and sustainability activities through a synergy between science cafés and science shops.

The cultural association "Caffé-Scienza" (science café) in Florence and Prato have accumulated a 14-year experience in organizing science café events, with several variations, and web and radio broadcasting support. Most of events concerns sustainability. We have also used this instrument for evaluating and discussing the results of European projects, and also to co-design new ones.

Two years ago, we joined the European project InSPIRES, which is fostering the science shop ideas. A science shop is a methodology by which universities and research centers can offer services to the population. If a request coming from the civil society can be answered by means of a scientific investigation, the shop contacts its researcher and invites them to propose the investigation to students. The results are then returned to citizens.

We are now experimenting with a more participative approach, derived in part by our experience with science cafés, in which the research proposal is discussed and elaborated in a public event, and citizens are sometimes directly involved in the investigation, in a way similar to citizen science.

We discuss here the main results of the first two years of activity, perspectives and obstacles we have met, and the possibility of expanding the services by means of the web, social nets and the Internet.

Keywords: Science shop · Science cafés · Public engagement · Sustainability

1 Introduction

The main difficulty encountered by many projects is that of involving citizens. This is particularly true or projects concerning sustainability themes, for which involved citizens are supposed to modify their habits. But this difficulty is present also for other topics. For instance, a lot of projects about Collective Awareness Platforms for Sustainability and Social Innovation (CAPS) started with quite innovative ideas but failed due to the lack of participation [1, 3, 7, 16, 22].

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 97–108, 2019.
https://doi.org/10.1007/978-3-030-34770-3_8

On the other hand, the participation of citizens to projects should be welcomed, even in the design phase. One would have in this way the possibility of anticipating doubts, fears, concerns, ethical questions, and so on.

Moreover, involving citizens in the process of research is one of the priorities of the European Commission, that invested many efforts in defining the roadmap of the Responsible Research and Innovation (RRI) principles [15].

The University of Florence is quite interested in sustainability themes, and in through its Center for the Study of Complex Dynamics (CSDC) and a supporting association (Caffè-Scienza) has contributed in organizing several science cafés. Recently, we participated in a European Projects (InSPIRES) [13] with the goal of developing a more participative methodology for science shops, exploiting the science café tool and related communications. The activities of the University of Florence in the field of sustainability are quickly revised in the following Section.

Another approach for a direct participation of the population is represented by those projects in which citizens are personally involved as "scientists" in the sense that they volunteer collect and/or provide data required for specific research, i.e., citizen scientists [23]. However, despite the amount of information emerging from citizen science projects, the practice is not universally accepted as a valid method of scientific investigation by the scientific community itself [8].

The promotion of participation can arise from a different direction, namely what in the language of the European Union is termed "science and society". In this framework, a well-established and democratic way of developing a shared knowledge is constituted by science cafés, which are described in detail in the next section.

In this methodology, the scientific knowledge is transmitted on a peer basis, and this practice favors the active participation of attendees. As we shall illustrate, in Florence we have tested several variations of this methodology, not only for science popularization but also for the dissemination of the results of publicly funded projects and the co-design of proposals.

Another interesting aspect for involving people is constituted by science shops. They are mainly hosted by universities and research centers, and are "windows" open to the population, that can signal problems or propose idea. Accepted themes are then investigated by university students, and the results are returned to the population. Science shops are illustrated more in details in Section "Science shops" and the related projects that are being carried out in Florence are reported in Sections "The pilot science shop project in Florence" and "Other science shop projects". Finally, conclusions are drawn in the last Section.

2 Sustainability at the University of Florence

In the last years many people of the Florence university community share a strong interest in environmental sustainability, and they are ready to change their lifestyle. The director of the University decided in 2015 to establish a group of work sustainability, to increase the communication and the research projects in this area.

Fig. 1. University of Florence, one of the fountains located in the University premises and in the first level the aluminum bottles.

Since 2015 the Sustainable University group was involved in significant initiatives that are inserted in the University of Florence's strategic plan. One of those is, at example the Tuscany students' card, that, including the subscription to public transport, represents a strong incentive for sustainable mobility.

Other important projects are the 20,000 aluminum bottles (see Fig. 1) distributed to the Florence University students that, with the seventeen water fountains located in various University buildings represents a strong action versus the "zero plastic bottle" cause, that is one very important point of "zero waste" policy, as it is well explain in an explicative video [24].

On the other point the University is working on the correct recycling of waste, first establishing the Sesto Fiorentino campus special garbage's collection points- "Eco-tappa" (they are now six in many University buildings), for small appliances, cartridges and toner for printers, spray containers, used batteries, expired medicines.

Moreover, they developed the two university gardens The first one is in the "Giardino dei Semplici" Botanical Garden of the University of Florence and it is prepared as demonstration gardens suitable also for small spaces and inspired by the "Ortobioattivo" concept [14]: interested people can participate concretely in the cultivation of vegetables, exchange experiences and practice in new techniques (see Fig. 2).

The second one is located in the educational complex of Viale Morgagni (student house), where students themselves supply the University canteen with products from their garden [25].

3 Science Cafés

One of the main defects of "standard" communication channels, like articles, radio and tv programs, web sites, and often also social media, is that they are (or are perceived) as "one way": the information flow from experts to participants.

In a conference there is a physical distance between the speaker/expert and the public, and also the choice of the place (a conference room) helps to ward off these two

Fig. 2. Different modalities of Science cafés: (a) in a conference room of the public Library "Oblate", in the center of Florence; (b) a Cine-science at the Cinema "La Compagnia", Florence; (c) a science café in the Central Market of Florence.

figures. In a certain sense, it is as if the conference takes place in the expert's home, and the attendees were guests.

As a result, the flow of information and interactions are almost always one-way, with a few questions from the audience, confined at the end of the intervention. Participation is therefore very selective and are unlikely to involve a generic public. From the point of view of the interaction, the conference is still better than other forms of dissemination such as articles, television and radio broadcasts, but the latter reach a wider audience.

In 1992, Duncan Dallas, a previous BBC Director, launched in Leeds the idea of a Science Café, anticipating the revolution in science communication from the "Public Understanding of Science" to the "Public Engagement of Science and Technology" [9]. A science café is a conference turned upside down [4, 5, 12, 17].

The advent of the Internet has also allowed to use tools such as podcasts to distribute the recording of events and even streaming, introduced mainly from the science café in Florence to be able to reach people that for geographical or heathy problem can't reach the venue of the event but want to participate at distance [19].

The cognitive aspects of science cafés are some of the most important aspects of this methodology as the need to actively involve the public in the transmission of information, avoiding the TV effect, and simultaneously the need to avoid taking premature position on hot topics, since there. Many of these aspects survive also when using virtual interactions.

In Florence, this activity is carried out by the non-profit association "Caffé-Scienza" (of which the authors are members), formed by academic and CNR researchers, but also many "ordinary" people. This association organizes, from 2000, the traditional science cafés (caffé-scienza) every week and other initiatives such as the junior science cafés (with high school students), or cine-science (movie-based events). As in other science cafés, our public is mainly composed by middle-age, highly educated people.

Nowadays, we organized more than 300 events, with several variations: in markets, in libraries, streamed on YouTube, with online participation, using movies (Cine-Scienza) and so on (Fig. 2). They have also experimented with radio support (9 years), and they have animated the Italian Network of Science Cafés [20]. Most of the times, they are discussing themes linked to sustainability like energy, transportation, pollution, wastes, health and so on. Science cafés have also been used as an instrument for evaluating and discussing the results of European projects, and also to co-design new ones.

4 Science Shops

A science shop ("shop" in the sense of "exhibition", they are more like "windows" or "counters") is a methodology by which universities and research centers can offer services to the population.

Citizens can express requests and needs, and if these themes can give origin to a research, they are assigned to university students as part of their Ph.D., master or B.sc. degree, or part of their traineeship or, finally, as part of an examination, always under the guidance of a researcher. The results are then restituted to the population.

Science Shops [11] are one of the practical realizations of the principle that the universities and research institutes should open their doors offering services directly addressed to citizens, clearly combining them with their fundamental mission of education, formation and research.

In fact, in addition to the classic missions of the universities, such as teaching and research, a third mission was added in the last years in Italy, concerning the transfer of research results to society. It regards both the technological transfer of research (patents, spin-offs, third-party contracts and agreements, intermediaries) and the production of social and cultural public goods (public engagement, cultural heritage, continuous training, clinical

In a certain sense, from the moment that in Italy the universities are funded at least in part by resources coming from the general taxation, this engagement provides a return of the investments to the civil society and the socio-economic structures of the society.

The activities belonging to the third mission contribute, in Italy and other countries, to the evaluation of the universities and research centers and thus concur to the establishment of their financing quotes.

In Italy, Legislative Decree 19/2012, which defines the principles of the self-assessment, periodic evaluation and accreditation system, and subsequently the Ministerial

Decree 47/2013, which identifies the indicators and parameters for periodic research evaluation and the third mission, recognized the third mission as an institutional mission of the universities, alongside the traditional missions of teaching and research. The products mentioned above therefore become part of the evaluation of each Italian University (VQR).

Actually, the possibilities offered by the Internet are not particularly exploited [17]. At present, the authors are participating in the European Projects about Science Shops InSPIRES (Ingenious Science shops to promote Participatory Innovation, Research and Equity in Science), trying to merge them with their experiences [13, 18].

One of the main goals of this project is that experimenting with a more participative approach, in which the research proposal is discussed in a public event like a science café, and citizens are directly involved in the investigation, in a way similar to citizen science, exploiting our experience with science cafés.

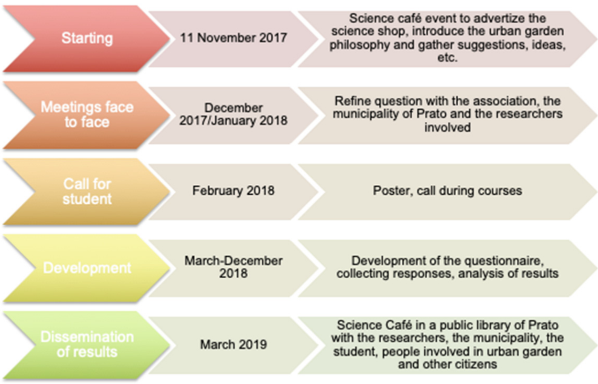


Fig. 3. Schedule of the urban garden science shop project.

The Florence Science Shop is trying to involve throughout the process not only the associations and the stakeholders involved in the research question, but the whole citizenship, opening the doors to citizens both in the beginning phase, and in the development phase and in the final dissemination phase.

5 The First Science Shop Project: Urban Gardening

The term "urban garden" combines two words that, in the common language, have opposite values: the idea of the vegetable garden is linked to the countryside, whereas with urban we normally refer to the cities and in general to the industrialized areas [26].

This practice has been traditionally used for economic reasons (substance agriculture), but now it has also acquired educational, nutraceutical, therapeutic and urbanistic aspects [10].

The theme of urban gardening is the core of the first project of the Florence Science Shop. It was planned according to the new methodology, trying to enhance the participative and collaborative approach during all the process, from the initial idea to the dissemination of results.

We received a manifestation of interest from the non-profit association Orti Dipinti in 2018, who runs a public “open” garden where citizens can learn how to deal with plants and gather aromatic herbs. The garden is used also as a didactic experience for children and therapeutic tool for mentally impaired people. They asked us to help them in promoting this idea, enlarging the participants to their and other experiences in this field.

Our first step (Fig. 3) was to organize a science café to illustrate the request and the spirit of the science shop and to collect questions on this specific topic, in order to set up a research project. The experts of our science café were Ugo Bardi, a physical chemist at the University of Florence and the University delegate for sustainability, Marina Clauser, from the Botanic Garden of Florence, and Giacomo Salizzoni, president of Orti Dipinti. Franco Bagnoli, who moderated the event, introduced the subject and the idea of Science shop to the public. Ugo Bardi, during his activity as the delegate for sustainability, introduced the idea of urban gardens run by students in the university and Marina Clauser is hosting school classes in the Botanic Garden in Florence.

Fig. 4. Urban garden in Prato, photo by UNIFI researchers.

Both citizens and experts proposed research issues and expressed their needs, like for example: to create a network for the exchange of information, the necessity to have a dialogue with the institutions, the requirement of financial support to urban gardens for social/recreational purposes, and the desire to have answers to questions about the new methods and techniques in horticulture.

After the event, other researchers from the Department of Agri-food Production of and Environmental Sciences of the University of Florence (DISPAA) joined the project.

We performed a first analysis [6] about the wide varieties of realizations of the idea of publicly accessible urban gardens open. Among them, those that are assigned by municipalities to retired or unemployed citizens are the most common and those that globally involve most citizens. So, we decided to further investigate this aspect.

From the stakeholders' point of view: the Municipality of Prato expressed its interest in monitoring its assigned public gardens (Fig. 4), and a thesis project on this subject started with the title "Impact of urban horticulture on water resources: the case of the social gardens of the Municipality of Prato".

An agriculture student, 24 years old, following the guidelines of the researchers, developed a questionnaire to evaluate both the agronomic aspects and the social/psychological impacts.

Finally, we organized a science café in close collaboration with the municipality of Prato in a public place (a municipal library), to return the results of the project to the population. Following the new idea that has permeated the whole process, we have invited as speakers not only the student and the researchers but also the Councilor for the Environment of the municipality of Prato and one of the elderlies who runs one of the analyzed urban gardens [6].

6 The "Perception of Science" Project

A second project was on the development of a survey about the perception of science. The research question was proposed by our Caffè-Scienza association (but not by ourselves). The proposers would like to know if science and the figure of the scientist have lost credibility in recent years, given the strong increase in people who believe in pseudoscience or overt scientific hoaxes (vaccines, chemtrails, etc.).

After a couple of meeting between the association's team, the Florentine Science Shop team and the psychologic department, the research team was formed, consisting of a researcher, a Ph.D. student and a graduating student.

Instruments currently present in the scientific literature have been selected, classifying them on the basis of their psychometric properties, their possibility of use with respect to more age groups (invariance of items with respect to the age of respondents), more or less pervasive use by part of the scientific community, to the administration procedure.

Simultaneously with the selection of the most suitable tools for the detection, a web-based application was developed using GoogleScript, capable of managing the online administration of the instruments using devices such as PCs, smartphones and tablets (Fig. 5).

The application allows any user who has the link to access in and to create an account in an anonymous but identifiable manner (use of nicknames) and, after this first phase, to respond to the selected tools as discussed above.

The app updates the instrument's average in real time whenever a new user responds to that particular survey. Therefore, its positioning, as the population of respondents increases, can change and hence the need to allow consultation of one's score at different times than the time of administration. The application was also designed to be able to add additional evaluation.

The participants at the perception of science survey were 381. The processing of data is still ongoing, and they will be presented in a science-café that we are organizing for the end of June

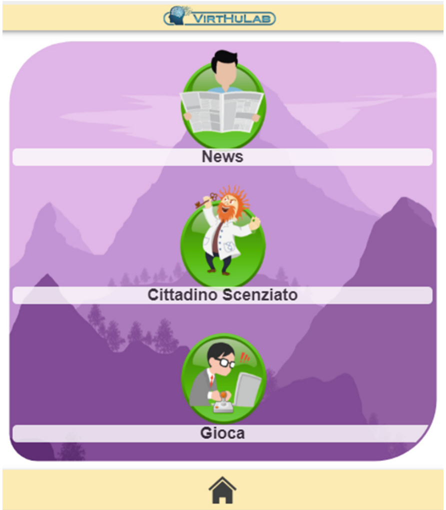


Fig. 5. The App interface for the survey about the perception of science.

7 Other Science Shop Projects

Before starting the InSPIRES project the authors collaborated in other projects closely related to public engagement and in which the science café as a method has been used for disseminating the results. As an example, we present a citizen science project on sustainability, involving the development of a compact low-cost air quality monitoring station, named AIRQino, and its hosting by citizens [21].

This device is based on a platform, developed by the Institute of Biometeorology (IBIMET-CNR) of Florence, constituted by an “Arduino” open-source hardware, on which low-cost sensors are integrated for monitoring the climate and air quality in urban areas. based on an “open” type of architecture able to be modified and adapted to different monitoring needs. It allows the creation of mobile stations for a more complete monitoring of quality in urban areas, in line with Directive 2008/50/ EC, which recognizes and regulates the importance of additional measures compared to those of fixed stations (Fig. 6).

A spatial data infrastructure was also developed that is composed of:

- a central GeoDatabase, for data storage and management
- a GIS engine, for data localization
- a WebApp for viewing in real time, querying and processing data

The AQM platform has been flanked by a platform developed by a local company, “Magenta s.r.l.” for real-time monitoring of traffic flows in urban areas. This system has been distributed to groups of citizens who agreed to host the instrument (outside a window overlooking some street) let the device use their own wi-fi for data transmission.

We presented the results of the research in a science café titled "Traffic Flow: A participatory approach to traffic monitoring (and more …)" (Science café, 2015).

Two other projects about sustainability are in progress at Florence Science Shop, both of them proposed by students' associations. The first one concerns the quality of air in the Florence University area. The scientific center of the University of Florence is located in an extra-urban area of the city, very close to the airport and to the city dump. An association of students of science contacted the science shop asking if it would be possible to perform a research to evaluate the air quality in that area.

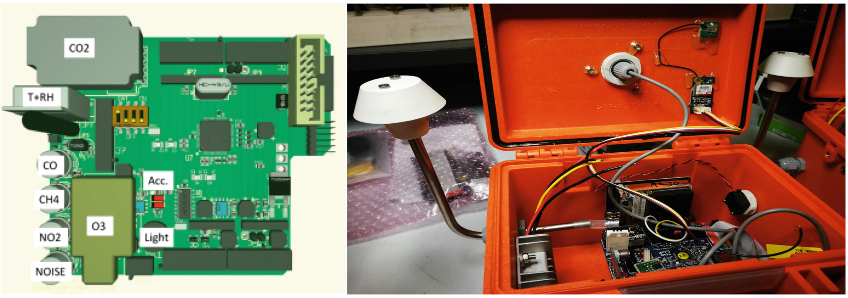


Fig. 6. (left) Hardware developed (Arduino); (right) Spatial data infrastructure composed of a central GeoDatabase, a GIS engine, a WebApp for viewing in real time.

After some meetings with researchers coming from the physic department and the CNR, National Research Council, four research projects for a master thesis are designed:

- Comparison/calibration of measuring systems of the mass of the pm10 with "standard" gravimetric measurements;
- Analysis of a large data-set composition of the pm10 championship in the area of the scientific campus;
- Sampling and analysis of the particulate with high temporal results;
- Development of a sampler with high temporal resolution.

8 Conclusions

The weakest aspect of many educational and research projects is the engagement part. Reaching citizens and having them participate in a constructive way in publicly funded program is difficult, and often the efforts and resources wasted in communication and promotion are not proportionate to that allocated for actually carrying out the activities.

One of the main defects of "standard" communication (articles, radio and tv programs, web sites, and often also social media) is that they are (or are perceived) as "one way": the information flow from experts to participants.

In Florence, we have a 14-year experience in organizing science café, and we have animated the Italian Network of Science Cafés. Most of times, we are discussing themes linked to sustainability like energy, transportation, pollution, wastes, health and so on.

We have also used this instrument for evaluating and discussing the results of European projects, and also to co-design new ones.

Recently, we joined and European Projects about Science Shops (InSPIRES), trying to merge them with our experiences [13]. A science shop ("shop" in the sense of "exhibition", they are more like "windows" or "counters") is a methodology by which universities and research centers can offer services to the population.

We are now experimenting to use the science shop methodology with a more participative approach, in which the research proposal is discussed in a public event like a science café, and citizens are directly involved in the investigation, in a way similar to citizen science.

In those first two years we learnt that sustainability themes attracted a lot of interests from students and citizens, as resulted from the participation in science cafés and related events. Thanks to the science shop students have a way of directly participating in sustainability researches and citizens can explain their needs and help with the research itself, coupled with citizen science initiatives.

Acknowledgements. The Florence Science Shop "La bottega della Scienza e della Sostenibilità a Firenze" was born thanks to the EU project InSPIRES, "Ingenious Science shops to promote Participatory Innovation, Research and Equity in Science" [13]. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 74167. The university co-gardening experience is supported by the Green Office of the University of Florence [2].

References

1. Arniani, M., et al.: Collective awareness platforms for sustainability and social innovation: an introduction (2014). http://booksprints-for-ict-research.eu/wp-content/uploads/2014/07/BS5-CAPS-FIN-003.pdf
2. Ateneo sostenibile: (2019) https://www.ateneosostenibile.unifi.it/index.php
3. Badii, A.: Collective awareness platforms and digital social innovation mediating consensus seeking in problem situations. In: Bagnoli, F., et al. (eds.) INSCI 2016. LNCS, vol. 9934, pp. 55–65. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45982-0_5
4. Bagnoli, F., Pacini, G., Politi, P.: Science is ready, serve it! Eur. News **42**(3), 4–5 (2011)
5. Bagnoli, F., Pacini, G.: Sipping science in a café. Procedia Comput. Sci. **7**, 194–196 (2011). https://doi.org/10.1016/j.procs.2011.09.070
6. Bagnoli, F., et al.: Urban gardening in florence and prato: how a science shop project proposed by citizens has grown into a multi-disciplinary research subject. J. Sustain. Dev. **11**, 111–119 (2018)
7. Bagnoli, F., Guazzini, A., Pacini, G., Stavrakakis, I., Kokolaki, E., Theodorakopoulos, G.: Cognitive structure of collective awareness platform. In: 2014 IEEE Eighth International Conference on Self-adaptive and Self-organizing Systems Workshops (SASOW), London, pp. 96–101 (2014). IEEE. https://doi.org/10.1109/sasow.2014.38

8. Bonney, R., et al.: Next Steps for Citizen Science. Science **343**(6178), 1436–1437 (2014). https://doi.org/10.1126/science.1251554
9. Dallas, D.: The café scientifique. Nature **399**, 120 (1999). https://doi.org/10.1038/20118
10. Duchemin, E., Wegmuller E., Legault A.-M.: Urban agriculture: multi-dimensional tools for social development in poor neighbourhoods. Field Actions Sci. Rep. **1** (2008). http://journals.openedition.org/factsreports/113
11. European Commission: EUR 20877 Science shops - knowledge for the community. Office for Official Publications of the European Communities, Luxembourg (2003). ISBN 92-894-6246-9
12. Ferris, D.: Science cafés tap nation's fascination with research and discoveries. Wired (2007). https://www.wired.com/2007/12/science-cafs-tap-nations-fascination-with-research-and-discoveries/
13. The InSPIRES project web site. http://inspiresproject.com/
14. Ortobioattivo: Semele cooperative (2018). http://www.ortobioattivo.com
15. Owen, R., Macnaghten, P., Stilgoe, J.: Responsible research and innovation: from science in society to science for society, with society. Sci. Public Policy **39**, 751–760 (2012)
16. Pacini, G., Bagnoli, F.: Results of a collective awareness platforms investigation. In: Bagnoli, F., et al. (eds.) INSCI 2016. LNCS, vol. 9934, pp. 19–26. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45982-0_2
17. Pacini, G., Bagnoli, F., Belmonte, C., Castellani, T.: Science is ready, serve it! Dissemination of science through science cafè. In: Bucchi, M., Trench, B. (eds.) Quality, Honesty and Beauty in Science and Technology Communication, PCST 2012 book of papers. Observa Science in Society (2012)
18. Pacini, G., Bagnoli, F.: InSPIRES: science shops 2.0. In: Kompatsiaris, I., et al. (eds.) INSCI 2017. LNCS, vol. 10673, pp. 69–78. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70284-1_6
19. Pacini, G., Bagnoli, F.: Science cafés in the Internet era. In: Kompatsiaris, I., et al. (eds.) INSCI 2017. LNCS, vol. 10673, pp. 79–89. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70284-1_7
20. Rete italiana dei caffè scienza: (2019). http://www.caffescientifici.it/
21. Science cafè "Traffic Flow: A participatory approach to traffic monitoring (and more …)", Florence, 29 May 2015. http://www.caffescienza.it/programma-2015-2016/chest
22. Sestini, F.: Collective awareness platforms: engines for sustainability and ethics. IEEE Technol. Soc. Mag. **12**, 53 (2012). https://doi.org/10.1109/MTS.2012.2225457
23. Silvertown, J.: A new dawn for citizen science. Trends Ecol. Evol. **24**(9), 467–471 (2009). https://doi.org/10.1016/j.tree.2009.03.017
24. Trevers, B.: Sustainability pills. Avoid buying plastic water bottles (2017). https://www.youtube.com/watch?v=H8EHqMUB0Bo
25. University co-gardening: Social initiative of the University of Florence (2017). https://www.unifi.it/art-2745-ateneo-sostenibile-presentati-gli-orti-in-viale-morgagni.html
26. Verde urbano: Caffè-scienza (2019). http://www.caffescienza.it/programma-2018-2019/verde-urbano

Use of Information and Communication Technologies for Obtaining Public Social Services in Russia

Veronika V. Romanenko(✉)

Saint Petersburg University, Saint Petersburg, Russia
nikar@yandex.ru

Abstract. The article deals with digitizing of state social services that started recently in Russia. We present an analysis of the legislative grounds and statistical indicators of the development of information and telecommunication systems. In particular, we address some aspects of the use of digital technologies in interacting with clients of social public services in one of the largest cities in the country, St. Petersburg. Our conclusions are based on the studies of website contents and on interviews with representatives of social services. The development of the digital economy in Russia and its new tools rapidly change the usual relationship between the state and the social client. During the current transition period, new electronic platforms are being created, e-governance is being developed and various types of assistance are being universalized.

As a result, it was concluded that the process of transforming the quality of social services due to digitalization is multidimensional and complex; it has not only positive effects (accessibility, time saving, bureaucracy reduction, automation of statistics collection, etc.), but is also associated with certain risks (unequal access to information, dehumanization of interactions between the client and the social service, technical failures, etc.). The employees of social service centers have a double responsibility: to develop strategies for online interaction with customers and at the same time to maintain the same standards of services for those who cannot use the Internet. The main limitation in the introduction of advanced technologies in the sphere of public social services is related to the financing of this process: in spite of new standards, it is not always possible to introduce technologies immediately, social service centers often need special facilities or more advanced computer equipment, and as well as additional staff training.

Keywords: Public services · Digitizing social services · E-governance · Electronic public services · Quality of public services

1 Introduction

In the modern world, information has become a resource in the social and economic relations, which are increasingly moving into the network space. A key factor in the transformation of the activities of market entities is the development of digital culture. Analysis of the changes taking place in the context of the development of Internet

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 109–120, 2019.
https://doi.org/10.1007/978-3-030-34770-3_9

technologies and the virtualization of social space is an important scientific problem. Its comprehension implies the formation of new conceptual approaches to the system of social services. On the one hand, in order to be comfortable and progressive, social services must switch to innovative configurations of public service delivery systems. On the other hand, the state social service system should neutralize risks that always accompany changes. The state has to accept the function of a guarantor that supports the development of appropriate technological platforms, effective communication between authorities, business and society, including the development of their network interaction [1].

Today, the Russian social services system has to become client-oriented and flexible as much as possible while maintaining stability and ensuring the rights of citizens [2]. There is a need to develop new legislative bases for the functioning of this system, taking into account the rapid development of technologies and of business and controlling new socio-economic inequalities. This article analyzes the efficiency of the social service delivery and accounting platforms recently introduced into the state service cluster, and also examines how information and communication technologies (ICT) are used in government social assistance institutions in St. Petersburg.

2 Significance of the Research

The transition to digital economy, including in the social services sector, and the creation of e-governance are not just new social phenomena that require scientific reflection, but also innovative processes that require constant analysis, development of predictive models and performance evaluation. The relevance of the topic is determined by the necessity for public attention to ongoing changes that have not yet become the subject of discussion from the viewpoints of various social groups those interests they affect.

The elaboration of digital technologies for automation of public services work has been performing in Russia for a long time, but until recently databases storing information about customers and services provided to them were fragmented and not always integrated even at the municipal level. Moreover, in order to receive a public service, it was usually necessary to collect a large package of documents and be physically present for registration in an appropriate office.

The application of innovative digital technologies on the national level should solve a number of critical tasks, such as the maximum simplification and speeding up the process of providing and receiving services, reducing the number of intermediaries and bureaucracy, improving and unifying the system of informing citizens, expanding the structure of digital services for citizens, simplifying the procedures of collecting statistics and forecasting for special services.

The priority of the digital communications development in the socio-economic life of the Russian society is reflected in several regulations adopted between 2010 and 2018 (presidential decrees, government regulations, amendments to existing laws in the field of social services [3–6], government programs – “Information society” [7], “Digital economy” [8]).

In these legal acts the concept of "digitalization" is encountered; however it is usually related to technical issues and is understood as a transition from the analog form of information transfer to the digital one, without affecting the underlying socio-economic influences on societal institutions. The study of these processes and the expansion of the definition of digitalization can be a topic for future sociological research. At the same time, in this study we consider the digitalization of social services as a transition to a digital method of communication, recording and transmitting data with the help of digital devices in the process of providing these services to the clients [9].

It is believed that the very concept of the digital economy was introduced in 1994 by the Canadian entrepreneur Don Tapscott (now Executive Chairman of the Blockchain Research Institute, Adjunct Professor at INSEAD and Chancellor of Trent University in Ontario) [10]. In his book "Digital Economy", he described the key characteristics of society in the digital age, such as globalization, virtualization, innovations and elimination of mediation institutions. Since then, a methodology has been developed to measure the level of development of various ICT indicators, a lot of research has been done in the field of computer sciences. However, in Russia, studies of the subject are considerably less numerous and mostly dated by 2015–2018. Thus, according to Rostelecom (Russia's leading long-distance telephony provider), the total amount of scientific articles, devoted to the topic of digital technology development, published in China, is amounted to more than 407.000 articles, in the USA – more than 234.200, in Russia – more than 41.000 [11]. Among them all, only a very small number is concerned with the intersection of social behavior and computational systems, include development of electronic social services.

This is not surprising, as the Government of the Russian Federation has recently chosen to ensure the accelerated introduction of digital technologies in the social sphere as one of the priority goals of the country's development (Presidential Decree of May 7, 2018) [3], but the prerequisites for the transition to digitalization of such catch-up sectors of the economy as state social services are already formed.

The rate of household access to the Internet in the country is approaching 80%, more than 60% of citizens over 15 years old use the Internet every day. Obtaining information and services in the electronic form has become a daily practice for the majority of Russians (see Table 1) [12].

Table 1. Internet development in Russia (according to National Research University Higher School of Economics)

	2012	2013	2014	2015	2016	2017
Rate of households with Internet access, %	60	67	70	72	75	76
Percentage of population using the Internet almost every day normalized by the total number of population aged 15–74 years, %	41	48	52	55	58	61
Percentage of population using the Internet to receive state and municipal services in the electronic form amongst the number of population aged 15–72 years that has received state and municipal services in the past 12 months, %	Data is absent	31	35	40	51	64

The International Telecommunication Union (ITU) report for 2017 showed that Russia takes 45th place (IDI = 7.07) out of 176 countries by the ICT development index (IDI). The first position is taken by Iceland (IDI = 8.98) followed by South Korea, Switzerland, Denmark, United Kingdom, with the last place is taken by Eritrea (IDI = 0.96) [13]. In addition, according to the ITU, in 2018 Russia took 26th place out of 175 countries in the ranking of countries by the ICT - Global CyberSecurity Index (GCI = 0.836), with the first place awarded to the United Kingdom (GCI = 0,931) [14].

According to the e-government development index (EGDI) for 2018, Denmark holds a leading position in the world in providing public services and information via the Internet, followed by Australia and South Korea. In this ranking Russia took 32nd place (EGDI = 0,7969) out of 175 countries [15].

These indicators show that the Russian Federation is on its way to intensive development of the digital economy and the formation of effective e-government services, which unavoidably concern the processes of social services delivery.

3 Research Methodology

The aim of this study is to analyze the role of information communication technologies in obtaining public social services in Russia. We are interested in the aspect of the demand for services by users, as well as how flexible the system is when interacting with the client. This task has been divided by us into several subtasks; we used different research methods for each one.

During the descriptive study we collected data on the work of the largest e-government site, the Unified Portal of Public Services GOSUSLUGI (www.gosuslugi.ru). We have used the following methods: statistics collection using web-analytics services Similarweb (www.similarweb.com), Alexa (www.alexa.com), Analysis (a.pr-cy.ru); analysis of secondary data from sociological studies; collection of statistics provided on the site.

We have also studied the operation of the United State Social Security Information System (EGISSO) portal (www.egisso.ru). To do that, the method of structured description of site sections and statistical data presented by its analytical system was used.

Those portals allow one to receive some government services online, to reduce time costs and to automate workflow, as well as to provide full information about the available types of social assistance together with the ways of obtaining them. However, there exist social services delivered only in the physical presence of the recipient, e.g. medical and social support for the disabled persons, for the elderly, or for families with children. This assistance is carried out by various institutions at the place of residence of the clients. 20 sites of comprehensive social service centers and rehabilitation centers for disabled people and disabled children were analyzed in order to study their attendance, usability and ways of interacting with clients. For this purpose, the service Analisys (a.pr-cy.ru) and personal observations of the testing person were used.

Additionally, conversations were held with the director of the integrated social services center of the population and the head of the social support department of the center of social rehabilitation for persons and children with disabilities, as well as with their specialists.

4 Results

4.1 Description of the Work of the Unified Portal of Public Services

Access to the Unified Portal of Public Services (GOSUSLUGI) and Unified State Social Security Information System (EGISSO) is subject to an identification and verification procedures (using The Unified Identification and Authentication System). Signing up requires some identification information such as first and last names, date and place of birth, phone number, social insurance account number, as well the individual taxpayer number. After completion the identification stage, the user gets an access to basic functions actually providing the information on the full range of services available online. The full functionality becomes available upon completion of the verification stage. Then the citizen gets an access to his or her taxes, fines, social assistance, ID documents such as the international passport, etc. All the information, inquiries, demands, and payments are accessible from any location in the world and do not require any physical presence. The result is the drastic reduction in physical waiting time and in the workflow.

The Unified Portal of Public Services GOSUSLUGI has been operating since the end of 2009. Today, according to the traffic analysis, it is one of the most visited websites in Russia (Country Rank = 13, Global Rank = 303, data of service of web-analytics Similarweb, available 05.05.2019). In April 2019, the number of unique site visitors reached 88 million (see Fig. 1). By the end of 2017, the number of site users counted 65 million [16], while in 2018 their number reached 86.5 million [17] which represents about 60% of the country's population.

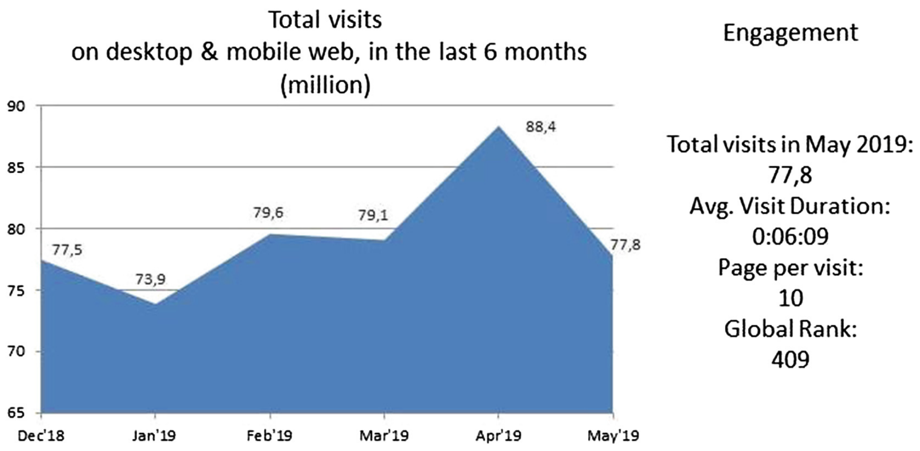


Fig. 1. Dynamics of the gosuslugi.ru site traffic (web analysis via the Similarweb service).

The government services portal GOSUSLUGI solves the following tasks:

- informing about state and municipal services;
- ordering government services in the electronic form whenever possible;

– registration for admission to various state offices;
– electronic payment of fines, court and tax debts, state fees, housing and communal charges;
– assessment of the quality of service.

Since the access to many social, medical and educational services is provided by regional organizations having the associated budgets, Russian regions progressively become connected to the GOSUSLUGI Government Services Portal. Depending on the region, these services can include getting child care benefits, kindergarten and school enrollment, and medical appointments.

The most popular site services are electronic payment of state duties, debts and fines. Thanks to a convenient payment system, discounts for early repayment of fines, and generally high transaction speed, Russian citizens made in 2018 48.2 million successful payments with the total of 56.4 billion rubles (more than $ 831,000). In addition to paying taxes, judicial debts and automobile fines, most popular services include getting a statement on the retirement account. Many other electronic document management services are also in high demand, such as submissions for obtaining a passport, obtaining or replacing a driver's license, vehicle registration and domicile registrations, as well as electronic appointments.

Social services on the site are mainly services of an informative nature (extract of the state of the retirement account, extract of the possibility of receiving social assistance, informing about the procedure for obtaining assistance, etc.). But some of the services can be obtained online by going to the regional subsystem of public services (apply for benefits, compensation, payments).

According to 2017 statistics, the most searched-for electronic state and municipal services required by population aged between 15–72 years were related to of services received in electronic form were services related to (in percent to the total population receiving services): health and medicine (42%), services of the state traffic police and the Ministry of the Interior (35%), taxes and fees (28%), housing and communal services (21%), education (15%), social security (7%) [12]. Social security is the least consumed online service, largely due to the complexity of its delivery to the target population consisting of people with low income, mothers with many children, refugees, disabled persons, etc.

4.2 Description of the Work of the Unified State Social Security Information System Website

To automate this process in 2018, the Unified State Social Security Information System (EGISSO) [18] portal was created. It is a Federal state information system that allows the citizens and authorities to receive up-to-date information on social support measures provided by the budgets of all levels ranging from an individual to the entire country, as well as to obtain information necessary for the authorities to plan social support measures.

Before different social services could classify the same service differently. People having several social statuses at the same time, for example a pensioner and a person with disabilities, could receive similar services in different institutions or benefits from different budgets. The decision to integrate and universalize personal data of the clients of the social services changes the very methodology of social service delivery. In addition to the criterion of targeted assistance addressed to a certain population category, the emphasis is put on the criterion of need: a unified system of registration of citizens allows both the client and all social services providers to see what type of assistance from the government can a particular person claim according to their level of income.

The system consolidates information contained in the federal, regional and municipal registers and other state information systems in the field of social protection and social services provision. Thus, all the information on pensions, social benefits, compensations, subsidies, benefits, social services, social assistance is accumulated in one database, regardless of the basis of provision.

The authorities, local governments and their subordinate organizations delivering social protection services are obliged to transmit information to the EGISSO. Analysts and authorities have a different level of access to the site and, on the basis of a deeper study of information, can plan budget expenditures and make predictive models. The personal data of the clients of the social services is protected, the transferred information is transparent, and the statistics of the services provision can be checked on the website.

Thus, according to the statistics from the EGISSO website, 32% of the population of the Russian Federation (over 47 million) are recipients of social services. These are mainly people with the status of a pensioner/a veteran and a disabled person. One person can receive social support for several reasons at once. According to web analytics (Analysis), during the year, the site was visited by 1.3 million unique users, which is about 3% of all social service recipients in the country.

The EGISSO website has been working not for long and, most likely, its services will still be improved. However, as noted above, certain kinds of social support cannot be obtained without the personal presence of the recipients. In addition, there exists groups of people who barely or do not use digital technologies e.g. elderly people. According to the Research University Higher School of Economics, in 2017, 55–59 years old citizens used electronic services in less than 50% of cases in comparison to the total number of state and municipal service uses by this age group. At the same time, this percentage for 60–72 years old citizens equals 30% while 20–44 years old users mainly access social services in the electronic form (more than 70% of cases of receiving public services) [12].

4.3 Analysis of the Experience of Public Social Service Institutions in St. Petersburg

In St. Petersburg, social workers indicate that the primary source of information on available services is a personal consultation. A common communication channel is a phone, mostly mobile, but also landlines. All 101 public organizations that provide social services in St. Petersburg have websites that contain up-to-date information on available services and activities (news, list of departments, types of assistance, etc.) [19].

In addition, specialists provide information to customers through social networks. In Russia, the most popular social network is Vkontakte (vk.com), where many social service institutions have created their own group pages. In these pages, news and current events of institutions are announced; there exists an opportunity to participate in discussions. However, this channel is hardly suitable for clarifying personal issues. Moreover, these groups in social networks are mainly supported by specialists from quality control or public relations departments, but not by client service specialists.

We have selected 20 websites of state budgetary social service institutions and rehabilitation centers for disabled persons and children in order to evaluate their convenience for clients. All but one site had a customer feedback form, and 6 of them had a forum or message board where frequently asked questions were posted. According to specialists working in state social institutions, customers rarely use feedback forms of communication; more often they clarify questions personally or by phone.

Site attendance varies, and most likely does not correlate with the number of people who were provided services. In 2018, the number of unique visitors to the websites of social service centers for the disabled and disabled children was greater than the total number of service recipients (see Table 2). In urban social service centers, the situation is different, perhaps due to the fact that a significant number of clients are people of retirement age.

All social services also had profiles on the social network Vkontakte, 15 of them had a message board thread where people could ask questions, all groups were active with the last message was posted today, yesterday, or a week ago. However, the number of participants in these groups is not high in comparison to the number of people who were provided services during the year (see Table 2, data is presented only for social services having the appropriate statistics). Informing the public about the work of social institutions should be as comprehensive as possible, and in this regard the role of social networks should not be neglected. However, a low number of social network users suggest that the online interaction tool is less important than personal consultations.

Table 2. The ratio of the number of recipients of social services and members of groups in the corresponding social networks

Site	Type of institute	The number of unique site visitors in 2018	The number of service recipients in 2018	The number of subscribers in social network Vkontakte
Kcson-kolp.ru	Comprehensive social services	9972	10428	405
csonkr.ru	Comprehensive social services	8736	15512	283
Kcson-kurort.ru	Comprehensive social services	5400	8025	70
Kcson-mosk.ru	Comprehensive social services	8402	16523	236
kcsonnev.spb.ru	Comprehensive social services	11532	17674	56
csr-spb.ru	Social services for persons with disabilities	18672	1486	543

(*continued*)

Table 2. (*continued*)

Site	Type of institute	The number of unique site visitors in 2018	The number of service recipients in 2018	The number of subscribers in social network Vkontakte
krcsr532.ru	Social services for persons with disabilities	17844	1709	249
reabcentr.spb.ru	Social services for persons with disabilities	33264	1563	460
csridiprim.ru	Social services for persons with disabilities	18552	228	540
Stbyst.spb.ru	Social services for persons with disabilities	16524	1983	358

Some institutions also provide consultations to clients through Skype (for example, counseling disabled persons and children with disabilities, who are often limited in mobility). However, according to the specialists, such consultations are not popular among their clients.

According to the law, starting from 2018, state social service providers should transmit information on the designated assistance measures to EGISSO. In addition, a classifier of social protection support was developed in order to unify the definitions of services provided by different departments. The system of control and accounting of social services in St. Petersburg in recent years has become centralized. This allows employees of the Committee for Social Protection of the Population and other bodies responsible for monitoring the operation of city social services to receive information about each client and social services received, as well as to see statistics on the provision of services in the entire city.

However, there are still problems experienced with the automatic consolidation of data about services and their consumers. From interviews with employees of social services we learned that the encountered problems include the following:

– technical issues related to maintaining the unified multiservice telecommunications network of the executive authorities of St. Petersburg;
– input errors;
– insufficient technical equipment when there are fewer computers than needed for effective work, specialists are forced to keep reports on paper, because they cannot timely enter the data; obsolete equipment, databases are loading slowly or may freeze;
– the database architecture is constantly being revised in order to simplify work, which leads to technical interruptions in work and the need to retrain personnel;
– some employees have insufficient computer skills.

Generally, these factors of human and organizational origin reduce overall work speed and produce distortion of information about the services provided.

5 Conclusions

The Russian Federation is entering a new level of providing public services via the Internet. E-governance is actively developing, as well as a portal for interaction between citizens and the state referred to as GOSUSLUGI. It should be noted that most popular public services are now paying taxes and fines. In this case, digitalization simplifies the process of economic interaction between the citizen and the state. As for social support, the digitalization of services helps reduce the institutions of mediation and the time expenses, and at the same time increases the awareness of citizens about their rights, but this process is still far from full automation.

Russia is characterized by a vertical management system whereas the concept of digital economy requires a more flexible approach to public administration. The process of digitization of social services leads to the unification of social services. Thus a unified classifier of state social support measures and a database of services rendered have been developed, which should be updated in accordance to this classification by specialists working in the field. On the one hand, aggregation of statistics on social services increases the efficiency of planning federal and regional budgets, and on the other hand, a universal approach to the provision of services makes a person in a difficult life situation a “consumer of services” according to a certain standard. That is, there is some dehumanization of the help system itself.

In addition, new inequalities appear, such as the inequality of access to information (older age cohorts do not actively use digital services). This leads to an increase in the burden on social services professionals, who must take into account the interests of all clients, keep online information, and at the same time, use familiar channels of communication for people excluded from the digital field.

There exists a problem of effective personal information protection, since all personal data of citizens (their income, their documents, the services they receive) are accumulated in the Unified Identification and Authentication System.

Actions are needed to increase the efficiency of the municipal public social services. These include leveling input errors caused by human factors. This can be achieved by increasing the automation of the process, when manual entry of services into the database is minimized. Timely staff training and computer literacy of specialists, as well as improved technical equipment of workplaces will also contribute to improving work efficiency. In addition, it will speed up the process of passage to purely electronic account for state services from the actual system in which the account is organized in the analog and digital forms at the same time.

Acknowledgements. The research was conducted in *Saint Petersburg University* with the support of the *Russian Science Foundation,* Project No *19-18-00246* “*Challenges of the transformation of Welfare State in Russia: institutional changes, social investment, digitalization of social services*”.

References

1. Alexeenko, O., Ilyin, I.: Digitization of the global world and role of the state in the digital economy. Inf. Soc. **2**, 25–28 (2018)
2. Borodkina, O.: International trends in Russian social work. Eur. J. Soc. Work **18**(4), 631–644 (2015). https://doi.org/10.1080/13691457.2015.1049588
3. Decree of the President of the Russian Federation of 7 May, 2018 № 204: On national goals and strategic objectives for the development of the Russian Federation until 2024. http://kremlin.ru/events/president/news/57425. Accessed 20 May 2019
4. Decree of the President of the Russian Federation of 09 May, 2017 № 203: On the strategy of the information society development in the Russian Federation for 2017–2030. http://kremlin.ru/acts/bank/41919. Accessed 20 May 2019
5. Resolution of the Government of the Russian Federation of 24 October, 2011 № 861: On the Federal state information systems providing electronic state and municipal services (functions). http://pravo.gov.ru/proxy/ips/?docbody=&nd=102151387. Accessed 20 May 2019
6. Federal law of the Russian Federation of 29 December, 2015 № 388: On amendments to certain legislative acts of the Russian Federation in terms of accounting for and improving the provision of social support measures based on the obligation to comply with the principle of targeting and the application of criteria of need. http://rg.ru/2015/12/31/kriterii-dok.html. Accessed 20 May 2019
7. Resolution of the Government of the Russian Federation of 15 April 2014, N 313: On the approval of the «Information society» state program of the Russian Federation. http://gov.garant.ru/SESSION/PILOT/main.htm. Accessed 20 May 2019
8. Order of the Government of the Russian Federation of 28 July, 2017, № 1632-p: On approval of the state program «Digitalization of the economy of the Russian Federation». http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf. Accessed 20 May 2019
9. Order of the Government of Moscow of 11 October, 2010 N 2215-РП: On the concept of providing residents of Moscow with telecommunication services for obtaining socially significant information by creation of conditions of equal access to the cable television and Internet resources
10. Tapscott, D.: The Digital Economy. Publisher: McGraw-Hill (1994). Personal site: http://dontapscott.com/books/the-digital-economy. Accessed 19 May 2020
11. Rostelecom: Global trends of digitalization monitoring. Report (2018). https://ict.moscow/static/monitoring_rt.pdf. Accessed 20 May 2019
12. Digital Economy Indicators in the Russian Federation 2018: Data Book. National Research University Higher School of Economics. HSE, Moscow (2018). https://www.hse.ru/primarydata/ice2018. Accessed 20 May 2019
13. ICT Development Index 2017: International Telecommunication Union. https://www.itu.int/net4/ITU-D/idi/2017/index.html. Accessed 20 May 2019
14. Global Cyber Security Index 2018: International Telecommunication Union. https://www.itu.int/en/ITU-D/Cybersecurity/Documents/draft-18-00706_Global-Cybersecurity-Index-EV5_print_2.pdf. Accessed 20 May 2019
15. Gearing e-government to support transformation towards sustainable and resilient societies: E-government survey, United Nations, New York (2018). https://publicadministration.un.org/en/Research/UN-e-Government-Surveys. Accessed 20 May 2019

16. Zemskova, I.A.: Changes in the quality of e-government services in view of digitalization of public administration. Vestn. Saratov Socio Econ. Inst. Plekhanov Russ. Univ. Econ. **3**(72), 23–28 (2018)
17. Unified Portal of Public Services. https://gosuslugi.ru. Accessed 20 May 2019
18. Integrated national information system of social services (EGISSO). http://egisso.ru. Accessed 20 May 2019
19. St. Petersburg State Institution: City Information and Settlement Center. https://iss.ktsz.spb.ru/ISS/esrn/index/servproviders.htm. Accessed 20 May 2019

Reinforcement Learning for Data Preparation with Active Reward Learning

Laure Berti-Equille[1,2(✉)]

[1] ESPACE-DEV/IRD, UMR 228, IRD/UM/UG/UR, Montpellier, France
laure.berti@ird.fr

[2] Aix Marseille Université, Université de Toulon, CNRS, LIS, DIAMS, Marseille, France

Abstract. Data cleaning and data preparation have been long-standing challenges in data science to avoid incorrect results, biases, and misleading conclusions obtained from "dirty" data. For a given dataset and data analytics task, a plethora of data preprocessing techniques and alternative data cleaning strategies are available, but they may lead to dramatically different outputs with unequal result quality performances. For adequate data preparation, the users generally do not know how to start with or which methods to use. Most current work can be classified into two categories: (1) they propose new data cleaning algorithms specific to certain types of data anomalies usually considered in isolation and without a "pipeline vision" of the entire data preprocessing strategy; (2) they develop automated machine learning approaches (AutoML) that can optimize the hyper-parameters of a considered ML model with a list of by-default preprocessing methods. We argue that more efforts should be devoted to proposing a principled and adaptive data preparation approach to help and learn from the user for selecting the optimal sequence of data preparation tasks to obtain the best quality performance of the final result. In this paper, we extend Learn2Clean, a method based on Q-Learning, a model-free reinforcement learning technique that selects, for a given dataset, a given ML model, and a preselected quality performance metric, the optimal sequence of tasks for preprocessing the data such that the quality metric is maximized. We will discuss some new results of Learn2Clean for semi-automating data preparation with "the human in the loop" using active reward learning and Q-learning.

Keywords: Reinforcement learning · Data preparation · Q-learning · Data preprocessing

1 Motivations

In a 2017 survey conducted by the data science community Kaggle, "dirty data" is the common answer for 49.4% members responding to the questionnaire, when asked about the biggest barriers faced in data science[1]. As mitigation, data

[1] https://www.kaggle.com/surveys/2017.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 121–132, 2019.
https://doi.org/10.1007/978-3-030-34770-3_10

preparation often accounts for about 80% of the work of data scientists, but it is perceived as the least enjoyable part of the data scientist's job. As scientists, they would rather be deriving new knowledge and insights. The paradox is that without principled and adequate data preparation, those new insights are suspect at best.

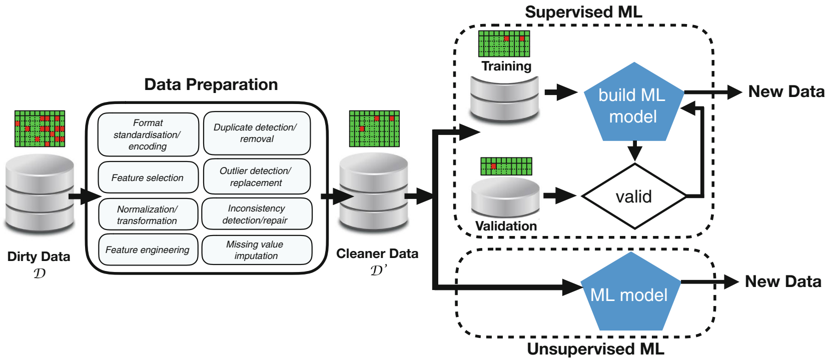


Fig. 1. A Generic ML workflow

Data preparation and data cleaning are known to be very challenging tasks that require detection and elimination of a variety of complex data quality problems, such as duplicate, inconsistent, missing, and outlying values. A wide range of methods for statistical analysis [13], constraint mining and checking [2], entity matching [5,9], and machine learning [7,11,16] are used nowadays for data quality checks, data cleaning, and data repairing [4]. However, most existing systems suffer from significant limitations.

First, data preprocessing includes many steps (e.g., normalization, transformation, encoding, discretization, imputation, deduplication, pattern or rule enforcement, feature selection, and engineering, etc.) but current systems generally provide support for a limited number of steps in the pipeline represented in Fig. 1.

Second, for a given data preprocessing step, current systems generally rely on a few by-default methods. The users have to try and test the methods to select the most appropriate ones in a tedious and time-consuming process. To extend the system with another method eventually more adequate, the users will have to write code or include other libraries, e.g., imputing missing values can be based on median, mean, most frequent values as default methods, but multiple imputations by chained equations (MICE) or K-nearest neighbors may be more accurate, but they are generally not included in the set of by-default methods provided by the systems.

Third, the systems do not recommend neither the most adequate preprocessing methods for a given dataset and a given ML task, nor the complete strategy as the full execution sequence of the methods: e.g., imputation of missing values can be done after or before deduplication, and these two different orderings may lead to a dramatically different preprocessed dataset. While some AutoML systems [3] can find the best model and configuration of hyper-parameters, the optimization has not been yet applied to the entire data preprocessing pipeline which remains fixed and based on by-default methods in a fixed (potentially nonoptimal) ordering.

Finally, the users may not know which preprocessing methods can be applied to optimize the final results downstream. This would require executing all possible methods for each task of preprocessing, as well as all the possible combinations of the methods with different orderings.

We argue that more efforts could be devoted to proposing a principled and adaptive data preparation solution to help the users in selecting the optimal sequence of data preprocessing tasks to obtain the best quality performance of the final result. As the first step in this direction, we have proposed Learn2Clean[2] in [1], a method based on Q-Learning, a model-free reinforcement learning technique that selects, for a given dataset, a given ML model, and a quality performance metric, the optimal sequence of tasks for preprocessing the data such that the quality metric of the final result is maximized. To integrate the "human in the loop", leverage existing knowledge and user's feedback during execution, we present a novel contribution enhancing Learn2Clean exploration strategy with active reward learning.

The paper is organized as follows. Section 2 presents a literature review and positions our contribution with respect to the literature. Section 3 presents Learn2Clean computational model based on Q-learning for data preprocessing and gives our notations. Section 4 discusses how the model is enhanced with active reward learning to integrate the "human-in-the-loop" and leverage the user's feedback. Finally, Sect. 5 provides some experimental results and insights in favor of our approach of adapting Q-learning with active reward learning for data preparation.

2 Related Work

Machine Learning-Based Data Cleaning. Several recent works that use machine learning to improve the efficiency and reliability of data cleaning and data repairing have been proposed lately [6,7,11,12,16]. In [16], Yakout et al. train ML models and evaluate the likelihood of recommended replacement value to fix erroneous or missing values. In [11], Rekatsinas et al. propose a framework for data repairing based on probabilistic inference to handle integrity constraints or external data sources seamlessly, with quantitative and statistical data repairing methods. Human input is often expensive and impractical to apply to entire large datasets. Machine learning can advantageously infer rules from a small set of examples cleaned by a human as shown in entity matching and deduplication by [6]. This approach can be coupled with active learning [12] to learn an accurate deduplication model with the fewest possible number of examples.

[2] https://github.com/LaureBerti/Learn2Clean.

Krishnan et al. in [7] propose a method for data cleaning optimization for user-defined ML-based analytics. Their approach selects an ensemble of methods (statistical and logic rules) for error detection and repair combinations using statistical boosting. However, to the best of our knowledge, our approach is the first one that leverages reinforcement learning for data cleaning and data preparation.

Reinforcement Learning. Reinforcement learning (RL) [14] has the potential to learn from interaction with the environment adaptively. It consists in discovering which actions produce the greatest reward by experience and estimating how good it is to take certain action in a given state, with the overall goal to maximize cumulated reward. Q-learning [15] is widely used because of its computational simplicity. In Q-learning, one does not require a model of transition functions and reward functions but learns directly from observed experience. This is an advantage as learning a model often requires exhaustive exploration that is not suitable for a large state space. Thus, Q-learning seems more suitable for exploring data preparation space. Furthermore, data preparation is a somewhat new application of reinforcement learning, and it raises new exciting challenges related to the dynamic update of the reward matrix at runtime. Typically, when the algorithm starts with a data preprocessing strategy, it picks and executes an action that will update the original dataset, and the set of the next valid actions and the corresponding reward matrix also change dynamically. Integrating the human in the loop in this context via active reward learning for automating data curation is again a novel contribution.

3 Learn2Clean Computational Model

Reinforcement learning methods can continuously learn from interaction with a dynamic environment [14]. Initially, with no prior knowledge of which action has to be taken in a given state, a conventional RL method takes an action in a starting state, receives a reward, moves to some next state, and repeats this procedure. It discovers which actions produce the greatest reward by trial-and-error experiences and estimates how good it is to take certain action in a given state to maximize the cumulated reward. Reinforcement learning methods are usually classified as:

- Model-based or model-free methods: model-based methods iteratively collect data samples randomly to learn the dynamics model, plan, recollect training data, and back-propagate the changes into a policy; model-free methods improve the value function directly from observed experience and do not rely on any predefined transition and reward models;
- On- or off-policy approaches: RL methods have to learn the optimal policy while behaving non-optimally, i.e., by exploring all actions. On-policy methods learn the best policy while using it to make decisions, whereas off-policy methods learn a policy different from what currently generates their behavior and actions.

3.1 Q-Learning for Data Preprocessing

Learn2Clean is based on Q-learning [15] which is a model-free, off-policy reinforcement method that learns the transition probability $T(s_{t+1}|(s_t, a))$ from the pair of current state s_t and action a to the next state s_{t+1}. It is a temporal-difference (TD) method that does not require prior knowledge about the environment's dynamics.

In our context of data preparation, the transition probabilities and the dynamics of the system are not given a priori because we do not know precisely which data preprocessing strategy would be the most adequate for a given dirty dataset and ML-based analytics tasks at hand. Furthermore, the space of possible states and actions can be very large if we consider all possible preprocessing methods with the range of their possible parameter configurations.

Table 1. Notations

Notation	Description
$M = \langle S, A, T, R, \gamma \rangle$	Markov Decision Process (MDP) with states S, actions A, transition function $T : S \times A \to Pr[S]$, reward function R, and discount factor γ
γ	The discount factor $\gamma \to [0, 1)$
π	A policy $\pi : S \to Pr[A]$ defined in Eq. (1) as the probability of selecting an action in a state
$Q^\pi(s, a)$	The Q-function represents the expected long-term reward of taking action a in state s, and following policy π for a given MDP as defined in Eq. (2) (also noted $Q(s, a)$ for simplification)
α	The learning rate $\alpha \to (0, 1]$ used in Eq. (2)
$\mathcal{S}$	Sequence of pairs (state, action), $\mathcal{S} = \langle (s_1, a_1), (s_2, a_2), \ldots, (s_p, a_p) \rangle$
$q(\mathcal{S})$	Quality performance metric for sequence of actions $\mathcal{S}$ for a given ML goal, e.g., accuracy for classification, silhouette for clustering, MSE for regression
$R^d(s, a)$	The deterministic reward function per (state, action) pair $R : S \times A \to [R_{min}, R_{max}]$
$R^u(\mathcal{S})$	The user-defined reward per sequence $R^u : (S \times A) \times \cdots \times (S \times A) \to [-\delta, +\delta]$
$R(\mathcal{S})$	The global reward defined in Eq. (3) also noted R for simplification
$\mathcal{D}$	The training dataset $\mathcal{D} = \{(\mathbf{q}_{1:n}, \mathbf{s}_{1:n}, \mathbf{R}_{1:n})\}$ where the quality metric value q_i for a sequence of actions s_i leads to a certain reward value R_i
$Pr(R\|\mathbf{g}, \mathcal{D})$	The probabilistic reward model given the goal $\mathbf{g}$ as the quality metric for a given ML model and the training data $\mathcal{D}$ as a sequence of data preparation tasks

Moreover, since we may not have generic solutions from previous datasets and tasks to learn how to clean, we believe that model-based reinforcement learning methods are not adapted.

Learn2Clean learns through experience in an unsupervised way. It explores from state to state until it reaches the goal (i.e., to maximize the user-defined quality metric).

Each exploration is equivalent to one training session in which the system explores the data preparation graph possibilities, receives the reward (if any) until it reaches the goal state. The purpose of the training is to enhance the decision of the system to get more training results. In other words, it has the choice between exploiting its knowledge by choosing the action with the highest estimated reward value and exploring its environment by taking any other action. Learn2Clean determines the next action (as the next preprocessing method to execute) without waiting until the end of the episode (i.e., the entire cleaning pipeline). This makes the method agile, computationally simple, and very performant compared to other data preparation approaches.

In [1], we propose a Q-learning reinforcement learning algorithm that can learn optimal actions for data preprocessing using heuristics for exploration of the action space. For exploration in RL, we apply the policy of Softmax action selection: the actions are weighted according to their respective Q-value estimates and sampled from the resulting distribution in the form of the Boltzmann distribution such as:

$$\pi = P(a \mid s) = \frac{e^{Q(s,a)/k}}{\sum_j e^{Q(s,a_j)/k}} \tag{1}$$

where k is a positive parameter such that a higher k value results in a more uniform distribution while a lower k value results in greedy action selection. Our notations are given in Table 1. The estimated value of taking action a in state s with policy π, denoted $Q^\pi(s,a)$ (or $Q(s,a)$ for simplification) is updated using reward $R^d(s,a)$, and the observed next state s' as:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\big(R^d(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)\big) \tag{2}$$

with α is a positive fraction such that $0 < \alpha \leq 1$, the step-size parameter that influences the rate of learning. When $\alpha = 1$, the system considers only the most recent information for learning. If α is properly reduced over time, the function converges [14]. The discount rate γ $(0 \leq \gamma < 1)$ determines the present value of future rewards. If $\gamma = 0$, the method is only concerned with the immediate reward. The action influences only the current reward. If γ approaches 1, the method considers future rewards more strongly. To maximize total reward, the system must select the action with the highest value (exploitation), but to discover such action, it has to try actions not selected before (exploration).

4 Exploration with Human-in-the-Loop via Active Reward Learning

To enable the exploration phase to experience other actions not taken before and increase the greater total reward in the long run in discovering better actions and also actions suggested by the human expert, we propose a new probabilistic reward model $Pr(R|\mathbf{g}, \mathcal{D})$, instead of a deterministic reward function $R^d(s, a)$ in order to modify Eq. (2). This allows us to integrate the human analyst in the loop and leverage the information about the certainty of the estimate to control the amount of human interaction required, as we only need to interact with the human expert if our model of the reward is not certain enough. The new architecture is presented in Fig. 2.

Reward Design. Both intermediate rewards and final rewards are used to guide the behavior of the system. Given a sequence $\mathcal{S}$, we define the final rewards as a sum over sequence-specific rewards and the user-defined rewards collected on-line for $\mathcal{S}$.

- The sequence-specific deterministic rewards R^d consist of the sum of deterministic scores for the sequence where invalid consecutive actions are penalized, e.g., normalization before imputation is penalized by -1, whereas normalization after imputation is valid with a score augmentation of $+1$. Valid sequences of actions are defined initially (as valid moves) in the initial reward matrix.
- The intermediate rewards R^u include step-wise validity rewards learned from the user. A small positive reward is assigned if the sequence of tasks does not violate the user's suggestions. Otherwise, a small negative reward is assigned.

The final reward score for a given sequence $\mathcal{S}$ is

$$R(\mathcal{S}) = \sum_{\forall(s,a)\in\mathcal{S}} R^d(s, a) + R^u(\mathcal{S}) \tag{3}$$

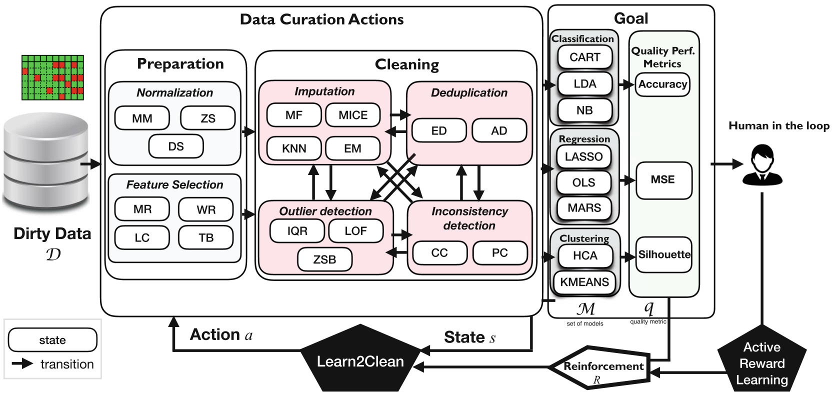


Fig. 2. Learn2Clean extension with active reward learning

Active Reward Learning. Our goal is to find a model $Pr(R|\mathbf{q}, \mathcal{D})$ that predicts the global reward given an observed quality metric outcome $\mathbf{q}$ and training data $\mathcal{D}$, which is partly obtained from the human expert annotating a set of sequences of actions to reach a certain ML goal in terms of quality performance. When modeling the reward, we have to take into account that the expert can give noisy samples of his implicit reward function and we also have to model the observation noise. Thus, we need to solve the regression problem:

$$R(\mathbf{q}) = f(\mathbf{q}) + \eta, \eta \sim \mathcal{N}(0, \beta), \tag{4}$$

where we assume zero mean Gaussian noise. Such a regression problem can, for example, be solved with Gaussian Process (GP) regression: $f(\mathbf{q}) \sim GP(m(\mathbf{q}), Cov(\mathbf{q}, \mathbf{q}'))$, where $m(\mathbf{q})$ is the mean function and $Cov(\mathbf{q}, \mathbf{q}')$ is the covariance function of the GP. Similarly as in [8], we use the standard squared exponential covariance function

$$Cov(\mathbf{q}, \mathbf{q}') = \theta_0^2 \exp\Big(-\frac{||\mathbf{q} - \mathbf{q}'||^2}{2\theta_1^2}\Big). \tag{5}$$

The set of hyper parameters $\boldsymbol{\theta} = \{\theta_0, \theta_1, \beta\}$ is found through Bayesian optimization [10]. We collect observations from the human analyst in previous iterations and add them to the training set $\mathcal{D} = \{(\mathbf{q}_{1:n}, \mathbf{s}_{1:n}, \mathbf{R}_{1:n})\}$ where the quality metric value q_i for a sequence of actions s_i leads to a certain reward value R_i[3]. We use Bayesian optimization to decide what reward R^{new} should be considered next. By the properties of GPs, $\mathbf{R}_{1:n}$ and R^{new} are jointly Gaussian and using the Sherman-Morrison-Woodbury formula (see [10] for details), the predictive posterior reward $Pr(R^{new}|\mathbf{q}, \mathcal{D})$ of a new goal $\mathbf{q}^{new}$ is then given by a Gaussian with mean $\mu(\mathbf{q}^{new})$ and variance $\sigma^2(\mathbf{q}^{new})$ as

$$Pr(R^{new}|\mathbf{q}, \mathcal{D}) \sim \mathcal{N}\Big(\mu(\mathbf{q}^{new}), \sigma^2(\mathbf{q}^{new})\Big). \tag{6}$$

To optimize the model of the reward function, we adapt the Expected Improvement (EI) from [8] that balances the greedy optimization with an exploration-exploitation trade-off parameter ξ such as:

$$EI(\mathbf{q}) = \Phi\left(\frac{\mu(\mathbf{q}) - \mu^* - \xi}{\sigma^2(\mathbf{q})}\right) \text{ with } \mu^* = \max_{i=1..B} \mu(q_i), \tag{7}$$

the best sample in the training set $\mathcal{D}$ over the budget B of queries submitted to the user and $\Phi(.)$ is the normal cumulative density function. The exploration-exploitation trade-off parameter ξ is chosen manually as it determines the amount of exploration during optimization and higher ξ values lead to more exploration. A suggested value by [8] is $\xi = 0.01$.

[3] We use subscripts to denote the sequences of actions annotated by the human analyst, i.e., $\mathbf{q}_{1:n} = \{q(\mathcal{S}_1), \ldots, q(\mathcal{S}_n)\})$ with related reward $\mathbf{R}_{1:n} = \{R(\mathcal{S}_1), \ldots, R(\mathcal{S}_n)\})$.

Algorithm 1. Bayesian Optimization for Active Reward Leaning in Learn2Clean

1. `INPUT:`
2. B budget of queries to the human expert
3. Top-k list of pairs (quality metric, sequence of actions) ordered by quality metrics
   $\mathcal{L} = [(q_1, s_1), \ldots, (q_k, s_k)]$
4. `FOR` $i = 1$ to B (with $B \leq k$)
5. `REPEAT`
6. Find the next sequence s_i in $\mathcal{L}$ by optimizing the expected improvement of quality EI over the GP: $s_i = \arg\max_s \Big(EI(q_s \mid \mathcal{D}_{1:i-1}) \Big)$
7. Obtain a possibly noisy reward sample $R(q_i) = f(q_i) + \eta$ from the objective function f
8. Add the sample to previous samples $\mathcal{D}_{1:i} = \{\mathcal{D}_{1:i-1}, (q_i, s_i, R_i)\}$ and update the GP.

5 Experiments

In these experiments, we study the impact of the parameters such as the step-size parameter α, the discount rate γ, and the training budget B, for a fixed exploration-exploitation trade-off parameter $\xi = 0.01$ on the global reward for selecting the data preparation strategy before a considered ML task. The code and datasets are available at: https://github.com/LaureBerti/Learn2Clean.

We evaluate how these parameters influence the learning performance of Q-learning in both cases: with versus without active reward learning. We used the House Price dataset from Kaggle[4] with 81 variables and 1.46k observations for clustering, regression, and classification. We ran Learn2Clean and obtained the results presented in Table 2 for various methods, namely: LASSO, OLS, and MARS for regression, KMEANS and HCA for clustering, and Naive Bayes and CART for classification. The quality metric, execution time (in seconds), and data preparation strategy are reported in the table. Data preparation strategies include various tasks such as K-nearest neighbors (KNN) or MICE imputation, outlier detection using InterQuartile Range (IQR), Z-score-based (ZSB), or LOF method, exact duplicate removal (ED), MinMax normalization (MM), and linearly correlated (LC) feature selection as illustrated in Fig. 2 (see [1] for more details about the preprocessing methods).

We compare regular Q-learning and active reward learning with the average number of steps. The average steps are calculated at every 100 episodes by dividing the total steps from the first episode to the last by the total number of episodes. Figure 3(a) shows the average number of steps with different step-size parameter α values for regular Q-learning. We compare $\alpha = .25, .5, .75, 1$ with a fixed $\gamma = .5$ and $B = 0$. As the α is smaller, the average number of steps also decreases. Lower step-size values perform better. This indicates that

[4] https://www.kaggle.com/datasets.

Table 2. Learn2Clean results on House Price dataset with regular Q-learning

ML task	ML model	Pred. Var.	Sequence of actions for data preparation	Metric name	Metric value	Time (s)
Reg.	MARS	SalePrice	$IQR \rightarrow MICE \rightarrow MARS$	MSE	0.231	134.5
	LASSO	SalePrice	$KNN \rightarrow ZSB \rightarrow LASSO$	MSE	1.44E−12	24.2
	OLS	SalePrice	$KNN \rightarrow ZSB \rightarrow OLS$	MSE	2.20E−25	45.1
Classif.	NB	SaleCondition	$LC \rightarrow MF \rightarrow NB$	Accuracy	0.518	20.1
	CART	SaleCondition	$MM \rightarrow AD \rightarrow CART$	Accuracy	0.4489	58.7
Clust.	HCA	//	$MM \rightarrow ED \rightarrow HCA$	Silhouette	0.844	32.2
	KMEANS	//	$MM \rightarrow MR \rightarrow LOF \rightarrow ED \rightarrow KMEANS$	Silhouette	0.679	29.3

accumulated experience affects value estimation more significantly than recent experience.

For the discount rate γ experiment, we fixed $\alpha = 0.25$ and vary $\gamma = .25, .5, .75$ for $B = 0$ (regular Q-learning). The average number of steps with different γ values are shown in Fig. 3(b). The lowest $\gamma = .25$ performs better. This means that immediate rewards are more important than future rewards with regular Q-learning. As episodes continue, a higher $\gamma = .75$ is slightly better. Relatively more steps are needed to achieve the same goal. In this case, future rewards are more significant than current rewards.

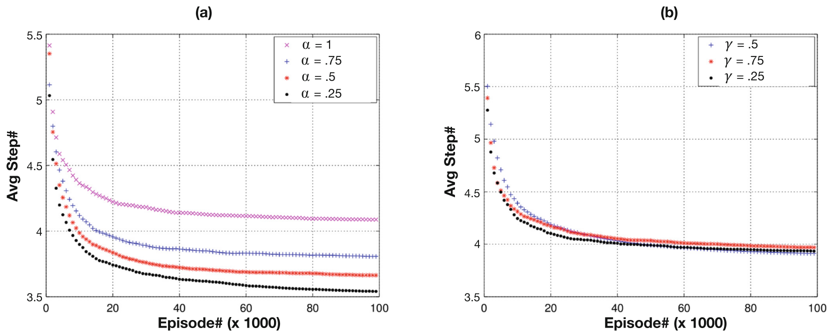


Fig. 3. Regular Q-learning: average number of steps with varying α with fixed $\gamma = .5$ in (a) and varying γ with fixed in $\alpha = .25$ (b) for $B = 0$ and $\xi = .01$.

We then compare regular Q-learning (with no human in the loop) with active reward learning involving the user. For experiments, we select the parameter values shown previously. The step-size parameter α is set to 0.25 because low learning rate seems to be the most appropriate to our problem. The discount rate γ is set to .25 as well.

Figure 4(a) shows the average number of episodes with different B values from 1 to 20, given $\alpha = .25$ and $\gamma = 0.25$. The average number of episodes first increases dramatically with B and then decreases after $B = 16$. In the beginning, exploration is relatively inexpensive but turns out to be less effective

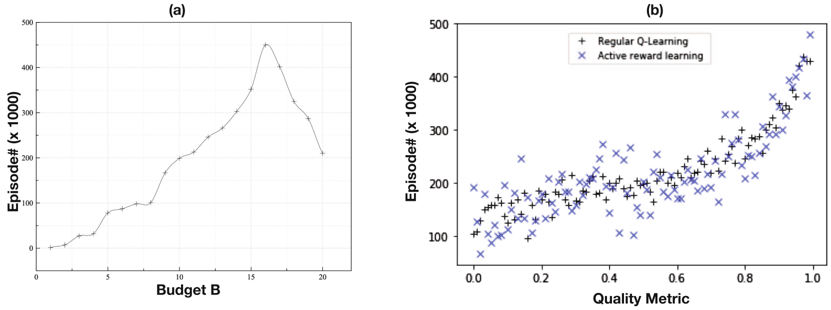


Fig. 4. Active reward learning: average number of episodes with varying B from 1 to 20 with fixed $\alpha = .25$, $\gamma = .25$, and $\xi = .01$ in (a) and quality performance in (b).

when the user tends to systematically disagree with the strategy recommended by the system via Q-learning. The more disagreements (as negative samples in the online training), the more steps are needed to converge and reach for a sub-optimal solution yet maximizing the quality performances. Eventually, when sufficient consensual knowledge is accumulated, exploitation is worthy. As immediate future work, we plan to investigate further the impact of adversarial user inputs on the Q-learning strategy to mitigate the problem of noisy inputs.

6 Conclusions

We explored the use of reinforcement learning for data preparation with integrating human in the loop via active reward learning. The evaluation results provide empirical support for the effectiveness of our approach for the task at hand. We are planning more extensive studies with datasets from various application domains to experience a wider range of data preprocessing scenarios. We are also currently extending this work to consider a more dynamic environment to scale the approach to larger and more complex spatio-temporal datasets with more complex data preparation strategies.

References

1. Berti-Équille, L.: Learn2Clean: optimizing the sequence of tasks for web data preparation. In: Proceedings of the The Web Conference 2019 (2019)
2. Fan, W.: Data quality: from theory to practice. SIGMOD Rec. **44**(3), 7–18 (2015)
3. Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F.: Efficient and robust automated machine learning. Adv. Neural Inf. Process. Syst. **28**, 2962–2970 (2015)
4. Ilyas, I.F., Chu, X.: Trends in cleaning relational data: consistency and deduplication. Found. Trends Databases **5**(4), 281–393 (2015)
5. Konda, P., et al.: Magellan: toward building entity matching management systems. PVLDB **9**(12), 1197–1208 (2016)

6. Konda, P., et al.: Magellan: toward building entity matching management systems. Proc. VLDB Endow. **9**(12), 1197–1208 (2016)
7. Krishnan, S., Franklin, M.J., Goldberg, K., Wu, E.: Boostclean: automated error detection and repair for machine learning. CoRR abs/1711.01299 (2017)
8. Matthew Hoffman, E.B., de Freitas, N.: Portfolio allocation for Bayesian optimization. In: 27th Conference on Uncertainty in Artificial Intelligence (UAI2011) (2011)
9. Navarro, G.: Approximate string matching. In: Encyclopedia of Algorithms, pp. 102–106 (2016)
10. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, Cambridge (2005)
11. Rekatsinas, T., Chu, X., Ilyas, I.F., Ré, C.: HoloClean: holistic data repairs with probabilistic inference. PVLDB **10**(11), 1190–1201 (2017)
12. Sarawagi, S., Bhamidipaty, A.: Interactive deduplication using active learning. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2002, pp. 269–278. ACM, New York (2002)
13. Schafer, J.: Analysis of Incomplete Multivariate Data. Chapman & Hall (1997)
14. Sutton, R.S., Barto, A.G.: Introduction to Reinforcement Learning. MIT Press, Cambridge (1998)
15. Watkins, C.J.C.H., Dayan, P.: Technical note: Q-learning. Mach. Learn. **8**(3–4), 1190–1201 (1992)
16. Yakout, M., Berti-Équille, L., Elmagarmid, A.K.: Don't be scared: use scalable automatic repairing with maximal likelihood and bounded changes. In: Proceedings of the ACM SIGMOD, pp. 553–564 (2013)

Can Machines Learn Whether Machines Are Learning to Collude?

Jonathan Cave(✉)

University of Warwick, Coventry CV4 7AL, UK
j.a.k.cave@warwick.ac.uk

Abstract. Online economic interactions generate data that artificial intelligence (AI), machine learning (ML) and deep learning (DL) can use: business predictive analytics, process optimisation and market power; consumer search and choice; and government gathering evidence and regulating harmful behaviour. In algorithmic collusion (AC), revenue management algorithms implement tacitly collusive behaviour. This paper summarises theoretical and empirical evidence and considers how ML methods affect AC and whether regulators' algorithms can help. It examines links between Internet regulation and competition policy.

Early ML literature concerned programmes 'learning' their environments, e.g. predicting rivals' prices to maximise profit by estimating prices/costs, identifying strategies or influencing learning. Here, ML is AI that self-programs to optimise specific objectives (data and model 'layers') and DL is many-layered ML. Increased depth makes behaviour an intricate convolution of data and programme history; invisible to programmers and inexplicable to others. ML by many firms may fail to converge or have unintended consequences.

Many models use simple ML algorithms to demonstrate behaviour consistent with collusion. It is not classically collusive without communication. Populations of simple AI can learn reward/punishment strategies that sustain profitable outcomes. This paper considers further variations taking into account strategic variations, finite-memory or dominance elimination and the impact of product characteristics and search. Simulation illustrates classic inefficiencies (overshoot, convergence to supracompetitive prices, cycles and endogenous market-sharing).

It is not clear what regulators could or should ban; can they detect AC or limit its consequences? We consider: restricting information available to firms; constraining price dynamics; coding standards that incorporate regulatory compliance in ML objectives; and algorithmic detection of specified anticompetitive behaviours. For instance, likelihood-ratio policy gradient reinforcement learning algorithms are more *likely* to converge to collusive behaviours when they take other firms' learning into account and *able* to shape others' learning with suitable prevalence of AI and network topology.

Keywords: Algorithmic collusion · Machine learning · Antitrust

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 133–152, 2019.
https://doi.org/10.1007/978-3-030-34770-3_11

1 Introduction

Algorithmic control of firm behaviour has been widespread and a subject of intense policy and academic concern. Algorithms can be used to elicit, collect and analyse information on customers, and thus to facilitate behavioural profiling, personalisation and other activities. These systems influence relations:

- Between firms and individual customers;
- Among firms competing over platforms and operating in extended value chains; and
- Between market participants and regulators seeking stability, efficiency and fairness.

This paper concentrates on the potential for pricing algorithms to facilitate tacit collusion. It considers whether such coordination should be proscribed or discouraged and therefore whether regulators can identify, demonstrate and remedy harmful practices without inhibiting useful innovation. But this is not separable from the effect of AI[1] and related technologies on other relations. We should thus consider how these technologies change our understanding of how markets should and do function – what is special about AI that merely automates what humans already do? We briefly consider these issues before moving on to discuss learning by and about firms.

2 Firms and Their Customers

Automated collection and analysis of information about individual customers raises privacy concerns and fears that control of such information and associated models may inhibit customer search. Nowadays, such data are subject to controls on access, retention, processing and use in decision making. However, such rules concentrate on *identifying* individuals (skirting issues of e.g. differential and collective privacy) and the problem of data that may be accurate and subject to consent but are used in conjunction with other (not personal) data or in derivative form (e.g. credit refusals). Moreover, data may be of use to parties other than those involved in the original collection. For instance, consumer interests may be harmed and markets rendered inefficient if relevant data are insufficiently or excessively mobile. This was a prime motivation behind the introduction in Europe of Open Banking and initiatives like Mydata; it underpins proposals for Open Communication. Inappropriately strong limits on data access or inadequate data governance can foreclose valid consumer interest in value generated by their data (explicitly addressed by GDPR). This in turn can produce selection bias and outright distortion and can inhibit useful innovation. Conversely, ML can in principle facilitate close to first-degree price discrimination. Not all of these possibilities are equally feasible, or damaging to individuals, markets or the economy as a whole.

[1] Here, AI refers to the gamut from statistical decision tools and scoring rules (only a data layer) through two-layer models (e.g. reinforcement learning, which combines data with an explicit coevolving model plane) to DL with multiple planes and largely implicit relationships. The latter two partially separate algorithm training and use.

A second source of concern is that data used by firms may go beyond those knowingly revealed by consumers. In financial services, credit decisions are increasingly based not only on relevant data meeting a legitimate-interest test but also on ‘harvested’ data that appear correlated with e.g. default probabilities. This kind of profiling raises additional concerns that cannot be handled by the individually-focussed privacy regulation apparatus. Two examples will illustrate the kinds of concerns raised.

First, decisions made about one individual may depend on models estimated from data on others. Insurance underwriting is entirely based on such comparisons. In highly automated contexts individuals cannot know what data are used to assess applications and therefore cannot challenge decisions or learn from the outcomes. This is not limited to service providers; fintech platforms providing credit intermediation often collect information that they use to show users only offers they are likely to get. Information is spread further than necessary and many decision rules, machine learning approaches and trained algorithms may be involved. In ‘thin’ markets, individuals will be aware that their data will be used to update models and may distort their data accordingly.

A second example is so-called ‘algorithmic bias’. It is by now commonplace to acknowledge that data sets used to train automated decision systems may contain implicit bias; this applies as well to the code. In the former case, sampling individuals into the data set and proximities between individuals established by the code can embed selection bias, especially when historical data (collected at a time when a given system was not used to make decisions and thus to generate data) are used for training. In the latter case, cultural assumptions and personal idiosyncrasies may be built into the optimisations and adjustments tools (e.g. objective functions used for updating model in light of new data; the layered structure in DL systems; or the extent to which recommendations are implemented automatically or relayed to human advisors).

One solution is to mandate explication. But this may be of limited effectiveness; a standard test for algorithmic bias is to compare results using data where only one nominally-irrelevant or legally-protected field (e.g. gender, ethnicity, religion) is changed. If results are different, bias is present[2]. But this only finds bias linked to identifiable data fields. This raises the question of why certain characteristics are protected. Bias represents simplified decision-making. It may use correlations that are accurate but unconscionable (e.g. location discrimination, traffic shaping, gendered insurance premia).

We can test the validity, ethical implications and proportionality of overt and measurable bias. It can be challenged; by individuals, politically or socially.

But suppose instead that complex data are fed to a deep learning algorithm. There may be no single characteristic or stable set of characteristics leading to adverse decisions. Those discriminated against have no way of understanding or rationalising their failures (e.g. to get a job or obtain credit). They may internalise this; feeling that ‘the system’ is against them and disengaging or reacting in damaging ways. This, in turn magnifies and/or embeds such biases deep in the data and algorithms of the system.

2 This test has been failed by e.g. algorithms used to make probation and parole decisions and to generate job interview offers on LinkedIn.

For individuals, this may convert incentive to labels. A person who fails to manage their finances effectively might once have felt a strong incentive to improve; with widespread automated credit scoring they will not be offered credit and thus will lack financial incentives to improve. Groups may respond collectively (political action, boycott, or bypass). It is important that these self-healing responses continue to function.

3 Interactions Among Firms

Firms continually observe the results of their own and each other's behaviour; they thus learn and adapt. Markets rely on this learning; customers are encouraged to search and firms to compete; the end result is some combination of static and dynamic efficiency.

But firms may recognise that it is better to coordinate actions rather than to compete. Done explicitly, this collusion is generally illegal. Such agreements are not self-enforcing; they by definition involve firms failing to maximise individual short-run profit in favour of enhanced collective and/or long-run profit in ways that reduce aggregate welfare. Firms therefore have an interest in ensuring that active firms adhere to the agreed discipline and that supranormal profits do not attract entry. Thus they would profit by a mixture of predatory entry-deterrence and intensive collusion.

Competition rules aim to prevent predation and explicit collusion. This requires a basis for identifying and proving violations, assessing net harms and relevant intent and crafting and enforcing appropriate remedies. This in turn involves an identification problem; does observed behaviour fall into the permitted or the prohibited category? For example, predatory behaviour must be distinguished from competitive response to entry and price distributions should be linked to valid estimates of cost. This may involve quantitative and other evidence sources and may include AI solutions to identification problems. There are some 'smoking guns' – patterns of behaviour that are so strongly associated with violations that they provide grounds for proceeding when detected. They may be dynamic e.g. lowering price when entry or a rival cuts price and raising it again after the perturbation has finished. Others are behavioural e.g. explicit communication or evidence of collusive or predatory intent, conduct showing clear punishment/reward behaviour and evidence that machine-supported strategic choices respond to shocks by prioritising price over volume as opposed to volume over price.

The problem is much harder when market relations are complex, imperfectly observed and highly dynamic. For instance, restricting output while prices are climbing could be taken as evidence that the firms were trying to create artificial scarcity or return to collusion after a 'punishment phase' but could be a competitive response to an exogenous fluctuation in demand; if prices rise exogenously it is rational to delay sales. Most jurisdictions regard the costs, burdens and uncertainties of a ban on conscious parallelism and tacit collusion as outweighing the advantages.

This judgement may not be time-consistent; it in effect encourages tacit (as opposed to explicit) means of collusion, which could in turn change the quantitative reason for not modelling and prosecuting such tactics. Important elements are the difficulty of identifying intent, the potential ineffectiveness of legal measures, the presumed difficulty of reaching and sustaining profitable arrangements for the cartel and the vulnerability of such arrangements to predation, entry and exogenous shocks.

AI-specificity may arise if technologies that change what firms know about each other, when they know it and the extent to which they can identify defections by members and thus credibly threaten punishment change the case for regulation.

To fix ideas, suppose firms use algorithms to keep an eye on each other and to adjust strategy. Suppose further that their collective operation implements a collusive outcome. A problem arises if there is neither explicit communication between the firms nor a demonstrable intent to write collusive algorithms e.g. if algorithms are purchased from a separate market layer (coders). One could treat this as vertical foreclosure.

It may be a natural result of upstream competition combined with network externalities; firms tend to buy algorithms that yield higher profits and the adoption of an algorithm that tends to yield collusive outcomes when it interacts with itself will increase demand for that algorithm. The result will be domination of the code market by profitably interactive algorithms (which learn how to collude with themselves or each other) and tacitly collusive outcomes downstream, without intent or overt communication.

In the absence of intent and control, such algorithms could also learn to implement (or converge to) more competitive behaviour. Certain features of their dynamic behaviour (e.g. punishment-and-reward behaviour) are associated with collusion and may be reinforced by the learning process itself. Conversely, our modelling suggests that 'signatures' of dynamic responses, tests of the propensity of such algorithms to collude or 'tipping' of a networked and algorithmic markets into collusive phases could be learned and used by regulators. Modifications of the code (e.g. adding to the updating and/or value functions terms that reward compliance with regulations or increases in consumer surplus) can lead firms to converge on less damaging behaviours. Finally, the networks through which firms observe each other and their overlapping patterns of contact are significant in two ways; convergence to collusion is faster and more effective in some networks[3] and 'seeding' key positions in sufficiently centralised networks with 'virtuous' algorithms can encourage convergence to less-collusive outcomes.

4 Firms and Regulators

What forms of algorithmic firm behaviour are legal and what forms should be?

4.1 Regulatory Modelling

How can violations be detected? Approaches that might be tested include the following.

[3] Specifically, if algorithm choice is viewed as a coordination game, certain (slow-growth) network structures show much faster convergence than grid-like or fully-connected networks.

1. Modelling the price distributions expected to emerge from different approaches. For instance, simple Q-learning algorithms in a pricing game can lead to imperfect coordination in simulations. Either firms learn to coordinate imperfectly (settling on less than fully collusive prices or experimenting too often and long) or converge to either collusive or competitive outcomes. These lead to different price distributions that could be learned by regtech algorithms and used to devise market tests.
2. Event studies. AI systems that continue to learn will respond to shocks in ways that can reveal e.g. the relationships between pricing and underlying costs or the implicit objectives used to drive the learning process.
3. Direct analytics and machine learning to solve identification problem. In favourable market conditions, a regulator could learn the collective behaviour of firms in order to evaluate competition implications. Of course, this one-sided learning depends on firms not seeking to learn about or to influence the regulator's models.
4. Block graph (Bayesian inference) approaches. If we regard even collusive market outcomes as having multiple causal pathways and if recognise that firms' learning may have multiple outcomes, the identification problem of determining *why* a market generated a particular set of data may be amenable to network inference methods.
5. A:B testing of trained and untrained algorithms. If firms are required to share their algorithms with the regulator, they can be tested to see whether (and under what conditions) they generate acceptable outcomes. This could be used in an *ex ante* regulatory approach with guidelines as opposed to *ex post* approaches.

To determine whether a given outcome is acceptable, the regulator will also need to learn firms' costs, market demand and – perhaps most importantly – what firms themselves know or believe about costs and demand.

A related issue is that results may not be deterministic. A given population of algorithms may converge to a collusive outcome following one realisation of consumer behaviour and a more competitive one under other circumstances. This is not new; it is one reason why so much attention is paid to entry as well as market conduct and to the composition of demand as mentioned above. In this setting we may have to reassess the degree of certainty needed to 'prove' violation or justify intervention.

4.2 Remedies or Disincentives

In deciding what should be regulated we should consider what remedies (or disincentives) can be implemented. Examples include the following.

1. Ex ante controls – some algorithms could be approved for specific applications. Use of approved algorithms could be linked to adoption of 'regtech' systems to provide real-time monitoring. This could be combined across systems to detect and assess emergent behaviours and systemic departures from effective competition, and limit their form, use and linkages (in the spirit of macroprudential regulation).
2. Controlling information sharing – the results of simulation and modelling depend strongly on who observed what information, so another control would change that pattern. This could break up clusters by maintaining uniform public data

repositories or placing structural limits on the complexity or breadth of data that can be shared or the speed with which they are reported or prices be adjusted.

3. Ethical guidelines – another tool in the regulator's kit is promulgating and enforcing guidelines ranging from explicit enforceable standards to more general statutory Codes of Conduct. Such initiatives are under active development in a range of UK agencies e.g. Digital Catapult (for start-ups receiving government support), NHS-X (health-related app development and commissioning), West Midlands Police (National Data Analytics Solution), financial services and contractors processing personal data for public service delivery. Associated to this are aids to regulatory innovation such as sandboxes and default "comply or explain" enforcement.
4. Explicability requirements – requiring algorithmic processing and decisions to be accompanied by a suitable explanation.
5. Coding practices and requirements for audit and/or proof of performance.
6. Direct market interventions – specific transactions or data signals to nudge the system away from inefficient behaviour or to produce relevant evidence[4]

4.3 Regulatory Impact Assessment

Any move to regulate algorithmic market behaviour should be assessed against better regulation principles. These include: an 'evaluate first' requirement; considering the accuracy and completeness of evidence behind the *status quo ante* assessment and the counterfactual against which the proposal will be measured; explicit statement of regulatory principles and objectives (including a framework for assessing direct and indirect impacts on static and dynamic efficiency, competition and innovation); modelling the interaction between these systems and other regulated firm behaviour; analysing the proportionality of requirements and potentially disproportionate impacts on e.g. SMEs; and likely levels and patterns of compliance.

Among firms, this analysis can involve a form of conjectural equilibrium. Firms that compete in prices will develop strategies that are response functions to past prices. Eventually, they may learn others' response functions or even reveal them directly, so strategies are responses to response functions, and so on. For machine learning in pricing algorithms, firms may begin by using reinforcement learning to explore demand and cost curves (regarded as exogenous and time-invariant and therefore capable of being learned in this way). Multiple firms can use these strategies, treating rivals as part of the environment. If they recognise other firms' learning, the problem changes - the 'environment' is no longer stationary and moreover responds to the firm's actions. In this environment firms should factor in how today's behaviour influences others in the future, combining learning with perturbing other firms to learn what they know and signalling. Regulators using rule-driven or learning strategies become part of the environment about which firms learn. This complexity argues for an experimental approach, and for impact assessments with explicit provision for monitoring and periodic review.

[4] Analogous to weakening uniform-price buyer collusion by artificial demands from the seller.

4.4 The Basic Set-Up

Firms

Firms are essentially playing a market game. This can be played on many levels. In each level, automated decision-making and machine learning can be used to automate and improve strategic interaction. AI serves as a faster, more consistent means of doing what firms would ordinarily do. Sometimes, quantitative speed and scale advantages lead to qualitatively different outcomes. For example, a cartel that is unsustainable because firms cannot detect and respond to defections with sufficient accuracy, focus and speed may become feasible with the use of suitable algorithms and information[5]. As a result, markets may experience 'phase changes' in which models of opponents' behaviour and the firm's own action space become more sophisticated. Strategies used by firms may include price, quantity, conjectural variation, memory, response speed, revenue management algorithms and machine learning approach.

Firms can pursue various objectives e.g. short- or long-run profit, reputation, supply chain dominance or IP. These can be static or evolving hybrids of simple objectives (e.g. $\alpha Profit + (1 - \alpha) Consumer\, surplus$). To develop a theory of whether these transitions are of policy concern, it is appropriate to describe strategy levels (previous paragraph) and see how inefficiencies could be generated at other levels e.g. whether competition in algo-writing 'tips' into norms that sustain tacit collusion at market level. Such models will be sensitive to what firms observe and seek to learn. When rivals, customers and regulators are learning, the underlying identification problem is complicated by issues of complexity and emergent behaviours[6], endogenous shocks (e.g. from experimentation), propagating risks and non-stationary demand and cost (e.g. with experience or 'social' goods, or learning by doing).

The next step is to specify (or ascertain) whether AI is used to: set prices or quantities; classify, categorise or identify customer types; structure interactions (e.g. influence preferences or elicit further information); adjust product characteristics; design contracts and services; identify/negotiate transactions; make suggestions to sales reps, etc. It is also useful to describe AI 'depth' (e.g. rule-driven, independent learning, interactive learning without accounting for others' learning, aware interacting learning, etc.) The method used to create and refine the learned model should also be specified; this can involve reinforcement/Q learning or some form of DL.

Regulators

At a simple level, regulators can use standard data science tools to detect (possibly high-dimensional) correlation among firms and potentially to match them to identified strategic patterns including potentially collusive learning. This is likely to be effective if costs and demand are stable for long enough to efficiently sample the distributions.

This analysis can be conducted on the assumption that firms behave in a simple fashion (e.g. Nash equilibrium or full collusion). However, firms may be tacitly colluding as mentioned above. The literature shows that simple ML algorithms can lead to supra-competitive prices without overt collusion, direct communication or collusive intent.

---

[5] For a currency exchange example, see e.g. Cave and Gelsomini [1].

[6] e.g. the – possibly simultaneous – deployment of AI systems at platform or intermediary level.

This is perhaps the most challenging part of the exercise; determining whether the patterns detected are evidence of an underlying collusive effect or intent and whether effective intervention is feasible and warranted. Note that effect or intent could justify intervention to address current inefficiency or prevent future collapse of competition.

But suspicious patterns may have innocent explanations. For instance, sudden price jumps followed by a slow return to a low price may be evidence of an Edgeworth Cycle[7]; firms usually price efficiently (near marginal cost) in a 'war of attrition' surviving on low margins. Eventually, one firm may raise prices (to milk customers who've ceased actively to search or cover a revenue shortfall). In response, the rest begin to raise prices, albeit by less because they need to retain existing clients and attract some from the first-mover. This is unstable, so a period of downward price competition follows. Without knowing true marginal costs, this may be indistinguishable from imperfectly collusive automated market behaviour.

Another example arises in a quantity competition setting if firms restrict production when prices rise and expand it when prices fall. Without knowing the depth of the market, we can't tell whether they are signalling willingness to collude, restoring discipline after a cartel punishment phase, bidding up prices by artificially restricting supply or treating demand as exogenous but persistent and delaying sales until price is higher.

Studying Firm Learning

If firms facing fixed (or stationary) demand sometimes learn to collude fully and sometimes don't, non-parametric analysis could be used to bound costs and thereby to identify supra-competitive prices and distinguish them from near-competitive ones. If, on the other hand, firms always learn partially to collude this identification problem is harder because prices will lie between Nash and Collusion, but with low dispersion.

5 Programme

This paper reports early progress on a programme to provide an analytic basis for integrating machine learning with real-world competition regulation. It proceeds in stages.

5.1 Learning About Prices and Quantities

The first stage models simple algorithms in games played on networks. The network programme starts with cliques (all firms observe and interact with each other), various static networks (stars, lines, small worlds, scale-free and slow-growth) and finally dynamic networks where firms endogenously choose other firms to monitor.

[7] See e.g. Maskin and Tirole [2].

5.2 Algorithm Dynamics

The next stage models algorithm adoption as a conventions game. Initially, there is a finite set of algorithms from which to choose; players can change algorithms (using the best according to current beliefs) or explore alternatives. The finite set includes algorithms that are best replies to themselves. They include collusive algorithms (which are payoff dominant) and imperfectly collusive or competitive algorithms. Thus, they converge to more or less collusive or competitive results when matched to themselves and produce inferior or non-convergent results when mismatched. The expected payoff to adopting an algorithm is found by considering the behaviour resulting from the algorithms chosen by the firm's rivals. The dynamics are reinforcement learning; firms choose currently-optimal algorithms with high probability and others with much lower probability. This leads to a Markov process over the algorithms adopted by all firms.

The process converges almost surely. In a clique, all firms converge to the same algorithm, which is not necessarily the optimal one. If the probability of choosing a suboptimal algorithm depends strongly on expected losses, the process converges to the risk dominant algorithm; other experimentation rules may have stable outcomes that are neither risk nor payoff dominant. Non-clique network structures may have stable outcomes where clusters have different algorithms and levels of collusion.

5.3 Reinforcement Learning/Q-Learning

Simulations conducted to date use a variant of the Q-learning approach to reinforcement learning. This is sketched below. It involves learning a finite approximation to the function that maps the firms' actions to outcomes in different states; an outcome involves a current return and a new state. Firms are not assumed to have prior knowledge. Q-learning involves discrete sets of states and actions; the firm's value from choosing an action is the current payoff plus the present value of future optimal play. Firms try an action and use the results to refine estimates of the value function. If the process visits all state/action combinations often enough and converges, firms will have learned all they need to behave optimally. If many firms (and also possibly customers and regulators) are learning and this is taken into account, the state space can grow very rapidly. What's worse, the impact of small changes in a firm's behaviour can also increase rapidly, requiring finer approximations and further increases in complexity. This can be artificially mitigated by imposing finite automaton learning or Markovian strategies.

Such structural limitations may themselves be collusive; a pre-commitment to finite memory or specific information (granularity) may increase the odds of a collusive outcome. This can be illustrated by two simple Prisoners' Dilemma examples (the game used in early discussions of mutually beneficial machine learning, which shares some features (inefficient dominant strategy one-shot equilibria) with oligopoly games).

First, suppose that firms play an undiscounted infinitely-repeated Prisoners' Dilemma with high-price (H) and price-cutting strategies (G). This has many equilibria (the one-shot dominant strategy one (G, G) being analogous to competitive behaviour). If the players commit themselves to one period of memory their strategy spaces collapse to a move at period 0 and moves in case the opponent chose H or chose G on the previous round - 8 pure strategies for each player. There are still multiple equilibria; the competitive one is now written GGG. Reinforcement learning in this game involves each player always using each strategy with positive probability, but sharply reducing weight on inferior ones, which corresponds to successive elimination of weakly dominated strategies. This version of the game proceeds until only the GGG strategy and collusive tit-for-tat (HHG) are left with significant likelihood. At this point, HHG defeats GGG and collusion emerges. Without finite memory, this could not be guaranteed.

A second example involves information asymmetry. Label the strategies A and B and suppose there are two states; state 1, where A is the collusive strategy and state 2 in which B is collusive. The first player knows the state; his move is observed by the other player. If he uses his information, the other player will learn this immediately; both players will choose the (dominant) competitive strategy. If the first player commits to ignoring his information, the other player may as well do the same and the players will partially collude – this is better for both than the outcome if they both learned and made optimal use of their information.

This opens up another collusion possibility. If the players train their algorithms on historical data and then use them with only minimal sampling to respond to changes in demand, the choice of a training data set can sustain collusion.

Overall, the outcome is determined by a combination of (training) data, architecture (observation, business model and market), algorithmic form and the (stochastic) environment. In many settings, initial variations in the algorithms chosen will be eliminated by their self- compatibility as described in the discussion of algorithm dynamics starting on page 10 and in line with the standard results on ‘tipping equilibria’ in network games[8]. Of course, there are differences between the algorithms (in the source code sense) and specific instances, given that they are trained on different data. But these differences should not be too great and will reflect salient details of the market structure[9]. Within this set of potentially- stable structures, firms could choose algorithms that deliver additional capabilities.

Our focus here is on collusion (maintaining internal discipline), but with suitable updating rules and training data they could also serve as predatory/punishment devices. In networked environments – where firms have different patterns of informational and market contact - they can be symmetric or aligned to distribute or share the costs of punishment to provide directed punishment and to reward members for carrying out punishments.

5.4 Q-Learning Basic Results

We start from a model in which only one firm learns. The same analysis can then be used to describe situations in which many firms learn, but where they ignore the learning behaviour of others (or model it as a stationary stochastic process for choosing their strategies).

[8] See e.g. Katz and Shapiro [3] and, for a discussion relevant to the outsourcing of pricing analytics, Galeotti and Goyal [4].

[9] See the discussion of learning in cliques (page 16) and networks (page 17).

When firms are aware of others' learning, they will take account of the fact that their choices combine current payoff, learning about the underlying market process associated to a given choice of action while simultaneously changing the state of that action. This is a Bandit problem[10].

Once the firm recognises that others are learning about the market and its strategies, we are in the world of interactive learning – in addition to current payoff, learning about exogenous processes and advancing the state of the chosen process, the firm's actions will change the information available to others. This is analogous to an epistemic inference (communication) game.

5.5 Isolated Learning

Suppose first that a firm is (or believes itself to be) the only one learning. In this model the state consists of (random) demand together with the (formulaic) behaviour of other firms.

The firm's is assumed to seek to maximise the PDV of a per-period objective

$$U = E\left(\sum \delta^t \pi_i(\sigma^t)\right) \tag{1}$$

Its optimal action is found and that value of choosing action s in a given state defined by solving the usual Bellman equation

$$V\left(s^t_{-i}|\omega^t\right) = \max_{s^t_i}\left\{E\left(\pi_i\left(s^t_i, s^t_{-i}|\omega^t\right)\right) + \delta E\left(V\left(s^{t+1}|\omega^{t+1}\right)\right)\right\} \tag{2}$$

If V and the state updating function are unknown, the firm can try to learn them by adjusting an approximation Q:

$$Q_i(s|\omega) = E\left(\pi_i\left(s^t_i, s^t_{-i}|\omega^t\right)\right) + \delta E(\max_{s_i} Q_i(s_i, s_{-i})|s, \omega) \tag{3}$$

Where we note that

$$V_i(s|\omega) = \max_{s_i} Q_i(s|\omega) \tag{4}$$

Thus knowing Q allows the firm to compute the optimal strategy. A simple reinforcement learning strategy starts from an arbitrary initial Q^t and state ω^t. The firm chooses an action (e.g. price, quantity) s. The 'usual' version sets $\pi =$ the firm's profit; this can be replaced by alternative objectives that the algorithm seeks to maximise. After choosing the action, the machine observes the current value of the objective π^t and the new state ω^{t+1}. This allows the algorithm to update the value of Q corresponding to that specific combination of state and action to:

$$Q^{t+1}(s|\omega^t) = (1-\alpha)Q^t(s|\omega^t) + \alpha\{\pi^t + \delta \max_{s'} Q^t\left(s'|\omega^{t+1}\right)\} \tag{5}$$

[10] Gittins [5].

ω^{t+1} is the next state reached from ω^t when the action s is chosen. The rest of Q (in this simple model) does not change. The parameter α measures the rate of learning. If this process converges (e.g. in a model where the state represents the non-learning behaviour of other firms and variations in demand) it will converge to the 'true' Q function and select the optimal action in every visited state. To be sure of this, things must be completely mixed (ergodic in a useful sense). However, because this means experimenting with things that do not appear to be (and generically are not) optimal in any given state, such experiments are costly and an optimal learning strategy might be a desired outcome (hence the connection to other forms of stochastic optimisation as in Gittens [5] or complex optimisation strategies, which go beyond our current scope). Because this is a complicated problem, we here settle for prescribing the 'exploration' or experimentation phase (what actions are tried in a given state and how often). For real policy applications, however, more will need to be known.

How will exploratory probabilities be chosen? The simplest approach is to choose the optimal action (with the current value of Q) with probability $(1 - \varepsilon)$ and to randomise uniformly across all actions that are currently suboptimal with total probability ε. This can be improved by letting the probability of a given 'experiment' be a decreasing function of the current expected payoff; a common expression is for the probability of choosing a given action to be log-linear in the expected cost (compared to the current optimum). As we shall argue, this can greatly affect the behaviour of interacting learning processes[11].

The updating function gives the transition probability to a pair consisting of current reward and the next state; it can be learned if it takes the form of a time-invariant distribution $F(\pi^t, \omega^{t+1} | \omega^t, s^t)$.

One form of learning is to choose the currently-optimal action with high probability and randomise across all others. This is analogous to the learning process in games used in analysing dynamic play of coordination games (cf. Kandori, Mailath and Rob [6], Young [7]). The experimentation can also be made to vary over time. For example, in this one-player environment, the costs of experimentation increase and the expected benefits decrease as the process converges, so the programme should experiment less frequently and less widely.

Convergence

This type of simple one-person learning process determines an irreducible Markov Chain and thus converges almost surely. More generally, convergence is guaranteed if the experimentation strategy

[11] In an uncoupled learning environment, this will increase the chances that play converges – if at all – to proper rather than perfect equilibria in some environments and to acceptable and predominated correlated equilibrium play in others. When the game involves choice of learning algorithms rather than prices or quantities this sampling rule of avoiding big 'mistakes' will force uniform selection of risk dominant algorithms (or a suitable generalisation) whereas a more adventurous rule such as uniform experimentation will have other points of convergence that are neither payoff nor risk dominant. The Q-determined experimentation process is sometimes called Boltzmann experimentation.

- Experiments less frequently over time;
- Makes completely mixed choices at any state that is visited infinitely often; and
- Chooses the optimal action with probability 1 and time tends to infinity.

Unfortunately, this convergence may take a very long time. This is because Q must be estimated for all combinations of states and actions, and because the 'lessons' (adjusting one part of Q) may have external consequences if the underlying distribution is not fixed or varies in a complicated way. These same 'externalities' are even more significant when the updating function is not time-invariant (e.g. because it reflects others' learning or if consumers have finite demand and are thus sampled without replacement).

As a practical matter, the specification of the state and action sets imposes a cost of its own, so the choice of a suitable approximation may also be important (and impossible to verify).

Finally, the speed of convergence may be affected by the nature of the transition function. Suppose that we represent it as a weighted network by considering the probability of transition from a given state and action to another state as a weighted link. For any given state, we can define the set of states that can be reached with at least some probability p as the p-neighbourhood of that state. A slow growth network has the property that the p-neighbourhoods of a state's p-neighbourhood overlap substantially with each other and the original state's p-neighbourhood. In other words, experimentation will densely explore the region around a given state within a short period of time, leading to durable learning before moving on to consider other areas. This systematic exploration makes convergence much faster than more diffused learning processes.

If the optimality of an algorithm takes the 'cost of learning' into account and the payoff function is concave, this may be dominated by an approach that quickly settles on a strategy that is only approximately optimal.

The multiagent version of this approach (in which each firm regards others as part of the market environment) is independent learning.

5.6 Independent Learning

As mentioned above, firms that use this kind of simple learning mechanism may be overlooking the fact that others (firms, customers, etc.) are also learning. Because learning by definition induces non-stationarity in the firm's behaviour, a firm that ignores others' learning will mistakenly assume that the environment about which it learns (here represented by Q) is time-invariant. If such learning converges, of course, this is a rational expectation. There is no guarantee that learning will converge, but if it does so it will converge to a Nash equilibrium.

When many firms are learning, but demand is stationary (albeit random) the natural model is a repeated game. However, without additional structure, the state space is the set of histories, which grows as the game progresses. This complication can be avoided by limiting the size of the state space (e.g. to finite memory or to finite automata) or otherwise ensuring that strategies do not become too complicated. An example is provided by positional games e.g. Noughts & Crosses, Nim, Go or Chess (within limits) in which the current state of play is a sufficient statistic for all prior histories.

This does not remove all complications. Consider common pool externality games, in which some of the stock of a resource is consumed in each period; the rest is renewed or updated before the game proceeds to the next round. Markovian equilibria are defined by looking for strategies that solve the relevant Bellman equation in a way that depends only on the current state.

Concretely, let S^t be the stock at time t, c_i^t be the consumption by player i and $U_i\left(c_i^t\right)$ be the associated payoff in period t. Renewal is given by the equation

$$S^{t+1} = f\left(S^t - \sum_i c_i^t\right) \tag{6}$$

A Markovian strategy is one in which consumption is determined by current stock. Stationary Markovian equilibrium strategies $\widehat{c_i}$ should solve

$$V_i(S) = \max_{c_i}\left\{U_i(c_i) + \delta_i V_i\left(f\left(S^t - \sum_{j\neq i} \widehat{c_j}(S) - c_i\right)\right)\right\} \tag{7}$$

One way to solve this is to 'explore' in the Q-learning style, but this need not converge. Solving this directly may be computationally challenging. A heuristic method is to start at an arbitrary endpoint and work backward, retaining the future optimal strategies. In this approach, the value function V_i^T and thus the optimal strategies depend on how long the game has left to run.

However, because of the strategic dependencies, even when the utility functions are concave and the renewal function is smooth things become unmanageable; with 2 periods left in the game the value functions become non-concave and with three they may become non-smooth or even discontinuous. General solutions are only possible for isolated choices of functional form[12], and can usually be improved on by non-stationary strategies that involve threats and punishments.

In the computer science approach to these issues a range of improved approaches to the independent learning set-up have been developed in recent years that have shown quite impressive performance (in terms of convergence, speed and profitability)in positional games[13] and simple video games[14].

This is a critical observation; it may be that reinforcement learning strategies can learn how to punish. But as yet there are no general convergence results for such processes – in other words, there is no reason to believe that ML-powered oligopolies will converge to a Nash equilibrium, even of a repeated game. Therefore, analysis of their behaviour should not be limited either to a 'comparative statics' analysis of whether or not they have chosen a socially inefficient equilibrium or a concentration on behaviour in the support of the long-term limiting distribution of a Markov chain.

[12] Specifically, a stationary Markov equilibrium to which learning might converge only exists when U is logarithmic and f is a power function or when U is linear and f is logarithmic. See e.g. Garcia [8] or Amir [9].

[13] See e.g. Silver *et al.* [10].

[14] E.g. the algorithms developed for the game of Pong in Tampuu *et al.* [11].

5.7 Mutual Learning

The Q-learning approach has the advantage of guaranteed convergence in non-strategic learning models. However, they require finite state and action spaces; finite approximations slow learning and convergence[15]. Moreover, they do not take others' learning into account.

To take into account learning by others, it is necessary to adjust the prediction of the next state following an action to incorporate other players' reactions. In conceptual analysis of dynamic games, equilibrium outcomes are identified ex post; this means that the strategies (including exploitation and experimentation) are common knowledge among the players and each strategy can be tested in order to see whether it is a best reply to the others'. This approach cannot be applied when learning others' strategies, but some refinement can be obtained if it is assumed that each player is playing according to an assumption about the kind of equilibrium behaviour being followed. More concretely, the Q about which the firm learns will be a payoff function for a game (a matrix in the case of a two-player game). In the standard approach a player with current belief Q^t will assume that the other player will play according to a Nash or correlated equilibrium in that matrix and predict the next state accordingly[16]. In order to do this, however, the algorithms would need to overcome several challenges:

- Other firms' payoffs (profits) as well as actions (e.g. prices) may not be observable, or only partially observable, depending on e.g. the market contact among firms – (e.g. if two firms share output or input markets, some inference regarding the rival's payoff can be drawn from the firm's own profits);
- The amount of information that has to be collected and analysed grows rapidly because rival's possible inferences must also be tracked;
- In the case of multiple equilibria, coordination problems (different algorithms assuming different equilibria for the current beliefs) can rise; and
- Firms may be unable to determine whether a rival's actions correspond to the exploitation phase (in which case they should be viewed as equilibrium or profit-maximising behaviour) or the exploration phase (in which case they carry little information about the current state of the rival's beliefs), again leading to potential miscoordination.

To take the possibilities of mutual learning fully on board, firms that are aware that others are learning may also re-evaluate their own choices. In a standard reinforcement learning approach there is a sharp distinction between actions taken for profit and those taken for learning; in a setting with multilateral or entangled learning these actions also influence others – in other words, they constitute *communication*.

---

[15] Smooth approximations to Q combined with deep learning can speed convergence by adjusting neighbouring parts of Q as well as the one explored. Another approach developed by Google updates separately the 'current state' and 'next state' Q^t terms on the RHS of Eq. (5).

[16] For the Nash case see Hu and Wellman [12]; for the correlated equilibrium see Greenwald, Hall and Serrano [13]; for an earlier survey see Shoham, Powers and Grenager [14].

- This implicit communication may provide the basis for future elaboration of a suitable policy response, even though it is the algorithms of one firm that communicate to the algorithms of others.
- In this way, firms' actions become loosely coupled, which may be sufficient to sustain supranormal profits and limit competition.

Aspects of this are covered in the computer science literature that uses various forms of deep learning, in which the deeper layers of the model correspond to other players' models and strategies.

Mutual Inference; the Importance of Structure

As mentioned above, when firms take into account others' learning, there is an aspect of communication in their interactions. The efficiency of learning depends, therefore, on who communicates with whom and what information each can observe. In an economic context, this may involve firms observing other firms prices or quantities – there is a clear difference between a model in which all prices can be observed and all firms share the same market and one in which price observations may be selective and market externalities are determined by a network of overlapping markets[17]. Even in a theoretical model, this may complicate normal analysis. In this section we give an example.

Suppose that n firms are competing in markets with uncertain demand, and that each firm has private information about the (time-invariant) distribution of demand. We may represent this by saying that the demand distribution depends on a single parameter $\omega \in [0, 1]$ and that firm i's information is represented by a partition Π_i^0 – in other words, when the true state is ω, the firm is informed that the true state belongs to the element $\Pi_i^0(\omega)$ that contains it. The firms' partitions (but not their actual private information) are common knowledge – each firm knows what each other firm could observe, but not what it has observed. The firms also have a 'language' in the form of a function giving the publicly-observable action $\sigma_i(E)$ the firm would take when it knew that the true state belonged to a set E[18]. For example, $\sigma_i(E)$ could be the amount produced or the price charged when the firm knows E. Learning proceeds as follows.

Learning in a Clique

In period 0, each firm takes the action $\sigma_i^0 = \sigma_i\left(\Pi_i^0(\omega)\right)$ corresponding to its initial private information.

It also observes the actions σ_j^0 taken by other firms. This allows it to 'refine' its information; instead of $\Pi_i^0(\omega)$, it now knows that the true state belongs to

$$\Pi_i^1\left(\omega|\sigma^0\right) = \bigcap\nolimits_{j \neq i} \left\{\omega \in \Pi_i^0(\omega)|\sigma_j\left(\Pi_j^0(\omega)\right) = \sigma_j^0\right\} \tag{8}$$

[17] Note that these networks of observation and of market interaction need not coincide; a firm can observe rivals' prices published on the Internet even if they sell in separate markets, and firms can be affected by others' prices and sales even when they do not observe prices directly.

[18] Formally, E must be expressible as a union of intersections of the various events observable by the firms – the field generated by the join (coarsest common refinement) of the partitions.

In other words, the partition Π_i^1 is the set of states that might have been the true state and that are compatible with the observed actions of the other players. Note also that

- each player will know how the other players will have revised their beliefs (in other words, player i will know Π_j^1); and
- For each player, Π_i^1 is strictly finer (contains more information than) Π_i^0

After t periods, the players will have partitions Π_i^{t} and take actions $\sigma_i^t = \sigma_i\left(\Pi_i^{\mathrm{t}}(\omega)\right)$. The updating can be written as:

$$\Pi_i^{\mathrm{t}+1}(\omega|\sigma^{\mathrm{t}}) = \bigcap\nolimits_{j\neq i}\left\{\omega \in \Pi_i^{\mathrm{t}}(\omega)|\sigma_j\left(\Pi_j^{\mathrm{t}}(\omega)\right) = \sigma_j^t\right\} \tag{9}$$

Because each partition for each player strictly refines the one that precedes it, this Bayesian learning process converges monotonically to a unique equilibrium set of partitions Π_i^* and associated actions σ_i^*. At this point, we can also identify the common knowledge of the agents with the meet (finest common coarsening) of the equilibrium partition[19] - in symbols $M = \bigwedge_i \Pi_i^*$. Of course, the actions are arbitrary; however, it has been shown[20] that the players will behave as though the common knowledge were the only information available – in other words they will agree and take actions $\sigma_j(M(\omega))$ if the action rule has the property that, whenever two disjoint events lead to the same action, this action will be taken when the player does not know which event occurred. In other words, this learning process converges to a state of effectively complete learning. One such action rule is a market choice (e.g. price) that maximises expected profit.

This model can be used for situations of random demand and private costs equally well. It recognises and exploits learning by others, albeit in a myopic fashion. The analysis only shows that the process converges. It is possible, for instance, that a firm might wish to manipulate the behaviour of others by departing from the action rule that they use to interpret its behaviour. Outright falsification (behaving in accordance with a state that the firm knows has not occurred) creates problems for the Bayesian algorithm, because it produces logically inconsistent information. However, the firm could decide to conceal information e.g. by deliberately coarsening the information structure on which it bases its actions. This will lead to a different equilibrium (albeit with the same common knowledge property). For instance, if a firm chooses the same initial strategy regardless of its private information, that information will never be revealed, which in turn will affect the ability of other firms to learn. Conversely, if each firm chooses an action that is different in every element of its partition, the process will converge immediately to full information revelation. This in turn is likely to have competition consequences. For example, full cost revelation may facilitate collusion; full revelation of customer information may facilitate competition and/or price differentiation, etc. This further illustrates that the same set of algorithms will behave differently depending

[19] Aumann [15].

[20] Cave [16].

on the information that is supplied to them, and thus that restricting the structure of pricing algorithms may distort but not prevent algorithmic collusion.

Learning in a Network
Now suppose that we use the same set-up in a networked market environment e.g. a circle of n firms in with each firm interacts (in terms of observation or market contact) with the firms on either side. The first phase of the game proceeds as before; the firms take their initially-optimal actions, observe those of their network neighbours and revise their beliefs according to Bayes' Rule. However, it is no longer the case that all of the Π_j^1 are common knowledge. Firm 3, for instance, will be able to take both firm 2's and firm 4's information into account, but firm 2 will not know what firm 3 learned from firm 4. In modelling Π_3^1, therefore, firm 2 will know only what firm 2 chose at round 0 and what firm 3 observed from firm 2. Of course, firm 2 will have some idea of what firm 4 might have done since it will know that the true state will have to be one that is compatible with firm 2's private information and with the observed actions of firm 1 and firm 3. So it can only form a conditional belief about firm 3's information in Π_3^1; for every state that might have led to a different action by firm 4 in period 0, there will be a different realisation of Π_3^1. This in turn means that the information conveyed by firm 3's next action is ambiguous (conditional on Π_3^1).

Cutting to the chase, this means that the state space on which learning operates will potentially expand as the game goes on. But this is not necessarily the case; firm 4's observation is not arbitrary. For example, if firm 4 knows strictly less than firm 3 (has a coarser partition), then firm 2 can work out Π_3^1 uniquely – it will be the same as if firm 3 was not in contact with firm 4. Alternatively, the market shared between firms 3 and 4 may be such that there is no difference between the optimal actions of firm 4 across states that firm 2 cannot distinguish (in which case again firm 2 can work out what firm 3 knows and therefore what firm 3's next action implies).

Finally, the network may help to trim the set of possibilities. Suppose that all the markets are replicas of each other (or sufficiently similar samples from a much larger population). This could happen if e.g. firms 2 and 3 compete on one platform and firms 3 and 4 compete on another but the users of the two platforms are drawn from the same distribution. Then the information generated in the 1:2, 2:3 and 3:4 markets will be substantially similar. There may be differences in pricing if the firm's costs are different, but these will be fixed factors and eventually all firms will 'learn' the demand distribution. Another form of trimming may come from the structure of the network. If the firms' contacts form a closed cycle and the algorithms converge, the same common knowledge as in the clique (full network) will be stable. It may not, however, be the only point of convergence.

6 Conclusion

This paper has summarised early results from a structured exploration of the implications of machine learning for competition policy. It is linked to a range of current policy areas including privacy, online search regulation, informational intermediation, digital profiling, algorithmic bias and the implications of artificial intelligence. Existing literature shows that algorithmic tacit collusion is possible but in many settings unlikely. This paper argues that there are many factors that could increase the likelihood and robustness of such outcomes. It also suggests that it may be necessary to reconsider the legal status and regulatory approach taken to tacit collusion, because it is no longer obvious that it can be dismissed as essentially harmless or regarded as disproportionately hard to tackle (because regulatory use of AI can to an extent help with identifying harmful uses). Of course, if the damage to market efficiency is inadvertent and an unnoticed, unintended or emergent consequence of algorithmic market practices, it may be neither possible nor efficient to hold user firms or programmers responsible; instead restrictions on information exchange and a range of detection approaches may help to diagnose problems.

References

1. Cave, J., Gelsomini, L.: Single-bank proprietary platforms (2010). https://efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2012-Barcelona/papers/EFMA2012_0482_fullpaper.pdf. Accessed 02 Sept 2019
2. Maskin, E., Tirole, J.: A theory of dynamic oligopoly, II: price competition, kinked demand curves, and Edgeworth cycles. Econometrica J. Econm. Soc. **56**, 571–599 (1988)
3. Katz, M.L., Shapiro, C.: Network externalities, competition, and compatibility. Am. Econ. Rev. **75**(3), 424–440 (1985)
4. Galeotti, A., Goyal, S.: The law of the few. Am. Econ. Rev. **100**(4), 1468–1492 (2010)
5. Gittins, J.C.: Bandit processes and dynamic allocation indices. J. R. Stat. Soc. Ser. B (Methodol.) **41**(2), 148–164 (1979)
6. Kandori, M., Mailath, G.J., Rob, R.: Learning, mutation, and long run equilibria in games. Econometrica J. Econm. Soc. **61**, 29–56 (1993)
7. Young, H.P.: The evolution of conventions. Econometrica J. Econm. Soc. **61**, 57–84 (1993)
8. Garcia, A.: Forecast horizon for a class of dynamic games. J. Optim. Theory Appl. **122**(3), 471–486 (2004)
9. Amir, R.: Stochastic games in economics and related fields: an overview. In: Neyman, A., Sorin, S. (eds.) Stochastic Games and Applications. ASIC, vol. 570, pp. 455–470. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-010-0189-2_30
10. Silver, D., et al.: Mastering the game of Go without human knowledge. Nature **550**(7676), 354 (2017)
11. Tampuu, A., et al.: Multiagent cooperation and competition with deep reinforcement learning. PLoS ONE **12**(4), e0172395 (2017)
12. Hu, J., Wellman, M.P.: Multiagent reinforcement learning: theoretical framework and an algorithm. In: ICML, vol. 98, pp. 242–250, July 1998
13. Greenwald, A., Hall, K., Serrano, R.: Correlated Q-learning. In: ICML, vol. 3, pp. 242–249, August 2003
14. Shoham, Y., Powers, R., Grenager, T.: If multi-agent learning is the answer, what is the question? Artif. Intell. **171**(7), 365–377 (2007)
15. Aumann, R.J.: Agreeing to disagree. Ann. Stat., 1236–1239 (1976)
16. Cave, J.A.: Learning to agree. Econ. Lett. **12**(2), 147–152 (1983)

The Online Bystander Effect: Evidence from a Study on Synchronous Facebook Communications

Andrea Guazzini, Enrico Imbimbo(✉), Federica Stefanelli, and Giulia Bravi

Department of Education, Languages, Intercultures, Literatures and Psychology, University of Florence, Via di San Salvi, 12, Building 26, Florence, Italy
{andrea.guazzini,enrico.imbimbo,federica.stefanelli}@unifi.it

Abstract. According to the Bystander Effect, the tendency to help a victim in a problematic situation decreases with the number of spectators to whom the request for help is made The main aim of this study is to investigate this phenomenon in an online environment characterised by synchronous communication. This paper evaluates how some subjective variables and the participants' familiarity with online interactions can affect the tendency to help others. The results of the study show that the Bystander Effect is present even in online environments. In particular, the results indicate that the tendency to help decreases not linearly with the increasing of the group size which received the help request. Moreover, the familiarity that a person has with online interactions has a significant association with his tendency to help. Finally, the general self-efficacy and social desirability appear as predictors of Bystander behaviour during online interactions with a victim.

Keywords: Bystander effect · Online social dynamics · Online prosocial behaviour

1 Introduction

Nowadays, in a world where more than half the population has access to Internet, and 2.5 billions of them last year used a mobile device to connect to social media (e.g. Facebook, Tweeter) [1], the pervasiveness and importance of human online interactions must be taken into account to understand the brand new social phenomena occurring within a world populated by a growing number of "ICTs users". The impact of the Information and Communication Technologies (ICTs) on human behaviours is currently an important field of study, in which psychological constructs such as "Altruism" [2,3],"Reciprocity" [4], "Conformity" [5] and "Participation" [4] attracted the attention of researchers both for theoretical research, and for many possible applications (e.g., crowdsourcing, crowdfunding, participatory citizenship). Altruism and reciprocity represent two of the main driving forces that make human beings as one of the most cooperative animals in nature [6,7], and many studies described their dynamics within a plethora of

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 153–167, 2019.
https://doi.org/10.1007/978-3-030-34770-3_12

scenarios [8–11]. A very famous example of such studies is the Bystander Effect, which refers to the social phenomenon that happens when individuals are less likely to help a victim when other people are present [12]. Alongside a large amount of studies that have examined this kind of effect in "real" environments and face to face interaction [13–19], only a few numbers of studies have explored the persistence of Bystander Effect in an online environment. Moreover, all the works that studied the Online Bystander Effect (OBE) focused their attention on asynchronous interactions between the group members, without any manipulation of the group salience [20–26]. Recent chronicles sadly confirm how the Web can support the emergence of extreme behaviours (i.e. suicide) through the use of online platforms, and how the specific features of the online interactions should be considered in order to understand the human online dynamics. In this study, to partially answer to the previous issues, the first step was to verify the existence of Bystander Effect in online synchronous interactions, within minimal group conditions, confirming its persistence even within an online environment.

The Bystander Effect was first investigated in the United States in the 1960s after the famous case of Kitty Genovese. The term refers to a psychosocial phenomenon that appears in emergency circumstances, and in the presence of other people, who, despite the repeated requests of the victim, decides not to give help. The phenomenon seems to correlate with the ambiguity of the situation and seems to be directly proportional to the number of spectators present at scene [12,27–30].

Literature provides two explanations of the Bystander Effect: the first refers to the fact that in an ambiguous situation, people tend to determine the severity of the event, partially based on the behaviour implemented by other spectators. If no one is perceived as getting alarmed, the situation will not be perceived as requiring an intervention [13,20,24,31].

The second explanation is that in the presence of others, people will tend to manifest a generalized apathy caused by the spread of responsibility [20,32].

Some studies in natural environments investigated the role of several personality variables in relation to the Bystander Effect [33,34]. In particular, openness and consciousness are positively correlated to the tendency to adopt help behaviors in emergencies [35]. Self Efficacy has been found to increase the probability to exhibit the Bystander Effect [36]. Interestingly, in real environments, many fundamental psychosocial traits and dimensions, such as Social Desirability, have been frequently found as not significantly related to the Bystander Effect [32].

In the past years, the Bystander Effect has been highly studied in both natural and laboratory contexts [37]. Today, with the advent of new technologies, as well as many of the psychosocial phenomena, also the Bystander have been re-examined to be modelled within online environments. The basis of these studies is a need to understand whether the characteristics of these new environments, such as partial or total anonymity, physical isolation and reduced identification, can alter the scientific evidence coming from real contexts [20–23,25].

Previous studies highlighted that the Bystander Effect also occurs in online contexts and how this phenomenon intensifies as the number of spectators in the group increases [23–25,38].

About this, some authors [24] demonstrated that if an asynchronous communication mode characterizes an online context, it is more likely to receive help if the request is made personally to only one interlocutor at a time, rather than if the request is made through a collective e-mail.

Similar results, always in asynchronous online communication conditions, showed that the Bystander Effect also manifests itself if the victim is a physically present person within the formal reference group of the recipient of the help request (i.e. university contest) [38].

Another study has shown how partial anonymity, in computer-mediated communication, influences the tendency of the actors involved to help [39].

Moreover, about the role of personality variables in relation to the Bystander Effect, some authors have pointed out that in an online environment the extent of the phenomenon is indirectly proportional to the age of the spectator [40].

Furthermore, most evidence has revealed that there are no significant differences in the amount of help provided in relation to the gender of the victim even in online environments [25,37].

Despite the latest evidence about the "Online Bystander Effect", the main results provided by literature regarding the Bystander Effect in real environments should be replicated in order to understand the impact of ICTs on this complex psychological dynamics.

2 Aims and Hypotheses

The main aim of this study was to investigate the Bystander Effect in online environments, for the first time in an "ecological" way for what concerns the modern WEB based environments (i.e., under synchronous communication, and "minimal group" conditions).

In accordance with the literature, first we hypothesized the existence of the Bystander Effect also within online communities, characterized by the constraints we defined previously [23–25,38], expecting to observe a reduction of "help actions" with the increasing of the size of the group, in which the help request was carried out during a web-based synchronous interaction. Moreover, we hypothesized an effect of age, following literature [40], such that older people should give more help than the younger even within online environments.

The second aim of this study was to analyse the impact of the subjects' variables (i.e., socio-demographic, psychological, and psychosocial dimensions) on the adoption of online help behaviours, in order to improve, confirm and test the preliminary results provided by the literature.

Finally, the third aim of this research was to shed light on how the familiarity with online interactions (e.g., technological fluency, computer literacy) and online environments, could impact the subjects' tendency to give help within such environments.

3 Methods

In this section the design of the study will be presented, including the sample characteristic, the modality followed in the experimental phase to collect data, the tools and psychological measures used, and the analysis implemented in order to verify the hypotheses.

3.1 Sample

The research was conducted in accordance with the guidelines for the ethical treatment of human participants of the Italian Psychological Association (AIP). The participants were recruited with a snowball sampling strategy. All participants signed informed consent and could withdraw from participation at any time. The sample was composed of 176 students coming from the School of Psychology of the University of Florence, recruited utilizing a voluntary census procedure. The sample was composed mainly by Italians (97.2%), and by females (73.3%), with an average age of 21.87 (SD 1.89). The subjects were approached face-to-face by the experimenter and were invited to send an e-mail if they were interested in participating in the study. The experiment was introduced using a fake story not to invalidate/affect data mining. They were told that they would take part in research, including simple questions of moral reasoning, for a study about quantum cognition. The participants were asked to have a Facebook account, while the part related to the Bystander Effect, was revealed only at the end of the experiment.

3.2 Experimental Procedure and Tools

The recruited subjects were contacted and instructed directly online, using a power point presentation of the fake-study. The subjects could participate to the experiment in an autonomous way, just using an ID code to participate. The subjects, depending on the codes they received, were randomly assigned to 60 different online groups (i.e., composed by 3, 4 or 7 people), created specifically as "Facebook Groups" on the web. In all of this group the participants didn't know each other and could not speak. Within each group a researcher's disguised helper was always present (i.e., a male or a female so to obtain two balanced conditions), that at the end of each session was instructed to ask for an help to the experimental community. In order to assess the psychological and socio-demographical dimensions of interest two surveys were developed, adopting both standard tools for psychological assessment and ad-hoc sections created to assess the factors under scrutiny. The first instrument that was administered was an online psychological survey created using Google form platform, investigating: demographic variables (e.g., sex, age, degree, years of education, family income), operational variables related to the usage of new technologies (e.g., frequency in the use of the various internet services, which devices are more used and how often, for which activities they are mainly used), the specific usage of social networks (e.g., which platforms are used most, the aim and the contexts in which they are most used and the importance of the interpersonal relationships created on these platforms), and the psychological dimensions of interest.

At the beginning of the experiment, two pairs of statements about moral reasoning were proposed to the subjects, and they were asked to assign a score of importance on a scale ranging from 1 to 100. Between the administration of the first and second couple of questions, the anonymous helper posted in the Group chat a help request to the other members in order to recruit subjects for his or her thesis project experiments. The accomplice used always the same standardize message as help request.

At the end of the experiment, the subjects were thanked for their participation, and a restitution survey, that also revealed the real intent of the study, was sent to them. The final survey had the target to get some opinions and perceptions about the experiment, as well as to determine if the subjects believed to the help request, how much they were affected by the group behaviours, and by the "online" conditions in which the social interaction occurred.

The preliminary survey, assessing the psychological dimensions was composed by the following psychological, psychosocial and personality scales:

- BEES (Balanced Emotional Empathy Scale, $\alpha = 0.87$) [41]
- FNE (a brief version of the Fear of Negative Evaluation Scale, $\alpha = 0.90$) [42]
- SRS (Social Responsibility Scale, $\alpha = 0.74$) [43]
- BSDS (Brief Social Desirability Scale, $\alpha = 0.60$) [44]
- I-TIPI (Italian Ten Item Personality Inventory) For personality dimensions such as: extroversion, agreeableness, conscientiousness, neuroticism and openness ($\alpha = 0.59$) [45]
- STAI (State-Trait Anxiety Inventory, $\alpha = 0.60$) [46]
- General Self Efficacy Scale ($\alpha = 0.90$) [47]
- Sense of Community Scale ($\alpha = 0.88$) [48]
- SIAS (Social Anxiety Scale, $\alpha = 0.88$) [49]

At the end of the experiment, in order to validate the collected data, a second survey was used to test the subjects opinion and perception about the experiment. The interface of the questionnaire opened with an explanation of the study to which the subjects had participated, and also revealed the real intent of the research related to the Bystander Effect. The questions proposed by the survey were all related to this part of the experiment, and investigated the probability of having understood in advance the aims of the study, which were the factors/elements that inspired her/his action to the message of the accomplice (whether or not to respond), and finally, if the presence or possible interactions with the other members of the group had influenced the decision to assist.

3.3 Data Analysis

The data analysis procedures have been organized within 2 main phases. During the first one the data mined from the experiments have been organized, codified, standardized and the descriptive statistics were produced, in order to verify the preconditions for the inferential analysis (i.e., Gaussian distribution for the continuous variables, and balance and minimal sample size for the discrete variables), and removing, where necessary, the outliers. Thus, is been assessed a sample size estimation in order to verify the power of the experiment results. During this phase two brand, new variables have been defined respectively as "Online Bystander Effect" (OBE) and "Provided Support". OBE has been defined as occurring when a subject didn't support the accomplishes nor with a help ($H_i = 0$) neither with a message ($M_i = 0$), instead, the variable "Provided Support" represents the cases in which a subject said he was willing to help the accomplice who had asked for help. As a consequence the variable OBE_i for the subject i can be expressed, as an approximation useful to detect a possible general law that underlies the effect, by the Eq. 1.

$$OBE_i = \begin{cases} 1, & \text{if } H_i \text{ and } M_i = 0 \\ 0, & \text{if } H_i \text{ or } M_i = 1 \end{cases} \tag{1}$$

Moreover, the answer response time of participants feedbacks have been collected and introduced in the analysis. Finally, in the second phase, a multivariate logistic regression analysis has been implemented to estimate the best model describing the OBE.

3.4 Sample Size Estimation

To obtain a robust approximation of the optimal sample size, disregarding the debate about the standard sample size estimation required for GLMM [50], it is been conducted a power analysis by reducing the hypotheses to the case of 1 proportion test, with one sample, and under a 2-sided equality hypothesis (Eq. 2) [51]. The results are reported in Table 1.

$$n_b = p(1-p)\left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{p - p_0}\right)^2 \tag{2}$$

with

$$1 - \beta = \Phi(Z - Z_{1-\frac{\alpha}{2}}) + \Phi(-Z - Z_{1-\frac{\alpha}{2}}) \tag{3}$$

and

$$Z = \frac{p - p_0}{\sqrt{\frac{p(1-p)}{n}}} \tag{4}$$

where, P_0 is the comparison value, Φ is the standard Normal distribution function, Φ^{-1} is the standard Normal quantile function, α is Type I error, and β is Type II error, meaning $1 - \beta$ is power.

The analysis revealed that approximately 151 participants would be needed to achieve 90% power ($1 - \beta$) at a 0.01 α level ($\alpha = 0.01$), so validating the sample size used in the experiment that is of 176.

Table 1. Using the variable provided Support as discrete dependent measure, has been designed a sample size estimation to evaluate one proportion from one sample, under a 2 sided equality hypothesis. It is been required a Power $(1-\beta)$ of 90%, and an Type I Error confidence level (α) of 1%. p_0 is actually the null hypothesis (H_0) of the statistical test applied to estimate the optimal sample size.

Dimension	Experimental Percentage	Error (p_0)	Sample size Required	Sample size Available
Provided support	13%(7%)	8%(15%)	151	176

4 Results

In this section we report the main results obtained from the experiment, organizing the section within three subsections. In the first, the descriptive statistics of the psychological and socio-demographical dimensions (Table 2), and of the operative dimensions (Table 3) are presented. In the second section the magnitude of the Bystander Effect in online environments, as well as the impact of group size on the Bystander dynamics is provided (Eq. 5). Finally, the best multivariate logistic regression model describing the dynamics of the Online Bystander Effect is provided (Table 4).

4.1 Descriptive Statistics

Table 2. Descriptive statistics of the sample

Sample descriptive statistics				
Variable	Average (SE)	Std. Dev.	Skweness	Kurtosis
Age	21,87(0.15)	1.89	2.43	8.33
Social Desirability	3.69(0.13)	1.75	0.25	−0.26
General Self Efficacy	28.28(0.35)	4.53	0.13	0.04
Consciousness	7.48(0.15)	1.90	−0.73	0.20
Openness	6.95(0.13)	1.74	−0.26	−0.57
Importance of Contacts on SNS	2.40(0.06)	0.78	0.22	−0.29
Message activity on SNS	2.68(0.07)	0.92	0.72	0.33
Number of Topics on SNS	3.52(0.14)	1.86	0.23	−0.23
Answer delay (Minutes)	22.88(4.27)	17.11	0.13	1.09
Forum Usage	16%(0.3%)	37%	-	-

In Table 2 the psychological and socio-demographical dimensions, as well as the variables describing the ICTs life of the experimental subjects are presented.

In particular, the WEB activity of the subjects assessed through the surveys filled during the experiment emerged by these data. Participants appear to usually give not very high importance to contacts on SNS ($M = 2.4, SD = 0.78$; $Min = 0, Max = 5$) and do not communicate very often with contacts in their SNS list ($M = 2.68, SD = 0.92; Min = 0, Max = 5$). They report discussing averagely 3 different arguments in their SNS life (M = 3.52, SD = 1.86). The 16% of subjects in the sample is active in online forums or discussion groups. The sample is characterized by average levels of Social Desirability (M = 3.69, S.D. = 1.75), General Self Efficacy (M = 28.28, SD = 4.53), Consciousness (M = 7.48 SD = 1.90) and Openness (M = 6.95, SD = 1.74). Time response during the experiment, for those who accept the request or at least interact with the needy participant, varied from a minimum of 2 min to a maximum of 50 min, with a mean of 24.88 min ($SD = 17.11$).

Table 3. Operative descriptive statistics of the sample.

Operative descriptive statistics		
Variable	N	Percentage
Answer to the request	16	9.5%
Provided Support	13	7.7%
Bystander Effect	136	80.5%

Data reported in Table 3 shows the descriptive statistics of the experimental observables. The 7.7% of participants showed a helpful behaviour, responding actively to the request addressed, while the 9.5% of the sample, even without a direct acceptance of the request, interacted with the accomplice in a supportive manner. As a consequence, the 80.5% acted as a bystander in the experiment, i.e. neither assisting who asked for help nor interacting with him. In order to study the bystander dynamics, three new variables can be defined. The first was related with the "Answer to the Request", and indicated all those subjects who responded to the help message of the accomplice, not distinguishing between those who offered support and who did not offer it. The second variable was defined as "Provided Support", and considers only those who have helped the accomplice. Finally, a last variable is introduced and labelled as "Bystander Effect", indicating all participants who, despite having read the message of the accomplice, decided not to respond to the message, and not to help.

4.2 Online Bystander Effect: Size of the Group

The OBE has been defined by this study, extending the definitions provided by classical literature, as the complex phenomenon that inhibits the subjects belonging to a large enough, from both giving help, as well as from just answering to the help request in an online environment. As previously described, in order to estimate the magnitude of the effect within our experiments, the Bystander Effect has been calculated as follows in Eq. 5:

$$Bystander = \left(1 - \frac{(\sum_{i=1}^{N} M_i + \sum_{i=1}^{N} S_i)}{2N}\right) 100 \tag{5}$$

Where, M_i represents the support message sent by the subject i, S_i is the support provided by the subject i, N is the total number of participants.

Experimental results confirm how, within the range of "small human groups", even in online environments the Bystander Effect appears to be consistent/robust even when the communication between subjects is synchronous and the group salience is enhanced; and it is related with the size of the group 1 (Fig. 1).

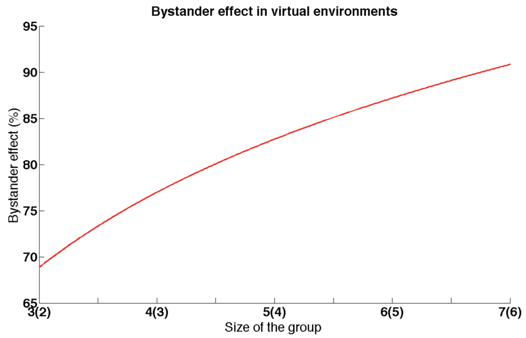


Fig. 1. The graph shows how increasing the size of the groups (x) the probability to observe the Bystander effect in the groups (Y) increases. The relation appears as to be best described by a non-linear function/trend.

A very preliminary fitting of experimental data supports the hypothesis of a non-linear relation (e.g., apparently a logarithmic relation), between size of the group and Bystander Effect (Eq. 6). In particular, the best fitting equation for the real data can be written as follows:

$$B_i = \alpha * \log(SG) + \beta \tag{6}$$

where SG is the size of the group, and the parameters of the best fitting equation for the experimental data are respectively $\alpha = 20$, and $\beta = 55$.

4.3 Online Provided Support: Multivariate Model

Bystander Effect is a phenomenon characterized by high complexity due to the high number of factors involved in its implementation and the interactions that take place between them. The implementation of a multivariate model represents a method to shed light on the main variables that could affect the occurrence of the Bystander Effect.

A multivariate logistic regression has been implemented in order to estimate the best model predicting Bystander Effect in online enthronements. The actual support provided during the online interactions has been studied taking into account all variables collected by the survey, and the groups' sizes.

Table 4. Logistic regression. Best model for Provided Support.

Best logistic model			
	χ^2	Df	Nagelkerke r^2
Model	48.15^{***}	10	0,59
Parameter	Coefficient (β)	Wald	Exp(β)
Group Size	−2.37	5.03^{*}	0.094
Age	0.24	6.48^{**}	1.267
Social Desirability	−0.56	4.12^{*}	0.569
General Self Efficacy	−0.15	2.44^{*}	0.858
Consciousness	0.42	2.40^{*}	1.518
Openness	0.78	5.53^{*}	2.183
Importance of Contacts on SNS	1.33	4.22^{*}	3.801
Message activity on SNS	1.70	3.40^{*}	5.468
Number of Topics on SNS	−0.76	5.26^{*}	0.465
Forum Usage	2.50	4.54^{*}	12.191

$^{***} = p < 0.001$, $^{**} = p < 0.01$, $^{*} = p < 0.05$

Table 4 shows the best predictors for a model aiming to prefigure the Actual Provided Support during our experiment. The best model appears to explain the 59% of the variance, including ten dimensions that can be divided into three main categories: Group Size; Demographic and Psychological dimensions; and Confidence with ICTs.

Size Effect and Demographic Dimension. The effect related to the dimension of the group on the tendency to engage in supportive actions results to be consistent ($\beta = -2.37$), the bigger a group in which a person is, the less he or she showed helpful behaviours. Age ($\beta = 0.24$) of participants results to affect the implementation of help actions directly. No significant effect of gender was found in the analysis.

Psychological Features Effects. Higher values of Social Desirability ($\beta = -0.56$) and General Self Efficacy ($\beta = -0.15$) increase the tendency not to help a person in the experiment. Moreover, the analysis carried out showed the effect of some personality traits in the display of helpful actions. In this study, high levels of openness ($\beta = 0.78$) and consciousness ($\beta = 0.42$) were found to increase the probability to behave in a supportive way with a person in need within a virtual environment.

Confidence with ICTs. Some aspect of ICTs usage were found to influence the analysed behaviour. In an online environment the importance given to contacts ($\beta = 1.33$), the number of message exchanges ($\beta = 1.70$) and the activity in online forums ($\beta = 2,50$) have shown to be directly related the propensity to answer to the request of support, conversely the number of topics addressed in social networking activity was found to decrease the probability to help someone in difficulty.

5 Discussion

This research aimed to investigate the Bystander Effect while the interactions between participants (victim and spectators) are online and synchronous, in a condition where group salience was enhanced. With the spread of Computer-Mediated Technology (CMC), many of the social phenomena studied in the past with face to face groups, have been resumed and studied in online groups. This new interest is justified by the evidence that, when the interactions between people are online, some factors that usually influence group dynamics to change their effect. This change occurs due to the peculiar characteristics of the online environment (e.g., anonymity and physical isolation). Indeed, how explained by the SIDE Model [52], when people interact online, and they are deindividualized, it is more likely that they comply with the group's norms.

The results have shown that even in this type of online environment the Bystander Effect is present, and as within real settings Sect. 4.3 it shows a positive relationship with the size of the group.

In particular, the research highlighted how the "tendency to help" in online environments appeared to decrease not linearly with the increasing of the size of the group in which the request was made (6).

Furthermore, evidence from this study suggests how some features of the subjects, such as their technological fluency and computer literacy Sect. 4.3, show a significant association with their susceptibility to the OBE. The main factors that appear to mitigate the OBE occurrence were the importance given to online social contacts, the frequency of online interactions, and the number of different arguments debated online (i.e., the degree of dispersivity of the SNS typical usage). In particular, those that are more interested in people met on social networks, to interact more with messages online, participate more in internet forums and discuss a smaller amount of topics on the social networks be less prone to incur in the OBE in online environments.

Besides, general self-efficacy and social desirability are significant predictors of the probability of being subject to the OBE during online synchronous interactions. People who care more about the judgement of others in social contexts and believe to be able to perform well in a task, tend to give less help.

Moreover, it has been found that high values of personality traits of openness and consciousness reduce the phenomenon analysed (i.e., OBE) and increase the tendency to implement aid actions.

Finally, age influences the OBE, indeed, in this study, older people were found to help more than younger.

This study suggests that some individual characteristics may serve as predictors to evaluate the probability of exhibit the Bystander Effect in online environments with synchronous interactions.

Some limitations can be ascribed to in the research presented in this paper. First of all, the sample used was composed mostly of young university students and females, limiting the possibility to generalize the results to the overall population. Another limitation of the research is represented by the lack of measurement to analyse the communication between participants outside the online group specially prepared for the experiment after they took part in it. A third limit is in the low degree of severity of the help request used by the accomplice that may have been not enough engaging in arising help answers in the participants. However, the adoption of this particular help request is justified by existing literature on this topic that used a request of similar severity [24,38]. This framework can be considered as the only suitable way to validate a procedure to study the effect of any possible factor (e.g., environments and group composition) on Bystander Effect within online environments.

Future researches may focus the effort in the investigation of the OBE expanding the sample widening the age of the subjects involved and balancing the sex in order to evaluate the possible role of empathy trait differently distributed among males and females population [53]. A further challenge, in the research on the OBE, could also be represented by the validation of an experimental environment where it is possible to monitor all communications between participants after the begin of the experiment.

Author Contributions. A.G. supervised the whole work. G.B. executed the experiments. A.G. executed the statistical analysis. E.I, F.S. and G.B wrote the manuscript. All authors read and approved the final manuscript.

References

1. Digital in 2019: Global Overview (16 April 2019). https://datareportal.com/reports/digital-2019-global-digital-overview
2. Leider, S., Möbius, M.M., Rosenblat, T., Do, Q.-A.: Directed altruism and enforced reciprocity in social networks. Q. J. Econ. **124**(4), 1815–1851 (2009)
3. Fowler, J.H., Christakis, N.A.: Cooperative behavior cascades in human social networks. Proc. National Acad. Sci. **107**(12), 5334–5338 (2010)
4. Dholakia, U.M., Bagozzi, R.P., Pearo, L.K.: A social influence model of consumer participation in network-and small-group-based virtual communities. Int. J. Res. Market. **21**(3), 241–263 (2004)
5. Postmes, T., Spears, R., Sakhel, K., De Groot, D.: Social influence in computer-mediated communication: the effects of anonymity on group behavior. Personal. Soc. Psychol. Bull. **27**(10), 1243–1254 (2001)
6. Nowak, M.A., Sigmund, K.: Evolution of indirect reciprocity. Nature **393**(6685), 573 (2005)
7. Warneken, F., Tomasello, M.: The roots of human altruism. Br. J. Psychol. **100**(3), 455–471 (2009)

8. Trivers, R.L.: The evolution of reciprocal altruism. Q. Rev. Biol. **46**(1), 35–57 (1971)
9. Gintis, H.: Strong reciprocity and human sociality. J. Theoret. Biol. **206**(2), 169–179 (2000)
10. Fehr, E., Henrich, J.: Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. CESifo Working Paper, No. 859, Center for Economic Studies and Ifo Institute (CESifo) (2003)
11. Nowak, M.A.: Five rules for the evolution of cooperation. Science **314**(5805), 1560–1563 (2006)
12. Latané, B., Darley, J.M.: The Unresponsive Bystander: Why Doesn't He Help?. Appleton-Century-Crofts, New York (1970)
13. Latané, B., Rodin, J.: A lady in distress: inhibiting effects of friends and strangers on bystander intervention. J. Exp. Soc. Psychol. **5**(2), 189–202 (1969)
14. Howard, W., Crano, W.D.: Effects of sex, conversation, location, and size of observer group on bystander intervention in a high risk situation. Sociometry **37**(4), 491–507 (1974)
15. Harari, H., Harari, O., White, R.V.: The reaction to rape by American male bystanders. J. Soc. Psychol. **125**(5), 653–658 (1985)
16. Bierhoff, H.W., Klein, R., Kramp, P.: Evidence for the altruistic personality from data on accident research. J. Pers. **59**(2), 263–280 (1991)
17. Levine, M., Prosser, A., Evans, D., Reicher, S.: Identity and emergency intervention: how social group membership and inclusiveness of group boundaries shape helping behavior. Pers. Soc. Psychol. Bull. **31**(4), 443–453 (2005)
18. Fischer, P., Greitemeyer, T., Pollozek, F., Frey, D.: The unresponsive bystander: are bystanders more responsive in dangerous emergencies? Eur. J. Soc. Psychol. **36**(2), 267–278 (2006)
19. Karakashian, L.M., Walter, M.I., Christopher, A.N., Lucas, T.: Fear of negative evaluation affects helping behavior: the bystander effect revisited. North Am. J. Psychol. **8**(1), 13–32 (2006)
20. Martin, K.K., North, A.C.: Diffusion of responsibility on social networking sites. Comput. Hum. Behav. **44**, 124–131 (2015)
21. Bommel, M.V., Prooijen, J.-W.V., Elffers, H., Lange, P.A.M.V.: Be aware to care: public self-awareness leads to a reversal of the bystander effect. J. Exp. Soc. Psychol. **48**(4), 926–930 (2012)
22. Voelpel, S.C., Eckhoff, R.A., Förster, J.: David against Goliath? group size and bystander effects in virtual knowledge sharing. Hum. Relat. **61**(2), 271–295 (2008)
23. Lewis, C.E., Thompson, L.F., Wuensch, K.L., Grossnickle, W.F., Cope, J.G.: The impact of recipient list size and priority signs on electronic helping behavior. Comput. Hum. Behav. **20**(5), 633–644 (2004)
24. Barron, G., Yechiam, E.: Private e-mail requests and the diffusion of responsibility. Comput. Hum. Behav. **18**(5), 507–520 (2002)
25. Markey, P.M.: Bystander intervention in computer-mediated communication. Comput. Hum. Behav. **16**(2), 183–188 (2000)
26. Postmes, T., Spears, R.: Deindividuation and antinormative behavior: a meta-analysis. Psychol. Bull. **123**(3), 238–259 (1998)
27. Clark, R.D., Word, L.E.: Why don't bystanders help? because of ambiguity? J. Pers. Soc. Psychol. **24**(3), 392–400 (1972)
28. Latané, B., Dabbs, J.M.: Sex, group size and helping in three cities. Sociometry **38**(2), 180–194 (1975)
29. Latané, B., Williams, K., Harkins, S.: Many hands make light the work: the causes and consequences of social loafing. J. Pers. Soc. Psychol. **37**(6), 822–832 (1979)

30. Greitemeyer, T., Osswald, S., Fischer, P., Frey, D.: Civil courage: implicit theories, related concepts, and measurement. J. Positive Psychol. **2**(2), 115–119 (2007)
31. Asch, S.E., Guetzkow, H.: Effects of group pressure upon the modification and distortion of judgments. Groups, Leadership, and Men, 222–236 (1951)
32. Darley, J.M., Latane, B.: Bystander intervention in emergencies: diffusion of responsibility. J. Pers. Soc. Psychol. **8**(4), 377–383 (1968)
33. Banyard, V.L.: Measurement and correlates of prosocial bystander behavior: the case of interpersonal violence. Violence Vict. **23**(1), 83–97 (2008). Please supply the page range for Reference [44] and also volume for ref. [31]
34. Harris, V.A., Robinson, C.E.: Bystander intervention: group size and victim status. Bull. Psychon. Soc. **2**(1), 8–10 (1973)
35. Dicke-Bohmann, A.K., Cox, C.B.: Predicting successful responses to emergencies: the emergency responsiveness scale. Current Psychol. **36**(2), 209–216 (2017)
36. Cramer, R.E., Mcmaster, M.R., Bartell, P.A., Dragna, M.: Subject competence and minimization of the bystander effect. J. Appl. Soc. Psychol. **18**(13), 1133–1148 (1988)
37. Fischer, P., Krueger, J.I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M., Kainbacher, M.: The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. Psychol. Bull. **137**(4), 517–537 (2011)
38. Blair, C.A., Foster, T.L., Wuensch, K.L.: Electronic helping behavior: the virtual presence of others makes a difference. Basic Appl. Soc. Psychol. **27**(2), 171–178 (2005)
39. Spears, R., Lea, M., Lee, S.: De-individuation and group polarization in computer-mediated communication. Br. J. Soc. Psychol. **29**(2), 121–134 (1990)
40. Vakhitova, Z.I., Reynald, D.M.: Australian internet users and guardianship against cyber abuse: an empirical analysis. Int. J. Cyber Criminol. **8**(2), 165–171 (2014)
41. Mehrabian, A.: Relations among personality scales of aggression, violence, and empathy: validational evidence bearing on the risk of eruptive violence scale. Aggressive Behav. **23**(6), 433–445 (1997)
42. Leary, M.R.: A brief version of the fear of negative evaluation scale. Pers. Soc. Psychol. Bull. **9**(3), 371–375 (1983)
43. De, C.D., Lange, P.A.M.V.: Why prosocials exhibit greater cooperation than proselfs: the roles of social responsibility and reciprocity. Eur. J. Pers. **15**, S5–S18 (2001)
44. Haghighat, R.: The development of the brief social desirability scale (BSDS). Eur. J. Psychol. 3(4) (2007)
45. Chiorri, C., Bracco, F., Piccinno, T., Modafferi, C., Battini, V.: Psychometric properties of a revised version of the ten item personality inventory. Eur. J. Psychol. Assess. **31**, 109–119 (2015)
46. Spielberger, C.D., Gorsuch, R.L., Lushene, R.E.: STAI Manual for the Stait-Trait Anxiety Inventory ("self-evaluation Questionnaire") (1970)
47. Luszczynska, A., Scholz, U., Schwarzer, R.: The general self-efficacy scale: multicultural validation studies. J. Psychol. **139**(5), 439–457 (2005)
48. Prezza, M., Pacilli, M.G., Barbaranelli, C., Zampatti, E.: The MTSOCS: a multidimensional sense of community scale for local communities. J. Commun. Psychol. **37**(3), 305–326 (2009)
49. Mattick, R.P., Clarke, J.C.: Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav. Res. Therap. **36**(4), 455–470 (1998)

50. Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M., Henry, H., White, J.-S.S.: Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. **24**(3), 127–135 (2009)
51. Chow, S.-C., Wang, H., Shao, J.: Sample Size Calculations in Clinical Research. Chapman and Hall/CRC, Boca Raton (2007)
52. Postmes, T., Spears, R., Lea, M.: Breaching or building social boundaries? SIDE-effects of computer-mediated communication. Commun. Res. **25**(6), 689–715 (1998)
53. Mestre, M.V., Samper, P., Frías, M.D., Tur, A.M.: Are women more empathetic than men? a longitudinal study in adolescence. Spanish J. Psychol. **12**(1), 76–83 (2009)

Collective Intelligence and the Geopolitical Crossroads in Central-Eastern Europe. Review of the Selected Research Methods

Rafał Olszowski[1(✉)] and Marcin Chmielowski[2]

[1] AGH University of Science and Technology, Kraków, Poland
rafal.olszowski@omni.pl
[2] ASBIRO Academy, Łódź, Poland

Abstract. Authors are the researchers working on the project "Collective intelligence on the Internet: applications in the public sphere, research methods and models of civic participation". This paper contains the first remarks of this research project. First, the authors show selected social problems of Central and Eastern Europe, particularly the Polish-Ukrainian relations. The phenomenon of the Ukrainians who are working, studying and settling in Poland is a context. Next, a present state of the research on collective intelligence (CI) on the internet and its compatibility with deliberative democracy theory is briefly presented. The main part of the paper is a critical review of the CI research methods in terms of their adequacy to the study of the public sphere initiatives concerning the selected social problems. The two separate problems come together under a single goal of enabling observation of Polish-Ukrainian interactions as manifestations of collective intelligence, and furthermore, the future use of collected results to stimulate the growth of bridge capital between Poles and Ukrainians. We focus on the possibilities of application of the selected methods in the social field we are interested in, as well as necessary changes and modifications. The selected methods are online deliberation and analytics with the use of IBIS model, using the indicators measuring social phenomena – CI Potential Index and UPVoCI scale, and finally a qualitative analysis with the use of CI Genome framework.

Keywords: Collective intelligence · Web 2.0 · Research methods · Migrations · Eastern Europe

1 Introduction

Collective Intelligence (CI) is an idea found in multiple disciplines such as computer and web science, decision-making, economics, political science, sociology, psychology and even biology [39]. The concept of CI is currently closely connected with the

Type of paper: Exploratory.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 168–186, 2019.
https://doi.org/10.1007/978-3-030-34770-3_13

development of ICT networks and entails utilization of interactions between users in online communities to create "added value" based on collective work and cooperation. The authors of this paper are currently involved in the research project "Collective intelligence on the Internet: applications in the public sphere, research methods and models of civic participation", which aim is a critical review of the CI research methods to examine their suitability for application in the public sphere, that is, among others: in citizen e-participation, in government crowdsourcing, and in direct e-democracy; practical analysis of selected projects related to the use of CI in the public sphere, and finally a theoretical reflection on the subject of existing and possible to implement models of civic participation specific for CI initiatives implemented in the public sphere and using the internet.

The geographical area, on which we consider these issues, is mainly Central and Eastern Europe. This region of the world is currently affected by several social problems - both the new ones, resulting from the contemporary geopolitical conditions: the political crisis in Ukraine and migration connected with it, as well as old ones, derived from the legacy of post-communism: low trust in political institutions and the loosening of social bonds, weakened norms of consensus and agreement, low quality of the public debate [17].

Our widely scoped goal is to answer several questions about the possible impact of the internet-based CI initiatives on the Central and Eastern European public sphere. We consider: whether and how the organization models specific to CI initiatives can help in solving emerging socio-political challenges of this region, among them: problems of social integration, migrations, and citizen participation? What new areas of the public sphere, as well as new methods of deliberation and collective action, can be reached through this?

However, the works described in this paper, focus on a more specific goal: collective intelligence and the geopolitical situation in the Central and Eastern Europe – migrations, conflicts, and social unrest. We assume that a certain type of CI projects implemented in the public sphere can ease the tensions related to these problems: reduce distrust, reduce polarization, and build communities (premises for this hypothesis can be found, among others in: [7, 44, 46]). Therefore we are looking for research methods that could confirm or deny this hypothesis. To be prepared for answering to these research challenges, we went through a critical examination of currently existing methods of researching and analyzing CI on the internet. The scientific literature on this subject has gone through an enormous growth in the last years. Nevertheless, each academic discipline takes its own research approach to the issue of CI and competing terminologies are often used to describe the same questions. For this reason, we were forced to focus our interest on a few selected analytical approaches for more complex analysis. During the process of the final selection, we had an additional goal: identifying the most adequate methods of research and analyzing of CI to single out the most important factors which would enable creation and development of online communities responsible for determining the challenges in the Central-Eastern European region.

2 Selected Social Problems of the Central-Eastern Europe: Polish-Ukrainian Relations

The situation of geopolitical crossroads, as highlighted in the title, is an apt metaphor for the current state of Polish-Ukrainian relations. They are burdened by the aforementioned conflicts in the past [15], beginning with the common and not exactly harmonious coexistence during the Polish–Lithuanian Commonwealth period [73], the competition during the Galician autonomy [38], the ethnic policy in interwar Poland [12], and the tragedy of genocide in Volhynia [54]. The context of contemporary Polish-Ukrainian relations, influencing both ordinary citizens and governments, are the war between Russia and Ukraine, with Poland supporting the latter side, and Ukrainian migration into Poland. The scale of immigration in comparison with other EU countries deserves an emphasis. Poland accepted ca. 800,000 Ukrainian immigrants – considering the total population, it is a significant number. Even considering the inconstant or oscillatory character of Ukrainian migration into Poland, the potential of permanent residence will remain, causing a lasting shift of ethnic composition in Poland. With the beginning of the Russo-Ukrainian war in 2014, the Ukrainian migration into Poland changed its character. While the percentage of immigrants from western parts of Ukraine, the former Galicia, and Eastern Borderlands, dropped, the number of migrants from East Ukraine increased, including the areas affected by war. This turn of events had a certain influence on Polish-Ukrainian relations, since East Ukraine is culturally more distant from Poland than western Ukrainian territories. However, there is a lack of quantitative and qualitative research which would help to expand this hypothesis [25]. Besides the history, there is another burden lasting on relations between Ukraine and Poland as countries, and Poles and Ukrainians as citizens. The discrepancy is also caused by sociological factors which can be analyzed based on social capital.

In comparison with Western Europe, levels of social capital in the former Soviet Bloc countries indicate a social deficit [33]. This state of facts also applies to Poland and Ukraine. Historical animosities lead to accumulation of cultural capital in the form of internal bonding and not bridging capital [5], a factor which weakens interaction. For this study, it is important to presume that in the relations between nations low levels of this type of social capital lessen the possibility of added value resulting from the emergence of collective intelligence in Polish-Ukrainian relations. Said relations are strained not by language or cultural barrier, or historical conflicts, but by a factor deriving from the cultural history of the region.

The present situation, an accelerated return to a multiethnic society which vanished in Poland after World War Two, poses important challenges both for Poles and Ukrainians living in Poland. It is obvious that both groups have their distinct concerns and genuine problems related to coexistence. The problems faced by Ukrainians were summarized to a certain degree in the expert talk “Current problems of Ukrainian immigrants in Poland” [3]. The report divides Ukrainians staying in Poland into three groups: the younger generation, those who seek international asylum, and temporary migrants. Although their problems and needs are divergent, it is possible to extrapolate a set of core issues which affect all groups. According to experts [3], they are:

1. In spite of obvious similarities, there are still linguistic difficulties related to poor grasp of Polish language among Ukrainians and, vice versa, low proficiency in understanding Ukrainian among Poles;
2. Integration problems resulting from the lack of willingness or possibilities of integrating both for Poles with Ukrainians and Ukrainians with Poles. Due to living and job conditions, actual isolation of both groups happens too often;
3. Adaptive difficulties related to migration to a new country, new surroundings, and other new conditions, the problem with solving "issues of daily life".

Another important problem is the image of Ukrainians and Ukraine in Polish internet discussions. The analysis of texts published in the Polish internet reveals a large percentage of negative or strongly negative opinions regarding Ukrainians. The percentage of such entries on the Facebook social portal is as high as 50%, a figure similar to comments published on internet portals. For internet fora said percentage amounts ca. 30%. The leading position belongs to Twitter, however, where ca. 60% of content related to Ukrainians and Ukraine shows them in a negative or strongly negative light [64].

This situation is an obvious outcome of unresolved issues from the past. It is necessary to take into consideration also the reality of information warfare and pro-Russian activity in the Polish-speaking Internet. From the standpoint of the Russian interests, stoking conflicts between Poles and Ukrainians gives many advantages.

The Internet in its Web 2.0 incarnation, being an instrument of participation and interaction between users bonding through social media, forms a space where creative intelligence can emerge with an effect more significant than in other spaces. This becomes possible mainly through the effect of scale [29]. It will not manifest itself, however, and the full spectrum of positive and pro-growth results of collective intelligence will not become apparent if the bridging capital remains scarce. The relation between bridging capital and Web 2.0 can obviously not be reduced to a simple reaction of causality, but it is most likely a case of feedback with the growth of one variable leading to another one growing - with one important restriction. Web 2.0 gives exceptional possibilities in revealing collective intelligence, but also in creating information bubbles which naturally increase the level of integral bonding capital [10]. The Web 2.0 projects with significant participation of Poles and Ukrainian migrants will emerge only when all dangers coming from overproduction of internal bonding capital become minimized and the level of bridging capital grows.

3 Analysis of the CI Research Methods in Terms of Their Adequacy to the Study of the Public Sphere Deliberation

As a part of the work carried out so far in the project "Collective intelligence on the Internet: applications in the public sphere, research methods and models of civic participation", we have made a broad review of the scientific literature on this issue.

Collective intelligence is a sphere where the collective creates possible solutions to its problems and evaluates them; assimilates, groups, explains and compares knowledge and ideas; finally – collectively makes decisions. The concept of CI is currently closely connected with the development of Web 2.0 and entails utilization of interaction between users in online communities to create "added value" based on collective work and cooperation, structuring of the debate, synthesis of knowledge, visualization of arguments, open innovation, reduction in the number of errors, and collective decision making. CI is found in multiple disciplines such as computer and web science, decision-making, economics, sociology, political science, psychology and even biology.

There are numerous definitions of CI in the reviewed scientific literature. One of the pioneers in the analysis of CI on the internet, Pierre Levy stated that CI is *a form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of the skills* [35]. Heylighen claimed, that CI is the ability of a group to *find more or better solutions than the whole of all solutions that would be found by its members working individually*. If there is no collective cooperation in a social structure, this structure has limited capability for solving a certain pool of problems: each person seeks solutions independently, and thus, there is neither positive nor negative interaction. He also noted, recalling an analogy with the life of insects, that the result of CI's work is often not a single solution, but a kind of shared memory, which he called the collective mental map (CMM) [23]. Malone proposed a simplifying definition, describing CI as *groups of individuals acting collectively in ways that seem intelligent*, at the same time highlighting goals, motivation and incentives, as well as cognitive processes related to this kind of work. Biological sciences evoke the term "swarm intelligence" as self-organized behaviour emerging from swarms of social insects, in which *accumulated interactions lead to an irreversible outcome* [18]. Moreover, while Woolley defined CI as *the general ability of a group to perform a wide variety of tasks* and proposed a general, single CI factor highly correlated with the level of cognitive diversity within a group [72], several psychologists challenge this thesis connecting CI more with an average individual intelligence level of the group and describing CI level rather as multi-dimensional vector than a single factor [6, 16, 57].

Collective intelligence frequently manifests itself when cooperation, competition or mutual observation gives rise to totally new solutions to the problems that the community faces or leads to an increase in the ability to solve complex problems. Empirical studies and theoretic simulations prove that a diverse collective can achieve better results in problem solving than a narrow group of specialists [24, 40, 60].

This knowledge has still not been efficiently used in relation to the public sphere. The ambition of numerous researchers and policy-makers is dissemination of CI projects to make citizens engaged as "co-decision-makers" on political matters. Moreover, they notice that since the methods of work organization and business management have undergone a real revolution in recent years (towards participative management, agile methods, turquoise management, etc.), methods of civic participation and political institutions are still deeply embedded in the 19th century (see: [47, 70]). They look forward for a new participative model of social life, whose characteristics include e.g. a relatively low entry barrier for civic involvement, strong support for creativity and sharing of one's own creations with others. Therefore technology can contribute to the generation of both new areas of the public sphere and new ways of deliberation and collective action. This possibility becomes particularly valuable if we consider the erosion of the traditional public sphere, as observed by Jürgen Habermas [21].

Theories describing the public sphere from the social sciences point of view emphasize the important role of debate in shaping civic attitudes, involvement in community life and creating social bonds. Among modern theories, the concept of deliberative politics derived from Habermas' thought is of particular importance - not only because it is very influential, but also because it emphasizes the manner of conducting the debate, which we consider to be significant for the problem we are investigating, i.e. Polish-Ukrainian relations on the Internet. Deliberation in this sense is: a form of communication that requires that people discuss a common problem and reach consensus on how to solve this problem, involving expression of reasoned opinion, judicious argument, equal participation, independence of judgment, critical listening, and earnest decision making. [58]. An important feature of deliberation from the perspective of our research is its rationalism: in a deliberation n*o force except that of the better argument is exercised* [14]. Contemporary theorists, however, redeveloped Habermas' thought, allowing in communication, apart from purely rational statements, also the emotions and interests of the participants [43], and emphasizing the role of external experts in assessing and moderating the debate [42].

It is worth noting that the theory formulated by Habermas has features of normative theory, and thus indicating an ideal communication situation unattainable in real-life interpersonal relationships. In our opinion, it would be appropriate for our research subject to reinterpret the theory of deliberation as prospective, and thus capable of changing social reality. The need to make deliberation a practical approach is also indicated by contemporary social scientists, who continue Habermas' thought [13]. From the point of view of our work, this is a significant shift in emphasis.

The use of CI research methods in conjunction with the analysis of communication as understood by deliberative theories will allow us to assess their usefulness for studying Polish-Ukrainian online communities as well as the potential of improvement of their work, which we understand as building bridging capital between Poles and Ukrainians. Therefore, our goal is not a simple monitoring of debates, but examining to what extent collective intelligence compatible with the deliberative communication is the dominant type of debate in the area we are interested in, or maybe it is a minority.

Although some researchers consider deliberative communication as an obvious way of implementing CI in the area of public affairs [32], we believe that it is not the only possible and necessary model to translate CI into social sciences language. For the description of some manifestations of CI, the agonistic model seems to be more adequate, in which competition, disputes and individual respect come to the fore [2, 8, 45], to the other projects more proper is an utilitarian model having features of free-market spontaneous order, evaluating primarily the usefulness of the debate results [1, 63], or a model in which collective wisdom is gradually accumulated and perfected through repetition [31, 46]. Nevertheless, the deliberative model is a determinant of expectations for our topic.

Reviewing the available research methods, used so far for capturing different aspects and components of CI, we took into account, among others: user behavior analytics, logical structure of arguments and deliberation analytics [26, 27, 29, 26–27], CI Genome – a framework designed to explain and classify the structures, processes, goals, and incentives emerging in various CI initiatives [40], measuring group performance on a wide variety of tasks [71, 72], semantic interoperability, semantic navigation, collective categorization and self-referential collective intelligence [34], computational models for simulating and measuring CI [60], issues of public decision-making, technologies of expertise, collaborative leadership [47], monitoring CI emergence using complex evaluation index – like Collective Intelligence Potential Index (CIPI) [55], understanding and measuring the user-perceived values of CI (UPVoCI) [69], studying financial markets as an ideal laboratory for CI [36], studying collective behavior of the animals and swarm intelligence [52], exploring the intersection of artificial and collective intelligence [41, 65], studying social influence, cognitive diversity, biases, decision frameworks and other important CI factors [9, 10, 24], analyzing of the behavior of online community participants using the Social Identity model of Deindividuation Effects (SIDE model) [62, 66], analyzing CI in politics as democratic reason [32], investigating relations between the strength of social relationships and distrust [44].

To clarify the area of our future research and choose proper research methods in relation to it, we have ranged the features of deliberation described in the theory of Habermas and his followers. Reviewing these deliberation features we were taking into account their objective verifiability (from fully objectively observable and independent of discretionary judgments, to the most dependent on subjective assessment). Then, we selected the most appropriate methods to confirm or deny their occurrence in the online communities. In addition, as we will show in the following sections of this paper, we have indicated technical possibilities of their application, indicated their strengths and weaknesses, as well as proposed changes and modifications matching the given method to the subject of the study.

The first group of identified properties of deliberation focus on a pure communication rationality, studied at a behavioral level, as objectively and non-discretionally as possible, including: active content assimilation and active evaluation of issues presented in the discussion; making consistent, non-biased arguments; equality of participation (not being dominated by individuals); conducting debate motivated by rational analysis of arguments, not prejudices or sympathies; ordered structure of the contributions [see more: 14, 21, 22, 56, 58]. Analytical software, dedicated to gather and process the data coming from a particular debate, seems to be the most adequate tool to examine these features.

The second group consists of qualitative features of deliberation, to which valuation in a graded scale of points is possible. Here it should be taken into account: the substantive value of the arguments [14, 22]; deliberative stance - that is, opening to a change of position and treating others as equals in mutual exchange of reasons [48], respect and culture of discussion [56]; sensitivity of community members [14]; inclusiveness or conditional exclusivity; sense of mission and shared responsibility [58]; civic engagement; independence of opinion [21, 22]; the opportunity to use expert knowledge [42]; decentralization [22]. The evaluation of these features would be to some extent possible using software, but we propose as the basic method an assessment with composite indexes and scales used in social sciences, used by the experts, or in the process of self evaluation, which is certainly more discretionary than machine assessment.

The last group includes the most subjective, qualitative deliberation assessments, not gradable, yet adequate to describe the incentives, interests and emotions related with deliberation process. In recent years, there has been an evolution of the position of deliberativists on these issues: many of them have accepted the opinion that the inclusion of emotions and self-interest is not only acceptable, but even necessary [42, 43]. This should be a way of analyzing the motivations of the participants of a debate, as well as their individual interests (disclosed or hidden), their emotions, cognitive styles, etc. Because these features are not subject to grading, but only a qualitative description, a general framework to prepare such an account is selected.

Below we present our selection of the methods to conduct research on all the above mentioned levels, as well as the tools that will allow to identify in Polish-Ukrainian debates an emergence of collective intelligence, and within it a presence of the specific features coming from the deliberation theory. We hope this multi-level analysis will enable an emergence of collective intelligence in our field of interest and a search for reinterpretation of the deliberative theory as a prospective one, i.e. one able to change the realities of social life.

3.1 Communication Rationality: Online CI Deliberation Analytics Based on Issue-Based Information System

The first selected method needs the use of analytical software, dedicated to process the data coming from a particular debate and gathered on the server, to identify communication rationality, studied at a behavioral level. This approach requires using dedicated IT environment - the analytics server that would calculate metrics of the online deliberation, measuring several aspects of users behavior, identify meaningful patterns in online deliberation and map these patterns to personalized attention-mediation recommendations for the deliberation participants, finally apply algorithms and statistical analysis to detect meaningful anomalies from those patterns. Using real-time online analytics would allow to measure the quality of the collective intelligence dynamics and gain knowledge, that can make the collaborative process significantly more effective.

One of the leading researchers in this subject, Mark Klein, based on his long-term experience and knowledge gained during field research [26–29] proposed a system, that *supports an iterative deliberation cycle. Participants interact by posting in a "deliberation map" (a tree structure made up of interleaved questions, answers and arguments). (…) Crowd-based idea filtering algorithms are then used to identify the most promising solution ideas generated by the crowd.* Theoretical background of this approach is based on a scheme developed by H. Rittel in the 1970s, called the "Issue-Based Information System" (IBIS) [49–51]. The core of this method is extracting from the deliberation such elements as: issues, ideas, arguments for and against an idea, decisions, and then advanced deliberation analytics, that *data-mine the traces of the crowd's activity and generate customized metrics, alerts and reports to help the participants, moderators and customers of the deliberation have a much clearer sense of where the deliberation is as well as where and how they can contribute best* [29]. Studies conducted with the use of this method were, among others: large-scale argumentation on the use of biofuels in Italy, conducted in the University of Naples, debate on the controversial questions about possible changes to Italy's election laws [26], Intel conducted deliberation on the possible use of "open computing" [41]. All of them showed promising results.

Looking for a technical solution to conduct this kind of analytics we considered using MIT-created system Deliberatorium or Catalyst Deliberation Analytics Server, which was created in the research project under 7th EU Framework Programme. Another option being considered is the development of our own software based on IBIS method. In this kind of software deliberation analytics take two forms: metrics and alerts. *Metrics provide summary overviews of the status of the deliberation, highlighting phenomena (such as controversy hot spots or balkanization) that would be difficult to identify simply by browsing the map on a post-by-post basis. Alerts provide user-specific notifications of what elements of the deliberation map a participant should consider in order to maximize their contribution* [28].

Using such a system to analyze large-scale online deliberation taking place in the Eastern European public sphere would allow for gathering a large amount of knowledge about the way the joined Polish-Ukrainian groups debate, identify hot spots, level of involvement in the discussion, useful contributions, and other key issues. Among the metrics already available in the existing software [28], we found the following as the most important for our target group:

- Maturity - indication how mature the discussion for an issue is. Estimated by gathering statistics on the topology of the branch (e.g. in terms of breadth and depth of issues, ideas and arguments) for the problem;
- Controversy - returning the controversy score for every post in a given branch, based on the ratings the posts received;
- Inequality - measuring to what extent the community support is unequal for the ideas under an issue;
- Clusters - returning clusters of posts that tend to be liked, rated, viewed together.
- Support consistency - measuring to what extent an idea's average rating is consistent with the ratings for the underlying arguments;
- Social graph - returning a graph showing which users have interacted (rated, commented on, responded to, or edited posts created by the other user);
- Groupthink - estimating the level of groupthink in the deliberation for a given issue. Groupthink occurs when a crowd converges prematurely on one solution, without giving adequate attention to competing ideas. It would be interesting to show whether there were competitive groups in the debate, characterized by a groupthink in their group.

We also considered extending the list of the metrics, that could be used to help us to draw maximum research benefits from the deliberation. Therefore we reviewed the most promising proposals of other research teams working on CI analysis. Below we listed some opportunities in this area:

- User efficiency - Calculated by dividing the number of a user's contributions by his total views of others' contributions. Metric proposed in the Gurkan et al. research paper "Mediating debate through on-line large-scale argumentation: Evidence from the field" [20];

- Modified "Groupthink" metric using the SIDE model in examining relationships within and between identified groups [62, 66]. It should be tested whether the processes indicated by this model will occur: whether Poles and Ukrainians recognizing each other will deepen rather than eliminate existing differences between them and whether there will be an effect of increased orientation towards their own national group.
- Social interaction strength, level of confidence, the density of social relations, scope of distrust, the value of interdependence – these indicators were used by the research team from the Politecnico di Bari (I. De Vincenzo, G.F. Massari, I. Giannoccaro, G. Carbone, P. Grigolini) in their simulation model of CI coming from statistical physics using the Ising-Glauber dynamics [44, 67, 68];
- Collective Mental Map (CMM) is a concept introduced by Heylighen, to register and store competitive discussion threads and alternative solutions in the form of a shared cognitive system [23]. Having sufficient data on the preferences expressed by individual users during the discussion, we could attempt to visualize deliberation in this form.

The undoubted advantage of this approach is the objectivity of the recorded data. However, a certain limitation is the fact that we only record numerical values of the deliberation, rather than assessing its substantive properties. The greatest disadvantage of this solution results from the fact that conducting and evaluating a deliberation requires the application of one specific kind of software. There are two options to solve this problem: first of them the use of analytics server to integrate several external communication platforms which form the framework for discussion about social issues. However, the use of analytics with different external platforms may become a source of numerous difficulties – from human and organizational reluctance to technical obstacles. Nevertheless, conjunction of external platforms (which could become, for example, the stage for a discussion about expenses in participatory budgeting involving both Poles and Ukrainians) would solve the question of recruiting participants and allow the achievement of scale effect.

The second possibility is a stand-alone organization of the entire deliberation process. This requires determination of subjects and rules of discussion, as well as the motivation of potential participants. This approach seems necessary at least in the initial phase of the work planned by us due to the necessity of test debates, required for a more accurate choice of indicators and determining the best possible conditions for the deliberation which would enable us to face diagnosed social problems.

Deliberation carried out in a laboratory environment may also help to identify the factors important for analyzing platforms to which we do not have access from the analytical server side (they must be studied using other methods described below).

3.2 Qualitative Features of Deliberation Valuated in a Graded Scale with the Use of Social Indicators: CI Potential Index, UPVoCI Scale

Another approach, which we found suitable for indicating qualitative features of deliberation which could be graded on a scale, needs adapting the typical social sciences method: an assessment using a complex indicator consisting of several weighted components. Composite indexes and scales used in social sciences are models, mainly quantitative and presented in numerical format, that attempt to measure complex social phenomena by selecting, combining and weightening several individual indicators that, individually, would not adequately describe the subject in question [4].

As the most promising solution of this type we identified the CI Potential Index (CIPI), that was developed in the research project "Social Technologies for Development Collective Intelligence in Networked Society" under the direction of A. Skaržauskienė [31]. Calculating the CIPI allows monitoring the CI emergence in online communities, virtual platforms, etc.

While creating this method, several major theoretical concepts concerning CI research and description were reviewed (among them: [9–11, 19, 24, 30, 34, 37, 53, 59, 70, 72]), and hypotheses regarding the influence of individual factors on CI potential in online communities were formulated. These hypotheses were verified during quantitative research and analyzed in qualitative research. On this basis, criteria, dimensions, interpretations, components, and indicators were determined.

The authors of this method elaborated several sub-indexes, covering various CI dimensions and created different components to measure each dimension. The main sub-indexes and some of their components are:

1. CI Capacity index, whose main dimensions are:
   a. Capacity for creativity; components: degree of diversity in the source of ideas and degree of diversity in engagement forms;
   b. Capacity for Aggregating Knowledge; components: degree of interdependence, degree of adequate supply of critical mass ("swarm effect").
   c. Capacity for decision making and problem-solving; components: degree of decentralization, efficiency of problem-solving, degree of independence.
2. CI Emergence Index, whose main dimensions are:
   a. Potential for self-organization; components: adequacy in form of self-organization to community task, degree of development of transparent structure and culture
   b. Intensity of Emergence; components: degree of development of new qualities in form of ideas, activities, structured opinions, competencies, etc. based on distributed memory system (Web intelligence).
   c. Potential for adaptivity; components: degree of development of improvements and learning processes within the community, development of life-long learning.
3. CI Maturity Index, whose main dimensions are:
   a. Maturity of Social Impact (behavioural); components: degree of civic engagement, degree of sustainability.
   b. Maturity of Social Motivation (psychological); components: level of maturity of social motivation of a community, level of social sensitivity of community members.
   c. Maturity of Social Orientation (cognitive); components: level of maturity of reaction to social issues, degree of diversity in cooperating partners and financing, level of maturity of generated content.

The main difference, compared to the earlier described method, is that using CIPI would allow analyzing all observed online communities, regardless of whether they are integrated into our analytic environment or not, without interfering in any way in their functioning, and even without revealing that they are subjected to research. Lack of obligation to integrate our server with the external platforms opens a large space for research and observation. Therefore we can analyze various aspects of the activities of online communities debating public issues and consisting of mixed multinational groups. This is important, because due to the subject of our research, we intend to monitor the activity of grassroots civic groups gathering many ethnic groups (including Poles and Ukrainians living in Poland, people from excluded societies, etc.) and the place of public debate organized, for example, by local governments in which they will take these groups.

However, the disadvantage of this method, like of all the composite social indexes, is it's prone to subjectivity and simplification of the reality, and sometimes errors and misrepresentations. It is a model, imitating reality based on subjective opinions of evaluators or self-assessment of community participants, necessarily numerous decisions (such as the choice and the weighting of indicators) rather than hard data showing the behavior of the collective.

What kind of changes and modifications did we find as worthy to consider? First of all, we draw attention to the fact that the authors of this method in the final version of CIPI did not use the CI Maturity sub-index due to difficulties in obtaining data. Meanwhile, this indicator, if it could be applied, could have a significant impact on qualitative features of deliberation, so we are planning to use it. In addition, the CIPI components and related evaluation questions should be reviewed: not all of them can be relevant to our topic, and the state of CI research since CIPI creation has gone forward. We consider as particularly important:

- Reinterpretation of CI Social Maturity Index. We recognize the importance of properly choosing reference points to illustrate complex relations, such as the relations between the Ukrainians living in Poland and native Poles, which could provide important knowledge before starting research in a non-designed Internet environment. The key questions chosen by us to demonstrate those relations are: the degree of mutual trust – defined as the conviction that the partner in a relation will act as we expect them to do and in respect to our interests, and the capability of long-term cooperation – defined as the ability to make compromises, the reciprocation of the partner in a relation, and then repetition of earlier partner moves by the players.
- Adding a component related to the user-perceived value of CI. Weng, Yang, and Hsiao [72] conducted very interesting research in this area. These authors proposed their own research tool in the form of measurement scale UPVoCI, thanks to which they were able to study the benefits of participants in enhancing interpersonal relationship, enhancing personal reputation, improving cooperative environment, etc.
- Verification of the possibility of using work from the research team of Politecnico di Bari (see: [44, 67, 68]). It may be particularly interesting to include to CIPI social interaction strength and social density indicators in the relational connection with the scope of distrust. It turns out that in some configuration of these variables the increase distrust level can bring salutary effects to CI efficiency. This statement needs to be proved or falsified.

- More nuanced measure for the cognitive diversity component: as demonstrated by Woolley [71] too high level of diversity may lead to a decrease in the level of CI: inverted U-shaped relationship between cognitive diversity and CI.
- The possibility of using the newly created Extended Turing Test (ETT) [61] as a tool for measuring the level of self-organization in groups. ETT is a new approach proposed by T. Szuba to measure the level of self-organization of the social structure and its effect on the collective intelligence of the group;

3.3 A Subjective Qualitative Assessment: CI Genome Framework and Its Modifications

As a complementary method to those described above, but allowing to capture the most subjective, qualitative deliberation assessments, we selected a qualitative analysis with the use of CI Genome framework. The CI Genome is a widely accepted concept presented by Thomas Malone, Robert Laubacher and Chrysanthos Dellacroas [40] from the MIT Center for Collective Intelligence. It is a framework designed to explain and classify the structures, processes, goals, and incentives emerging in various CI initiatives. The method is based on the development of the taxonomy of CI building blocks (referred to as CI "genes") and the identification of the set of behavioral patterns under specific projects described on their basis. It has highlighted four elements which predict the formation of the principles of Collective Intelligence: aims, incentives, structure and participants and along lines of queries: who contributes and undertakes the activity? Why do they perform a specific task? What motivates people to take part? How do they operate? Research by the MIT involved such analysis of about 250 projects [40]. As Wise [70] summarizes: Malone et al. *propose the existence of three motivators or genes which serve to explain the reason why individuals contribute to CI systems. The motivators are Money, Love Or Glory. Where money is a traditional extrinsic motivator and the most commonly occurring one in a hierarchical gene, Love and Glory represent more intrinsic motivators which reflect an individual's enjoyment of doing an activity or their desire to boast about their achievements. (…) What question can be reduced to two simple genes: to create and to decide. (…) The How genes classify CI systems along two lines: collection and collaboration.*

Malone also called for further investigation into the application of these genomes and an expansion of the genome framework to include emerging organizations. This method, which was originally used primarily to study commercial projects, has been therefore expanded by a team led by S. Wise [70] for the purpose of analyzing the public sphere initiatives through the addition of new genes specific for the projects related to public goals and analysis of over 120 more examples.

In our research, we take into account the use of this method to describe phenomena and processes that are not directly transferable to numerical indicators. Many situations, especially those related to the motivations of individuals and complex organizational processes in groups cannot be measured with the use of parameters, but are described in context. However, we are aware that the CI Genome Framework method is controversial because of its age. Therefore, we realise that it may not stand the test of time anymore. At the same time, it is attractive to us, because of its flexibility and ability to describe the incentives, interests and emotions related with deliberation process. With the reservation made, it can be treated as an additional layer of analysis and perhaps the necessary supplementation of our imaginarium with a qualitative rather than a quantitative tool.

However, it cannot be assumed that by using this tool in its original form, we will be able to accurately capture CI manifestations in Polish-Ukrainian groups. We can, however, suggest adding a new "gene" to this model, one that in our opinion would better adapt this tool to the problem we want to investigate. This gene would be a "contestation" gene which we understand as flaunting a disagreement with the existing situation, a peculiar protest, an expression of frustration. According to our pre-tests with randomly selected working groups, the contestation gene would be one of the most important factors motivating the participants of the group we are interested in. The contest can be creative or destructive, but it could certainly increase the sensitivity of the CI Genome Framework method. Its development would be useful due to the specific character of Central and Eastern Europe marked by street revolutions (Polish Solidarity movement, the Ukrainian Orange Revolution and Revolution of Dignity) based on contesting the existing situation.

4 Conclusions

The presented paper deals with two overlapping research problems. The first of them is a social issue: the relations between Poles and Ukrainians living in Poland. This subject has been described relatively well so far, mostly because it touches a question simply relevant to public entities shaping immigration policy in Poland. The researchers interested in the job market, political geography, law, and other areas also participate in the search for conclusions in this field of research. However, in most cases, the research remains in the Polish and Ukrainian circles. The second issue of the article is an attempt to select research methods suitable to determine to what extent collective intelligence compatible with deliberative communication is a dominant type of debate in this social field. The two separate problems come together under a single goal of enabling observation of Polish-Ukrainian interactions as manifestations of collective intelligence, and furthermore, the use of collected results to stimulate the growth of bridge capital between Poles and Ukrainians. Currently, we are still not able to reach that final point, however, we are convinced that our research should continue and that it can become an invitation to other researchers and incentive to test our directives in practice.

We are already able to evaluate the usefulness of CI research methods in the relevant context, as well as to point out its strengths and weaknesses. The application of deliberation analytics using Issue-Based Information System seems to be the most suitable choice, due to its ability to identify communication rationality, studied at a behavioral level, as non-discretionally as possible. This method will allow us to isolate the space of interaction between the relevant subjects, that is between Poles and Ukrainians communicating on the Internet. The disadvantages mentioned earlier are not obstacles which would prevent laboratory research, but they may prove detrimental in attempts to carry out research in the real existing Internet. Despite this, we assume the relative usefulness of other tools discussed in the main text. In regard to qualitative complex indicators measuring social phenomena – CI Potential Index, UPVoCI scale, and CI Genome framework – we detect their usefulness specifically in those areas where the online analytic server fails, in real-life, outside-the-laboratory conditions. However, this leads us to another problem to solve, which is the acquisition of data necessary to use the CI Maturity Index in practice. The authors of the CIPI method in its final version resigned from using it due to difficulties in data collection.

As a result of further research on our target group, we expect the following results

- Collecting sufficient data to confirm or deny the hypothesis that certain type of CI projects implemented in the public sphere can ease the tensions related to Polish-Ukrainian problems: reduce distrust, reduce polarization, and build communities;
- Verifying the suitability and correctness of using IBIS-based deliberation metrics in our research subject, as well as extending "Groupthink" metric using the SIDE model;
- Reinterpreting the CI Social Maturity Index by showing relations as complex as the relationship of Ukrainians living in Poland and being in it recently, and Poles themselves: the degree of mutual trust, the ability to long-term cooperation;
- Confirmation or denial of the relations observed in the recent laboratory experiments, like the relation of social interaction strength and social density indicators in the connection with the scope of distrust. The laboratory research stated that in some configuration of these variables the increase distrust level can bring salutary effects to CI efficiency. This statement needs to be proved or falsified;
- Having sufficient data on the preferences expressed by individual users during the discussion, we could attempt to visualize deliberation in the form of Collective Mental Map, as proposed by Heylighen [23].
- Checking the possibility of using the newly created Extended Turing Test [61] as a tool for measuring the level of self-organization in groups;
- Checking the validity of adding a new gene to the CI Genome Framework, that could better adapt this tool to the problem we want to investigate. That would be a Contestation gene.
- Due to the intercultural context of our research, i.e. the problem of interaction between Poles and Ukrainians, it seems necessary to check whether the reality requires greater nuance approach to cognitive diversity: as the Woolley et al. [72] study shows, too much diversity can lead to a reduction CI level, but at the same time some level of it increases it.

Our conclusions should pass a falsification test in artificial environments of projected laboratory experiments. If it turns out to be a success, we will gain a theoretical basis for research aiming at finding a solution for a problem which really exists in Central and Eastern Europe and affects many people, both Poles, and Ukrainians. Mutual cooperation will improve the living standards of both groups, but it is necessary to research ways to stimulate it.

Acknowledgements. This paper reports some of the research results obtained within the research grant UMO-2018/28/C/HS5/00543 - "Kolektywna inteligencja w Internecie: zastosowania w sferze publicznej, metody badania i modele partycypacji obywatelskiej", funded by Narodowe Centrum Nauki (National Science Centre, Poland).

References

1. Aitamurto, T., Chen, K.: The value of crowdsourcing in public policymaking: epistemic, democratic and economic value. Theory Pract. Legis. **5**(1), 55–72 (2017)
2. Arendt, H.: The Human Condition. The University of Chicago Press, Chicago (2018)
3. Aktualne problemy ukraińskich migrantów w Polsce. https://pl.naszwybir.pl/wp-content/uploads/sites/2/2014/03/fnw_aktualne-problemy_wnioski.pdf. Accessed 31 May 2019
4. A theoretical framework for the DSI index. https://digitalsocial.eu/images/upload/7-A_theoretical_framework_for_the_DSI_index.pdf. Accessed 26 May 2019
5. Barber, B.: If Mayors Ruled the World. Yale University Press, New Haven (2014)
6. Bates, T.C., Gupta, S.: Smart groups of smart people: evidence for IQ as the origin of collective intelligence in the performance of human groups. Intelligence. **60**, 46–56 (2017)
7. Becker, J., Porter, E., Centola, D.: The wisdom of partisan crowds. PNAS **116**(22), 10717–10722 (2019). https://doi.org/10.1073/pnas.1817195116. First published May 13, 2019
8. Benhabib, S.: Models of public space: Hannah Arendt, the liberal tradition, and Jurgen Habermas. In: Calhoun, C. (ed.) Habermas and the Public Sphere. MIT Press, Cambridge (1992)
9. Bernstein, A., Klein, M., Malone, T.W.: Programming the Global Brain. Commun. ACM **55** (5), 41–43 (2012)
10. Bonabeau, E.: Decisions 2.0: the power of collective intelligence. MIT Sloan Manag. Review **50**(2), 45–52 (2009)
11. Bothos, E., Apostolou, D., Mentzas, G.: Collective intelligence for idea management with internet-based information aggregation markets. Internet Res. **19**, 26–41 (2009)
12. Brzoza, C., Sowa, A.: Historia Polski 1918–1945. Wyd. Literackie, Kraków (2009)
13. Chambers, J.: Deliberation and mass democracy. In: Deliberative Systems. Cambridge University Press, New York (2012)
14. Cohen, J.: Deliberation and democratic legitimacy. In: Matravers, D., Pike, J.E. (eds.) Debates in Contemporary Political Philosophy: An Anthology. Routledge, in Association with the Open University, London (2003)
15. Copsey, J.: Remembrance of things past: the lingering impact of history on contemporary polish-ukrainian relations. Europe-Asia Studies **60**(4), 531–560 (2008)
16. Credé, M., Howardson, G.: The structure of group task performance—a second look at “collective intelligence”: comment on Woolley et al. (2010). J. Appl. Psychol. **102**(10), 1483–1492 (2017)
17. Czerwiński, K.: Przemiany obywatelstwa we współczesnych demokracjach – wyzwania dla edukacji. Podstawy Edukacji nr **7**, 39–59 (2014)
18. Gordon, D.H.: Collective behavior in animals: an ecological perspective. In: Malone, T.W., Bernstein, M.S. (eds.) Handbook of Collective Intelligence. The MIT Press, Cambridge, London (2015)
19. Goyal, A., Akhilesh, K.B.: Interplay among innovativeness, cognitive intelligence, emotional intelligence and social capital of work teams. Team Perform. Manag. **13**(7/8), 206–226 (2007)
20. Gurkan, A., Inadoli, L., Klein, M., Zollo, G.: Mediating debate through on-line large-scale argumentation: evidence from the field. Inf. Sci. **180**, 3686–3702 (2010)
21. Habermas, J.: The Structural Transformation of the Public Sphere. Polity, Cambridge (1989)
22. Habermas, J.: The Theory of Communicative Action, Vol. 1: Reason and the Rationalization of Society. Trans by. T. McCarthy. Beacon Press, Boston (1987)
23. Heylighen, F.: Collective intelligence and its implementation on the web: algorithms to develop a collective mental map. Comput. Math. Organ. Theory **5**(3), 253–280 (1999)

24. Hong, L., Page, S.: Groups of diverse problem-solvers can outperform groups of high-ability problem-solvers. PNAS **101**, 16385–16389 (2004)
25. Jaroszewicz, M., Małynowska, O.: Najnowsza migracja z Ukrainy do Polski (nie)stały fenomen, http://www.batory.org.pl/upload/files/Programy%20operacyjne/Forum%20Idei/Najnowsza%20migracja%20z%20Ukrainy.pdf. Accessed 26 May 2019
26. Klein, M., Spada, P., Calabretta, R.: Enabling deliberations in a political party using large-scale argumentation: a preliminary report. In: 10th International Conference on the Design of Cooperative Systems (2012)
27. Klein, M., Iandoli, L.: Supporting collaborative deliberation using a large-scale argumentation system. In: Conference on Online Deliberation. Berkeley (2008)
28. Klein, M.: The CATALYST deliberation analytics server. MIT Working Paper (2015)
29. Klein, M.: Towards crowd-scale deliberation. SSRN Electronic Journal, January 2017
30. Krause, S., James, R., Faria, J.J., Ruxton, G.D., Krause, J.: Swarm intelligence in humans: diversity can trump ability. Anim. Behav. **81**, 941–948 (2011)
31. Landemore, H., Elster, J. (eds.): Collective Wisdom, Principles and Mechanisms. Cambridge University Press, Cambridge (2012)
32. Landemore, H.: Politics, Collective Intelligence, and the Rule of the Many. Princeton University Press, Princeton and Oxford (2012)
33. Legatum Institute, Eastern Europe: Russia holding back the region. https://www.prosperity.com/feed/eastern-europe-russia-holding-back-region. Accessed 20 May 2019
34. Levy, P.: From social computing to reflexive collective intelligence: the IEML research program. Inf. Sci. **180**, 71–94 (2010)
35. Levy, P.: L'intelligence collective: Pour une anthropologie du cyberspace. Editions La Decouverte. English Edition: Collective Intelligence: Mankind's Emerging World in Cyberspace, Plenum (1997)
36. Lo, A.W.: The wisdom of crowds vs. the madness of mobs. In: Malone, T.W. (ed.) Handbook of Collective Intelligence. The MIT Press, Cambridge (2015)
37. Luo, S., Xia, H., Yoshida, T., Wang, Z.: Toward collective intelligence of online communities: a primitive conceptual model. J. Syst. Sci. Syst. Eng. **18**(2), 203–221 (2009)
38. Łazuga, W.: Kalkulować. Polacy na szczytach CK Monarchii. Zysk i S-ka, Poznań (2013)
39. Malone, T.W., Bernstein, M.S. (eds.): Handbook of Collective Intelligence. The MIT Press, Cambridge, London (2015)
40. Malone, T.W., Laubacher, R., Dellarocas, C.: The collective intelligence genome. MIT Sloan Manag. Rev. **51**(3), 21–31 (2010)
41. Malone, T.W.: Superminds. Little, Brown, New York, Boston, London (2018)
42. Mansbridge, J.J.: A minimalist definition of deliberation. In: Heller, S., Rao, V. (eds.) Deliberation and Development, pp. 27–49. World Bank, Washington, DC (2015)
43. Mansbridge, J.J., et al.: The place of self-interest and the role of power in deliberative democracy. J. Polit. Philos. **18**(1), 64–100 (2010)
44. Massari, G., Giannoccaro, I., Carbone, G.: Are distrust relationships beneficial for group performance? The influence of the scope of distrust on the emergence of collective intelligence. Int. J. Prod. Econ. **208**, 343–355 (2019)
45. Mouffe, C.: Agonistics: Thinking the World Politically. Verso Books, London, New York (2013)
46. Mulgan, G.: Big Mind. How Collective Intelligence Can Change Our World. Princeton University Press, Princeton, Oxford (2018)
47. Noveck, B.S.: Smart Citizens, Smarter State the Technologies of Expertise and the Future of Governing. Harvard University Press, Cambridge (2015)

48. Owen, D., Smith G.: Two types of deliberative system. Presented at the 2013 Annual Meeting of the American Political Science Association American Political Science Association, Chicago (2013)
49. Rittel, H.: Dilemmas in a general theory of planning. Policy Sci **4**(2), 155 (1973)
50. Rittel, H.: Second Generation Design Methods. In: Cross, N. (ed.) Developments in Design Methodology, pp. 317–327. Wiley, Chichester (1984)
51. Rittel, H., Noble, D.: Issue-based information systems for design. Institute of Urban and Regional Development, University of California, Berkeley. OCLC 20155825 (1989)
52. Roy, S., Biswas, B., Chaudhuri, S.S.: Nature-inspired swarm intelligence and its applications. Int. J. Mod. Educ. Comput. Sci. **12**, 55–65 (2014)
53. Schut, M.C.: On model design for simulation of collective intelligence. Inf. Sci. **180**, 132–155 (2010)
54. Siemaszko, W., Siemaszko, E.: Ludobójstwo dokonane przez nacjonalistów ukraińskich na ludności polskiej Wołynia 1939–1945. Von Borowiecky, Warszawa (2000)
55. Skaržauskienė, A.: Social Technologies and Collective Intelligence. Mykolas Romeris University, Vilnius (2015)
56. Steenbergen, M.R., Bächtiger, A., Spörndli, M., Steiner, J.: Measuring political deliberation: a discourse quality index. Comp. Eur. Polit. **1**(1), 21–48 (2003)
57. Stefanelli, F., Imbimbo, E., Carletti, T., Guarino, A., Burette, F., Guazzini, A.: Amathematical model of Collective Intelligence. Accessed 18 Oct 2019
58. Stromer-Galley, J: Measuring deliberation content: a coding scheme. J. Public Deliberation **3, 2**(1) (2007)
59. Surowiecki, J.: Wisdom of Crowds. Anchor Books, New York (2005)
60. Szuba, T.: Computational Collective Intelligence. Wiley-Interscience, New York (2001)
61. Szuba, T.: Formal and computational model of Adam Smith's Invisible Hand. In: 2018 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn.2018.8489091
62. Tanis, M., Postmes, T.: Short Communication. A social media identity approach to trust: Interpersonal perception, group membership and trusting behavior. Eur. J. Soc. Psychol. **35**, 413–424 (2005)
63. Tucci, C.L., Afuah, A., Viscusi, G. (eds.): Creating & Capturing Value Through Crowdsourcing. Oxford University Press, Oxford (2018)
64. Tyma, P. (ed.): Mniejszość ukraińska i migranci z Ukrainy w Polsce. http://www.oup.ukraina.com.pl/images/stories/DOCUMENT/raport_mniejszosc_ukrainska.pdf. Accessed 23 May 2019
65. Verhulst, S.G.: Where and when AI and CI meet: exploring the intersection of artificial and collective intelligence towards the goal of innovating how we govern. AI Soc. **33**(2), 293–297 (2018)
66. Vilanova, F., Beria, F.M., Costa, A.B., Koller, S.H.: Deindividuation: from Le Bon to the social identity model of deindividuation effects. Cogent Psychol. **4**, 1308104 (2017)
67. Vincenzo, I., Giannoccaro, I., Carbone, G., Grigolini, P.: Criticallity triggers the emergence of collective intelligence in groups. Phys. Rev. E **96**, 022309 (2017)
68. Vincenzo, I., Massari, G., Giannoccaro, I., Carbone, G., Grigolini, P.: Mimicking the collective intelligence of human groups as an optimization tool for complex problems. Chaos, Solitons Fractals **110**, 259–266 (2018)
69. Weng, S., Yang, M., Hsiao, P.: A factor identifying study of the user-perceived value of collective intelligence based on online social networks. Internet Res. **28**(3), 696–715 (2018)
70. Wise, S., Paton, R.A., Gegenhuber, T.: Value co-creation through collective intelligence in the public sector. Vine **42**(2), 251–276 (2012)

71. Woolley, A.W., Aggarwal, I., Malone, T.W.: Collective Intelligence and group performance. Curr. Dir. Psychol. Sci. **24**(6), 420–424 (2015)
72. Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W.: Evidence for a collective intelligence factor in the performance of human groups. Science **330**(6004), 686–688 (2010)
73. Żurawski vel Grajewski, P.: Duch pyszny poprzedza upadek. Rozważania o naturze procesu rozpadu unii. Ośrodek Myśli Politycznej, Kraków (2012)

Information Diffusion and Processing

A Network Neutral Alternative to Free Basics

Ramneek[1](✉), Patrick Hosein[2], and Sungin Jung[1]

[1] Electronics and Telecommunications Research Institute, Daejeon, South Korea
{ramneek,sijung}@etri.re.kr
[2] The University of the West Indies, St. Augustine, Trinidad
patrick.hosein@sta.uwi.edu

Abstract. The evolution of mobile communication systems and the availability of affordable smart-phones and advanced applications have led to a rapid increase in the number of connected devices and cellular subscriptions. Although the number of connected mobile devices exceeded the total population of the world in 2014, two-thirds of the world's population still has no access to the Internet. Internet.org, also known as Free Basics, is a Facebook led initiative aimed at providing Internet access to the two-third of the world's population who do not have it. Through partnerships with mobile operators in different parts of the world, it aims at providing the benefits of connectivity to the poor population who cannot afford a mobile data plan. However, it has a number of associated issues that can lead to the violation of Network Neutrality principles. In addition, network operators have to rely on the design principles of Free Basics applications to ensure minimum network usage so that paid customers do not suffer significant performance degradation. Due to these drawbacks, there is a need for a network neutral solution that benefits both the operators and the end users, without any third party intervention. In this paper we propose a controlled Low Throughput approach, which we henceforth call LTP, for providing free Internet access to those who need it while providing performance guarantees to users of paid subscriptions. In particular, the proposed approach is also Network Neutral.

Keywords: Network Neutrality · Free Basics · Digital divide · Zero-rating · QoS

1 Introduction and Background

At present, approximately 3 billion people in the world have access to the Internet [24]. These account for roughly 40% of the total population of the world. The rapid evolution of Internet technology and an increase in the affordability of smart devices has resulted in a drastic increase in the mobile data traffic and the number of connected devices on the Internet. Although the number of connected devices on the Internet surpassed the world's population in 2014,

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 189–203, 2019.
https://doi.org/10.1007/978-3-030-34770-3_14

two-third's of the world's population still has no access to the Internet [3]. This is of concern because Internet adoption is growing at a very low rate of about 9% per year. There are a number of factors that contribute to the current digital divide between the users and the non-users of the Internet [30]. The main factors include, high cost of data plans and devices, digital illiteracy, lack of robust infrastructure and security issues. Widening the internet access is a necessary condition for bridging the digital divide, but not sufficient. The mass connectivity can be further impeded by the lack of infrastructure and lack of sufficient knowledge to be able to comprehend the information accessible via access to the Internet. Due to the large investment required for the global infrastructure needed to provide Internet access, data plans offered by operators are quite expensive. However, a majority of people in developing countries do not have enough disposable income to spend on such pricey data plans. According to a study by Deloitte [4], the main factor behind the low Internet penetration rate is income level. Since the amount of video content is increasing rapidly and accounts for the majority of Internet traffic [3], data subscription plans are expected to remain high in the future. Therefore, it is important to provide affordable services in order to narrow the digital divide and to extend the benefits of the infrastructure and technology evolution to the poor.

Most of the solutions offered by service providers are based on zero rating of a few popular services. As users will naturally be inclined towards using zero rated services instead of paid ones, such proposals might be favorable for the zero rated content providers, but are not network neutral. Moreover, providing access to a few walled-garden services may stifle the development of economic and network neutral access in such areas which are already lagging far behind. One such initiative, known as Free Basics (formerly known as internet.org), was started by Facebook with the primary aim of connecting the unconnected to the Internet [6]. It is supported by a number of companies such as Ericsson, Samsung, Mediatek, Nokia, Qualcomm and a number of mobile operators.

Free Basics by Facebook provides free access to a diverse set of applications over the Internet, with the support of mobile operators, across a number of developing nations in Asia, Africa and Latin America. It aims at making the Internet affordable as well as increasing awareness by providing free access to a range of services such as AccuWeather, BBC News, BabyCenter and MAMA, Dictionary.com, Wikipedia, Facebook, Messenger, Facts for Life, etc. According to internet.org, they have managed to connect around 15 million people to the Internet who would not have been connected otherwise. In addition, through their connectivity endeavor, they have helped many application developers who have joined the Free Basics ecosystem, to increase their customer base and overall social impact. By accelerating Internet adoption, they have helped such applications to establish themselves in the ever growing mobile market [5]. For example, the number of daily searches for SmartBusiness, a South African platform for helping people to launch and run businesses, has increased five-fold since its inclusion in the free basics platform. Similarly Maya, a website for women healthcare support, receives 18 times more traffic every day with more than 70% of it originating from Free Basics [5].

Although Free Basics is claimed to be a non-profit initiative aimed at reducing the cost of Internet access, many have voiced their objections because it violates the basic principles of Network Neutrality as the offered services are favored over competitors. It also violates the principle of openness as the users cannot freely choose what they want to access over the Internet [11]. To address these issues, they re-branded internet.org as the Free Basics platform where any application or website that conforms to the design principles of Free Basics could be a part of the platform [26]. However, Facebook still maintains the role of a gatekeeper as all data for these apps must flow through Facebook's internet.org proxy. In addition, supporting zero-rated applications is against the principle of Network Neutrality [19,22]. Due to its gate keeping approach, many dominant partners in India, Brazil, Indonesia and others have withdrawn their support [14].

Therefore there is need for a Network Neutral solution that provides free or cheap access to the Internet without any third-party intervention. In this paper, we propose Low ThroughPut (LTP), a network neutral approach that benefits both the mobile operators and the underprivileged end users by providing free Internet access while taking the mobile operator's resource efficiency into consideration.

The rest of the paper is organized as follows: Sect. 2 provides an overview of the principles of Network Neutrality; Sect. 3 includes a brief description of the Free Basics guidelines, as an example of the existing solutions; Sect. 4 describes the proposed approach in which we describe how both Consumers and Operators can benefit without a third party, followed by simulation results in Sect. 5; Sect. 6 provides a brief discussion about the related work, followed by conclusions in Sect. 7.

2 Network Neutrality

The principle of Network Neutrality advocates an equal treatment of all Internet content by network service providers without any kind of blocking, discrimination or paid prioritization. The users must be free to access all applications and content available on the Internet, irrespective of its source or origin. Network Neutrality aims at protecting the openness of the Internet, by limiting any kind of preferential treatment of any content, and ensuring that network providers do not block or limit access to any application or website [12,21]. Over the last few years there has been widespread discussion on the issue of Network Neutrality. It is of importance due to its technical and political implications and the potential impact it can have on future Internet design and development [2,13].

Supporters of Network Neutrality assert that there has been incredible innovation and growth in Internet technology and applications in the past due to the openness of the Internet and, to ensure that it continues in the future, any kind of discrimination or paid prioritization should be banned. They argue that, in the absence of Network Neutrality regulations, network service providers might be tempted to discriminate (charge based on the content provider) or block content from certain application providers, which can lead to a transformation from the market that supports innovation to one where the success of the application or service depends on money paid to network providers [16]. Moreover, the absence of Network Neutrality can have adverse effects on the emerging Internet of Things (IoT) based applications. Although connected IoT devices produce small amounts of data, the sensitivity of data to latency could cause serious issues if service providers are allowed to block or throttle traffic at their own will. ISPs could easily prioritize traffic from their own solutions or throttle the traffic from their competitors, creating an unfair playing field.

On the other side of the debate opponents, including mainly network providers, allege that if differential charging is prohibited by law, then they might lose the incentive to invest in network infrastructure and it will be unreasonable to apply the same traffic management rules to content providers whose websites or apps contribute to heavy data traffic and to other small players producing significantly less traffic [35].

Telecommunications regulatory authorities worldwide, including those in Belgium, Netherlands, Russia, Canada, India, Japan, Argentina and Brazil [31], have been investigating various issues related to Network Neutrality violation and framing laws and regulations for limiting the ability of Internet access providers to throttle, block or discriminate between the content that is available over their networks. In the United States of America, the Federal Communications Commission (FCC) [7], proposed rules for an open internet in October 2009 and asked for reviews. Finally, the FCC's Open Internet rules were adopted in February 2015. However, the FCC repealed network neutrality rules in 2017, stating that they were only focused on resolving the anti-competitiveness among the service providers and that intervention of government in internet issues could stifle its growth and innovation. This received a lot of opposition during the open comment period and the debate is still ongoing [32].

In the case of the European Union, some member states including the Netherlands and Slovenia have already crafted laws to safeguard Network Neutrality while the debate is still ongoing in other member states [15,23]. The European Parliament and the Council of Ministers passed the first version of the Telecommunications package in 2009 with an agenda to preserve the open and neutral Internet. It also required network providers to ensure some minimum QoS to end users and provide QoS information transparently to them. Although it did not impose any legal restrictions on the member states, it fired the Network Neutrality debate.

As networks have advanced rapidly and are now capable of satisfying different types of QoS requirements of different applications, more fine-grained Network Neutrality rules and regulations are required. A number of such rules have been proposed in the literature [21]. While the all or nothing approach that bans all kinds of discrimination are over restrictive and can hinder the evolution of the network, allowing all discrimination defeats the purpose of the rule as it would allow blocking. A Case-by-case approach on the other hand bans discrimination that is harmful under the antitrust standard. Such approaches might fail to cover some important aspects of discrimination as the decision about what is considered as harmful or unreasonable is left for the future. Moreover, it must be noted that the network neutrality debate is not a debate against QOS, and proper rules for differentiating between acceptable and unacceptable forms of QoS are required.

Recently, the focus of the Network Neutrality debate has shifted to zero-rated applications, i.e., whether zero-rated applications should be prohibited by Network Neutrality rules or not. Zero-rated applications, also known as sponsored or toll-free applications, are those whose data usage is not counted against the monthly bandwidth cap of the user. These applications, provided free of charge by the mobile network operators, are gaining popularity in a number of countries. Currently more than 45% of network service providers in the world are providing at least one zero-rated application to their subscribers [33]. The main advantage of providing zero-rated applications is that it subsidizes the cost of data for customers, especially for those who have never accessed Internet services and applications. This can result in the expansion and acceleration of Internet access leading to an enhanced awareness as well as social and economic growth.

Some supporters believe that zero-rating is less harmful than other forms of discrimination as zero-rated applications are provided at the same QoS level as regular applications. However, there is a price-based discrimination which results in more users being attracted to such applications than their paid counterparts [22]. The network service providers either charge the content or application providers for zero-rating their content or it may zero-rate certain applications without charging any fee from content providers [19]. In the former case, it can be a big disadvantage for small application or content providers who may not be able to pay the price for being a part of such schemes. Such practices impose entry-barriers for small players, thereby suppressing innovation. In the latter case, although network operators might not charge for zero-rating certain applications, it is a form of anti-competitive discrimination as the content providers, whose apps are not being zero-rated, might face difficulties in attracting users in the presence of free substitutes. Therefore zero-rating is against the fundamental principles of Network Neutrality.

3 Free Basics

In this section we briefly describe Free Basics as an example of currently used solutions. The Free Basics platform can be accessed through the FreeBasics App for the Android platform or by visiting FreeBasics.com on a mobile device. The websites or apps that are a part of Free Basics are provided for free, i.e. without any data charges. If the service is provided for free, more people will be connected to the Internet and might convert into paid customers once they experience the benefits of being connected. Hence, in the long term, this effort can benefit mobile operators as well. Therefore, it is important to design the applications in such a way that they enhance the user experience. As the majority of people cannot afford smart-phones and might still be using feature phones, it is important to design the applications so that they work well on all kinds of mobile devices. In addition, one should ensure that the data consumption by such free applications remains low so that there is no quality degradation for paid subscription customers. In order to address the above two issues, the Free Basics platform has some technical guidelines and terms of use which must be complied with, by the websites or application, to become a part of the Free Basics program.

For a website to be listed in Free Basics, it must not use Java script, SVG images, WOFF font types, iframes, videos, large images, and flash and java applets. Large data-consuming content such as videos and high-resolution images of size greater than 1 MB are not permitted. All services are required to be tested for different use cases such as throttled networks, disabled JavaScripts, etc., using the Chrome Device Emulator and the Opera Emulator [5]. In addition, all traffic from within Free Basics is routed through the internet.org proxy as shown in Fig. 1. This is done to standardize the traffic flow and make it easy for operators to identify the services that are a part of Free Basics and zero-rate them. The services that are a part of Free Basics are neither cached nor hosted by internet.org. Free Basics proponents argue that they are not anti-network neutral [27], as the principles of universal connectivity and Network Neutrality do not conflict with each other, and can co-exist. However, it is a closed domain, and grants access to only a limited number of Internet destinations. As it has the potential to influence a large percentage of the world's population, it can represent the entirety of the Internet for them. This may deprive them of the additional benefits they may get from the other content and competing services. In addition, although Facebook allows any application to be a part of Free Basics if it follows the rules and guidelines, it still maintains a role of gatekeeper. All data for these apps must be accessed via Facebook which benefits from harvesting the large amount of user data flowing through its proxy [14,25,29].

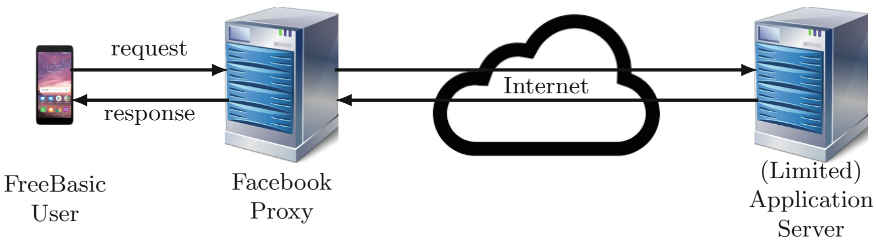


Fig. 1. The free basics architecture

4 The LTP Approach

In the LTP approach free Internet service is provided but at a restricted Quality of Service. Therefore, instead of relying on application restrictions for low data consumption as in the case of Free Basics, we limit the throughput that is provided to LTP users. By doing so, network operators can manage their resources through proper scheduling and ensure that premium or paid subscription users do not suffer performance degradation as the number of LTP users increases. Unlike Free Basics, where users have access to a handful of applications, LTP users can access any website on the Internet as illustrated in Fig. 2. In this way, any website or application that wants to extend its service to these free users can develop an LTP version of their app (using similar techniques used for Free Basics Apps but not limited in any way to those choices) so that they can provide satisfactory service to LTP users. It must be noted that it is essential for data intensive application providers to optimize their apps for restricted QoS in case of congestion, as it is not possible to support high traffic generated by these apps for free, without penalizing the paid user's QoS. This can be achieved by using a mix of static and dynamic content, prioritizing content, and replacing the multimedia content by a lite version of it to maintain the user experience even in the case of restricted QoS provisioning by LTP. For instance, Google recently announced that it will be launching a special version of its popular products such as Youtube and Chrome, designed to provide more efficient service to people with low-bandwidth network connectivity [34]. Such applications can be provisioned efficiently over the LTP channel without any restrictions. Note there is no need for any third party gate-keeping.

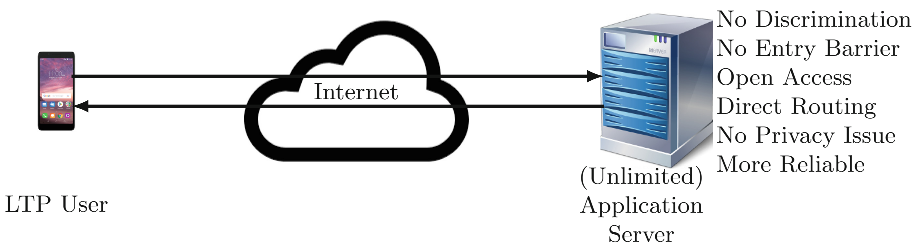


Fig. 2. The LTP architecture

In addition, the operator can allocate a fixed percentage of bandwidth to LTP users. For instance, if 5% of the operator's bandwidth is reserved for LTP users, the throughput assigned to these users can be dynamically varied to achieve this limit. In this case, the average throughput provided to the LTP users can be increased if the number of users decreases and vice-versa. Note that paid subscription users are not allowed to switch to LTP usage to avoid scenarios where one switches if the number of LTP users is low. However, LTP users can switch to a paid subscription whenever they desire. As LTP users can access the whole Internet the operator can also set a data cap for such users. This can help to ensure that some greedy LTP users do not occupy bandwidth for extended periods of time.

Network operators can use different scheduling and load management techniques as well as tiered service models and pricing strategies for implementing the LTP approach. For example, the operator can categorize the users into two service classes: Paid users (those who have subscribed to a data plan) and LTP users (free users who have no data plan). At any point in time let T_p denote the total throughput of all paid users and let T_f denote the total throughput of all LTP users. At each scheduling instant, we compute T_p and T_f. Let x denote the fraction of capacity that is reserved by the mobile operator for the LTP users. At each scheduling instant, if $\frac{T_f}{(T_p+T_f)} > x$ then a paid user is scheduled, else, a free user will be scheduled. If a paid user is to be served but none require service then a free user, if available, will be served. Similarly if a free user is to be served and there is none, then a paid user can be served.

If $T_f + T_p$ is less than capacity then all traffic can be accommodated and the above algorithm has no effect. However, when $T_f + T_p$ reaches capacity then the algorithm ensures that T_f is no more than x times the capacity. In this way we can ensure that the QoS achieved by paid subscription users is not significantly affected as the number of LTP users increases. The paid users can be scheduled by the network operator using any QoS-based scheduling approach in order to provide their QoS guarantees [8,20].

The LTP users can be scheduled using the Proportional Fair algorithm. In this case the throughput achieved by a user is proportional to their achievable rate. Users with good channel conditions therefore achieve a higher average throughput than those with poor channel conditions. Of course other approaches can be used such as always serving the user in the best channel conditions. This approach will maximize the total throughput of the LTP users but at the expense of fairness.

When the network is lightly loaded, the LTP users get similar service to that of the paid users. However, as the loading increases the free users will get no more than a fraction of x of the available bandwidth. Since this fraction of the total bandwidth is shared by all active LTP users we need to maintain some lower bound on throughput to ensure acceptable service. One approach to do this is as follows. If the average LTP user throughput drops below some threshold (e.g. 32 kbps) then we block new LTP connections. Once the average throughput again rises above this threshold then we can continue to accept new LTP connections. In addition, to ensure some degree of fairness among LTP users, we can also do the following. If the average LTP user throughput drops below a second threshold (below that of the above threshold) then, in addition to blocking new LTP connections, we also drop one or more existing LTP connections. These users can be chosen based on either the length of time they have been active or the amount of data they have downloaded on average. This will allow new LTP users to be added even under congestion. In addition, a daily or monthly data cap can be set by the operator to control the amount of data used by free LTP users. In this way we can (i) guarantee that no more than a fraction x of available bandwidth is allocated to LTP users, (ii) guarantee that free users achieve some minimum bandwidth if they are admitted into the system and (iii) ensure that LTP users are fairly treated and that no user can hog the system.

The advantages of the LTP approach over the Free Basics approach can therefore be summarized as follows: (1) Operators have better control over the performance of their paid subscribers since the rate offered to LTP users varies depending on congestion (this is not the case for Free Basics); (2) The LTP approach is Network Neutral since LTP users can access any Internet application (content providers can choose to develop a LTP version of their site as they do for a mobile version of their app to enhance the performance experienced by LTP users); (3) The reliability and performance of LTP users is better since there is no single point of failure (like the Free Basics proxy) and there is no triangular routing (as with routing through the Free Basics router); (4) In addition, in order to avoid abuse of the system (such as frequent downloading of large video files) the operator can provide data caps (e.g., 5MB per day). This cap can also be used to control the bandwidth provided to LTP users in order to limit the effect of the service on the performance of paid subscription users. Note that this allows the LTP users to experience services that they would not have been able to access via the Free Basics approach. For example, they can view a video file during a low congestion period (say late at night) since they will achieve acceptable download rates during lightly loaded periods. This exposure will increase the likelihood of switching to a paid subscription; The content and application providers can leverage the unrestricted access and level playing field provided by LTP to attract more customers by providing applications optimized for restricted QoS and also encourage the migration to paid subscriptions; (5) Finally there is no third party (like Facebook) involved which increases the data privacy and security afforded to LTP users. Hence, the proposed LTP approach ensures affordability (by providing free access to the internet, and by providing tiered QoS plans that can be chosen by the user based on his QoS requirements and budget), accessibility (by providing unrestricted access to all the applications and services) and quality (by guaranteeing a minimum percentage of bandwidth for LTP users that are admitted to the system while ensuring fairness).

5 Simulation Results

In this section, we illustrate the performance advantages of LTP over Free Basics through simple examples. With Free Basics, free users and paid subscription users are treated equally since one can assume that the free users will not require significant resources because of the constraints placed on the applications as dictated by Facebook. Therefore free users will have high throughput but over very short periods. When there are many such users, the capacity, as seen by paid subscription users, will be highly variable. In the case of LTP, since the total throughput of LTP users is limited, then the capacity, as seen by the paid subscription users, has much less variability. Therefore, even if the average capacity made available to paid subscription users in both cases is the same, the variability differs. However, the throughput performance of transmission protocols such as TCP drops with increasing variability [10]. The increased variability in available bandwidth results in an increase in the variability of packet round trip times (RTT). The TCP Re-transmission Time Out (RTO) value is a function of the RTT but also depends on the variation of the RTT. As the RTT variation goes up, the RTO value is increased and this results in decreased TCP throughput.

We illustrate this with a simple simulation. We simulate a single paid user and n free users for both Free Basics and LTP cases. These users are communicating at the same time over a shared network of capacity 600 Kbps. The simulation was performed using the Network Simulator-3. The paid user is downloading a large file and hence its TCP connection is continuous. In addition, we simulated ten free users each downloading a small file of size 60 Kbs at periodic intervals of time. The time between transmissions is assumed to be exponentially distributed with a mean value of 10 s.

In the case of Free Basics, the network capacity of 600 Kbps is shared equally by the paid and the free users as there is no restriction on the throughput. The users are just restricted to small sized files in order to minimize the data consumption. However, in the case of the LTP approach, the total throughput available to free users is restricted to a fraction $x = 0.1$ of the total capacity. Hence the LTP users are restricted to $60/n$ Kbps. The TCP throughput achieved by the paid user in the case of Free Basics and in the case of LTP is shown in Fig. 3. We clearly see the fluctuation in capacity offered to the paid user in the case of Free Basics when compared to the LTP case. We also computed the exponentially filtered throughput of the paid user in both cases and plotted them in Fig. 4. We find that the average TCP throughput of the paid user in the Free Basics case is lower because of the greater estimate of the RTO timeout.

6 Related Work

Compliance with the fundamental principles and guidelines of Network Neutrality is an important issue in networking as it has many political, economic and social implications associated with it. While expanding internet access is essential for narrowing the digital divide and for expanding the knowledge economy, proper measures must be taken to prevent network operators and service providers from unfairly monetizing the access and usage of their network resources. Recently, the focus of the Network Neutrality debate has shifted from the pros and cons of no-blocking and no-discrimination to zero-rated services and applications. Although zero-rated applications are provided at the same QoS as other applications, they violate the principles of Network Neutrality due to their discriminatory effect. If the network operators charge the content providers in exchange for zero-rating their applications then it imposes a threat to innovation as it is a form of preferential treatment and many of the newer players might not be able to afford it. In addition, operators may provide preferential treatment to such applications and lower the quality for others in order to attract more content providers to pay for zero-rating their applications. In some cases, the operators can zero-rate some applications without charging the content providers, but it is still against the principles of Network Neutrality. For instance, the zero-rating

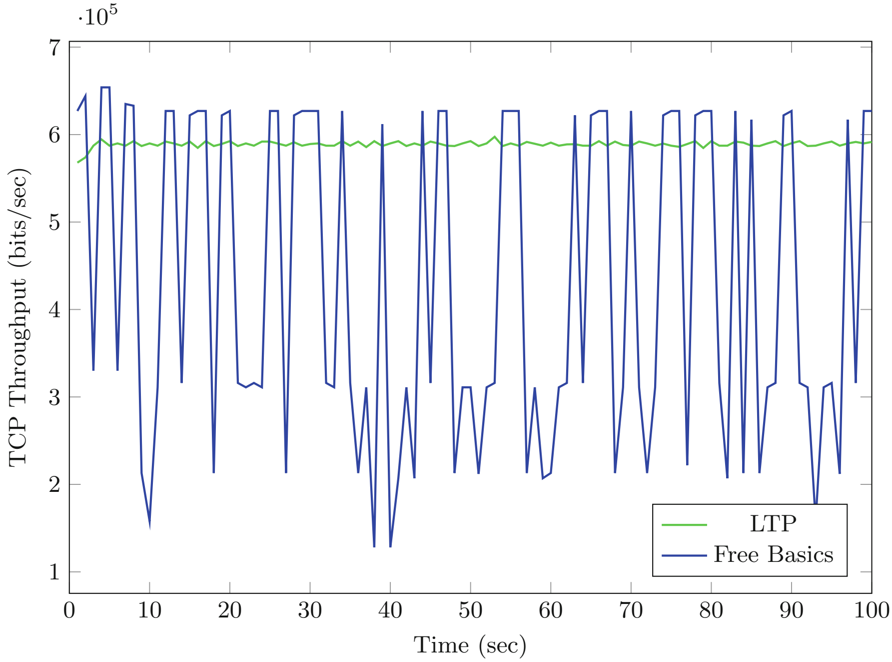


Fig. 3. Instantaneous TCP throughput of the paid user

of Free Basics by different operators does not serve the purpose of providing internet access to the undeserved communities. Instead, there is a need for a Network Neutral solution that provides a low cost or limited access to the entire Internet than to a small number of applications.

A number of Network Neutral alternatives to zero-rating and Free Basics have been proposed. For example, Aircel launched a program called Free Basic Internet (FBI) in India, aimed at providing low-speed Internet access at 64 Kbps to all its subscribers, free of charge [28]. Although this initiative is Network Neutral, it has some limitations. Since the speed is too low, video-streaming, online gaming, video calling and other high bandwidth applications are not supported. In addition, although it provides free access to the entire Internet at a low-speed, similar to our proposed mechanism, there are some performance issues due to which it might not be feasible in the long run. If a large number of free users access the network simultaneously, network congestion can increase rapidly leading to a degradation of performance for the paid users. This can increase the network churn in the long run. In contrast, in the proposed LTP approach, the free users are not only restricted to a low throughput, but there is also a restriction on the percentage of bandwidth assigned to them. Hence, the average throughput achieved by the paid users in the case of LTP is expected to be higher and less bursty as compared to those in the case of Free Basic Internet.

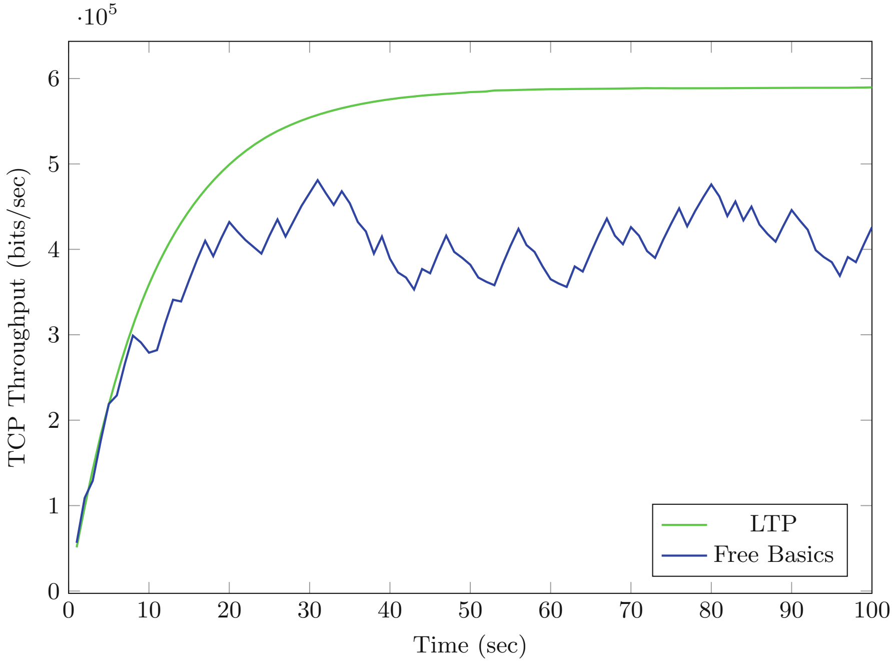


Fig. 4. Filtered throughput of the paid user

Another alternative is an application called mCent that was launched by a company called Jana. It is an Android application that is supported by various operators through a backend billing system [17]. The app users are reimbursed for downloading and using certain apps, and the free data achieved can be used in accessing any content or application on the Internet. The app developers in turn pay the fees to Jana depending on the number of users that downloaded and used their apps. Similarly, Gigato is another application where users get free data when they use other apps from Gigato's partners [18]. Although these app developers do not play the role of a gatekeeper to decide which apps the users must be allowed to access, the developers of different apps still have to pay for their applications to be advertised through these platforms. In addition, the apps are supported only through the Android platform and users who cannot afford these smart phones will not be able to take advantage of the service.

Mitchell Baker of Mozilla proposed another Network Neutral alternative to zero-rating known as equal-rating [1]. The first option is to provide some minimum amount of data for free or at discounted prices, giving the sponsoring companies a "brought to you by" acknowledgment. The second option includes providing unrestricted data in return for watching advertisements. Currently this option has been implemented by Grameenphone, a mobile phone operator in Bangladesh, where users can receive 20MB of unconstrained data per day after watching an advertisement. In this case, the traffic from the equal rated users gets the same treatment as the traffic from the paid subscribers, and hence, there can be performance issues similar to Facebook's Free Basics. Although a small amount of data is provided for free, if a large number of users are online at the same time, it can increase the congestion in the system thereby lowering the achieved QoS for those who have paid plans.

Another solution, as discussed in [9] is to decouple the Internet access from any content or platform and instead limit it by volume or time. However, such a scheme may face serious performance issues if a large number of free users are online at the same time thereby consuming a lot of bandwidth. This can result in bursty periods where the paid user's performance is affected adversely. Therefore, by limiting the throughput as well as total data usage as in the case of LTP, we aim at addressing these performance issues while conforming to the principles of Network Neutrality.

7 Conclusion

Accelerating Internet adoption and awareness by providing the service for free is a novel idea but it should not be done at the cost of people's freedom to access and use information. Although Free Basics by Facebook is aimed at reducing the cost of Internet access and extending the benefits of connectivity to the under-privileged, it has a number of flaws due to which it has faced strong opposition ever since its inception. The main issue is related to the violation of Network Neutrality. Instead of just providing free access to Facebook and a handful of other applications for free, the mobile network operators should provide low cost, limited access to the entire Internet. Therefore we proposed LTP, a network neutral alternative to Free Basics which is beneficial to all involved parties.

Acknowledgment. This work is supported by an IITP grant funded by the Korean government Ministry of Science and ICT (MSIT) under the project titled *Research on High Performance and Scalable Many-core Operating System* (Grant No: 2014-3-00035).

References

1. Baker, M.: Zero rating and the open internet, May 2015. https://blog.lizardwrangler.com/2015/05/06/zero-rating-and-the-open-internet/
2. Beckert, B.: Network neutrality from an innovation research perspective. In: 2011 50th FITCE Congress (FITCE), pp. 1–5, August 2011. https://doi.org/10.1109/FITCE.2011.6133426
3. Cisco: Cisco visual networking index: global mobile data traffic forecast update, 2017–2022 (2017–2022). http://www.cisco.com
4. Delloitte: Value of connectivity: economic and social benefits of expanding internet access (2014). https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-tmt-value-of-connectivity-tmt.pdf
5. Facebook: Free basics (2015). https://developers.facebook.com/docs/internet-org

6. Facebook: Our mission-internet.org (2015). https://info.internet.org
7. FCC: Open internet — federal communications commission (2015). http://fcc.gov/openinternet
8. Hosein, P.A.: QoS control for WCDMA high speed packet data. In: 4th International Workshop on Mobile and Wireless Communications Network, pp. 169–173 (2002). https://doi.org/10.1109/MWCN.2002.1045716
9. Telecommunications Regulatory Authority of India: Consultation paper on differential pricing for data services, December 2015. http://trai.gov.in/consultation-paper-differential-pricing-data-services
10. Jain, M., Dovrolis, C.: End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput, vol. 32. ACM (2002)
11. Jordan, S.: Some traffic management practices are unreasonable. In: 2009 Proceedings of 18th International Conference on Computer Communications and Networks, pp. 1–6, August 2009. https://doi.org/10.1109/ICCCN.2009.5235365
12. Jordan, S.: Implications of internet architecture on net neutrality. ACM Trans. Internet Technol. **9**(2), 5:1–5:28 (2009). https://doi.org/10.1145/1516539.1516540
13. Jordan, S., Ghosh, A.: A framework for classification of traffic management practices as reasonable or unreasonable. ACM Trans. Internet Technol. **10**(3), 12:1–12:23 (2010). https://doi.org/10.1145/1852096.1852100
14. Karr, T.: Global internet activists give thumbs down to facebook's internet.org — billmoyers.com, May 2015. http://billmoyers.com/2015/05/12/internet-zuckerberg-facebook/
15. Kitsing, M.: Network neutrality in Europe. In: Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance, ICEGOV 2011, pp. 313–316. ACM, New York (2011). https://doi.org/10.1145/2072069.2072126
16. Lee, R.S., Wu, T.: Subsidizing creativity through network design: zero-pricing and net neutrality. J. Econ. Perspect. **23**(3), 61–76 (2009). https://doi.org/10.1257/jep.23.3.61. http://www.aeaweb.org/articles?id=10.1257/jep.23.3.61
17. Olson, P.: This app is cashing in on giving the world free data, July 2015. http://www.forbes.com/sites/parmyolson/2015/07/29/jana-mobile-data-facebook-internet-org/#294d28b029a6
18. POst, H.: Gigato tries to make internet access affordable with data rebates, July 2015. https://www.huffingtonpost.in/entry/gigato-app-data-rebates_n_55bbb899e4b0d4f33a02b5ed
19. Ramneek, Hosein, P., Choi, W., Seok, W.: Disruptive network applications and their impact on network neutrality. In: 2015 17th International Conference on Advanced Communication Technology (ICACT), pp. 663–668, July 2015. https://doi.org/10.1109/ICACT.2015.7224879
20. Ramneek, Hosein, P., Seok, W.: Load metric for QoS-enabled cellular networks and its possible use in pricing strategies. In: 2014 IEEE Symposium on Wireless Technology and Applications (ISWTA), pp. 30–35, September 2014. https://doi.org/10.1109/ISWTA.2014.6981190
21. van Schewick, B.: Network neutrality and quality of service: what a non-discrimination rule should look like. In: 67 Stanford Law Review, pp. 1–166 (2015)
22. van Schewick, B.: Network neutrality and zero-rating, February 2014. http://apps.fcc.gov/ecfs/document/view?id=60001031582
23. Scott, B., Heumann, S., Kleinhans, J., Global Commission on Internet Governance and Centre for International Governance Innovation and Chatham House: Landmark EU and US Net Neutrality Decisions: How Might Pending Decisions Impact Internet Fragmentation? Paper series, Chatham House (2015). https://books.google.co.kr/books?id=KYhtjwEACAAJ

24. Stats, I.L.: Number of internet users (2016) - internet live stats (2016). http://www.internetlivestats.com/internet-users/
25. TImes, H.: Trai to reliance communications: Put free basics on hold for now, December 2015. http://www.hindustantimes.com/tech/trai-puts-free-basics-on-hold-for-now/story-Cnm2juiFb1OQZzRwxmgT7O.html
26. Times, I.: Facebook rebrands internet.org as free basics-adds new features, September 2015. http://timesofindia.indiatimes.com/tech/tech-news/Facebook-rebrands-Internet-org-as-Free-Basics-adds-new-features/articleshow/49101242.cms
27. TImes, L.: Mark zuckerberg seeks to ease net neutrality concerns over internet.org, April 2015. http://www.latimes.com/business/technology/la-fi-tn-mark-zuckerberg-net-neutrality-20150417-story.html
28. Trak.In: Aircel launches free basic internet for all which is actually net neutral, free and fair, October 2015. http://trak.in/tags/business/2015/10/16/aircel-free-basic-internet-net-neutral-free-fair-facebook/
29. Tribune, I.: Neither good nor free, January 2016. http://www.tribuneindia.com/news/comment/neither-good-nor-free/182383.html
30. West, D.M.: Digital divide: Improving internet access in the developing world through affordable services and diverse content (2015). http://www.brookings.edu/~/media/research/files/papers/2015/02/13-digital-divide-developing-world-west/west_internet-access.pdf
31. Wikipedia: Net neutrality by country-Wikipedia, May 2019. https://en.wikipedia.org/wiki/Net_neutrality_by_country
32. Wikipedia: Net neutrality in the United States, May 2019. https://en.wikipedia.org/wiki/Net_neutrality_in_the_United_States
33. Wireless, F.: Report: 45% of operators now offer at least one zero-rated app (2015). http://www.fiercewireless.com
34. World, P.: Google rolls out more low-bandwidth versions of its products, September 2016. http://www.pcworld.com/article/3124908/google-rolls-out-more-low-bandwidth-versions-of-its-products.html
35. Yoo, C.S.: Beyond network neutrality. Harv. J. Law Technol. **19**(1) (2005). http://jolt.law.harvard.edu/articles/pdf/v19/19HarvJLTech001.pdf

Intransitiveness in Games and Random Walks

Alberto Baldi[1], Giulia Cencetti[2], Duccio Fanelli[1,3], and Franco Bagnoli[1,3(✉)]

[1] Department of Physics and Astronomy and CSDC, University of Florence, via G. Sansone 1, 50019 Sesto Fiorentino, Italy
franco.bagnoli@unifi.it

[2] Mobs Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, TN, Italy

[3] INFN, sez. Firenze, Florence, Italy

Abstract. In this paper we introduce the concept of intransitiveness for games, which is the condition for which there is no first-player winning strategy, and the second player can statistically win. We show that the game can be converted into a random walk on a graph, i.e., a Markov process, and therefore we can extend the intransitiveness concept to such systems.

The end of the game generally consists in the appearance of a pattern chosen by one of the player. In the language of random walk this corresponds to an absorbing trap, since once that the game has reached this condition the game itself comes to an end. Therefore, the intransitiveness of the game can be mapped into a problem of competition among traps.

We analyse in details this problem for the Penney game (an extension of the heads or tails game which is intransitive for sequences longer than three), for walks on a circle and for a scale-free network, reminiscent of the structure of the world wide web. We also introduce several variations: traps can be partially absorbing, the walk can be biased and the initial distribution can be arbitrary.

We found that the intransitiveness concept can be quite useful for characterizing the properties of a graph, and that the consequences of the above mentioned extensions are not trivial.

1 Introduction

All around the world people compete with each other, in amusement games as well as in real life. We compete for space, visibility, benefits in general. Also animals compete for food and commercial activities compete for the amount of customers. We shall refer to all kind of competitions with the term "game".

In some games there exists an optimal strategy which always wins and some players are smart enough to find it. However, life in general is more complicated and usually (hopefully) we do not have a unique winner and if one wins a hand there can always be someone else who wins in the following match. In game theory this defines a non-transitive game. Typical example is the game rock-scissors-paper: rock beats scissors, scissors beats paper, paper beats rock.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 204–216, 2019.
https://doi.org/10.1007/978-3-030-34770-3_15

Even in this kind of games, if a player knows the competitor's move in advance, he/she can always find a strategy to beat him/her.

In this article we will explore intransitive games in their most general definition and we will analyze the intrinsic characteristics which those systems have in common.

Our analysis starts by the investigation of a mathematical game, called the Penney game [7], which provides a general and simple definition of "globally intransitive" systems. The Penney game can be reformulated as a random walk on a specific network, where the nodes represent the possible choices which the players can bet on. When the betting choices have been made, the selected nodes become targets and can be represented by absorbing traps for the random walker.

In our analysis we will observe that the intransitiveness property is strictly related to the directedness of the graph which represents the physical support of the random walk. This will bring us to explore the simplest example of directed network: a one-dimensional lattice with periodic boundary conditions.

The network topology in general plays a key role in defining the outcome of the competition in the game, as well as the strategic positioning of the trap nodes in order to interfere with the flux of agents to control their destination.

Moreover, the intransitiveness of a system is related not only to the topology of the network that we use to schematize it, but also to the ability of traps to absorb walkers. For this reason we will also explore the effect of partially absorbing traps.

The relation between intransitiveness and absorption capacity is explored at varying the parameters of the system, both in the Penney game and in the one-dimensional lattice. In the last example we will also observe the outcome of the competition at varying the level of asymmetry in the cycle.

These first two models are at the same time simple and representative of the main characteristics of intransitive games. However, the devised formalism is general and we only need the adjacency matrix of the network to fully describe the competitive behavior. Consequently, it can be applied to many different cases, from competition for the best positioning of a shop in a urban network, to the visibility of sites in the Internet network. We explored this last case in the last section of the article, before conclusions, where we analyze an artificial scale-free network in order to simulate the process on a Internet-like topology.

2 Intransitiveness

Binary relations are often characterized by transitivity, a property of a relation $\sim$ defined on a set $\mathcal{I}$, such that

$$A \sim B, B \sim C \Rightarrow A \sim C \qquad \forall A, B, C \in \mathcal{I}. \tag{1}$$

and a relation is not transitive if this condition is not valid for at least a triple. Instead, a relation is intransitive when

$$A \sim B, B \sim C \Rightarrow \neg(A \sim C) \qquad \forall A, B, C \in \mathcal{I}, \tag{2}$$

where "$\neg$" denotes the negation (NOT).

Many games are based on transitive rules, but there are also games with intransitive rules. This is the case of the rock-paper-scissors game: rock wins against scissors, which in turn wins against paper, which in turn wins again against rock. Games like this one are defined "globally intransitive" and are characterized by the fact that, if a player know the strategy of the opponent, he/she has a high probability of winning by choosing the optimal strategy.

In order to quantify the intransitiveness degree of a system, we introduce the victory matrix V and the index of global intransitiveness σ. The first one is defined as a square matrix of the size of the number of possible choices of the players (three for the rock-scissors-paper game) and the generic entry V_{ij} is the probability that choice i wins against choice j.

The index σ is defined as:

$$\sigma = \min_{i=1\dots N}\left\{\max_{j=1\dots N} V_{ij}\right\} - \frac{1}{2}, \tag{3}$$

Indeed, $\max_{j=1\dots N} V_{ij}$ for a specific i corresponds to the winning probability of i in the best scenario where it can win the game. Then, by considering the minimum over index i we find the worst case, i.e. the winning probability of the index i whom best strategy is the most losing. In order to have global intransitiveness this quantity needs to be larger than one half. So, assuming that the second player chooses his strategy after the first one:

- if $\sigma > 0$, the system is globally intransitive and the choice of the first agent can be countered by an optimal choice of the second, who statistically wins;
- if $\sigma = 0$, the two players can "tie", because the system permits at least two equivalent strategies;
- if $\sigma < 0$, the system is transitive and so, if the first agent makes the optimal choice, statistically wins against the second.

3 From Games to Random Walks

It is possible to reformulate a game as random walk on a directed network where the nodes represent the possible choices of the players and the links describe the competitive relations. Let us illustrate it for the Penney game.

3.1 The Penney Game

This game is an extension of coin flip game, in which two players are asked to bet not on a single coin flip, but on a contiguous succession of heads and tails, whose length l is fixed. This gives to each player a choice among 2^l different sequences of heads and tails. The game is won by the player whose sequence will be extracted first.

Every possible succession has the same probability to appear in the first l coin tosses, and considering infinite tosses every succession appears the same number of times. These observations might suggest that betting on a sequence rather

(a)

$$\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(b)

Fig. 1. (a) The graph of the Penney same for sequence length $l = 3$. Every node is a possibly winning sequence, the links go from one sequence to the sequences which can be obtained with a coin flip. Each link is weighted 1/2. (b) Weighted adjacency matrix (Markov matrix) M of the system, where the nodes relative to the sequences are ordered as binary numbers.

than on another is completely equivalent but this is not true and the reason is that the game has a start. Indeed, once the first succession of l coin tosses has been extracted, the second one has the first $l - 1$ symbols in common with the first one, and so on for the following successions. After any succession there exist only two possibilities for the following succession (given by the extraction of head or tail in the subsequent coin toss). This ultimately implies that there exist specific paths in the space of the 2^l successions that one is forced to follow and the starting sequence (which is actually random) represents the initial condition which affects all the following extractions. The Penney game can be therefore represented as a random walk process on a directed network with 2^l nodes, where the directed links connect successions which can be extracted in sequence. The coin extraction process is therefore figuratively represented by the moves of a random walker which travels the network by following the directed links. In Fig. 1 the network and the relative adjacency matrix for the case $l = 3$ are shown.

3.2 Partially Absorbing Traps and Biased Walks

This change of perspective allows to extend the following analysis also to other competitions situations, like shops competing for customers. In this case we can think of the network given by streets, approximating customers as random walkers looking for a certain product. If they enter the first shop selling that product, shops act as absorbing traps, and an interesting question is whether, given a city map, there exists locations that are robust with respect to competing shops ($\sigma < 0$) or not ($\sigma > 0$) [5].

Another example is that of the searches by web crawlers. A robot is downloading web pages searching for some piece of information, by navigating the links (and eventually jumping to other pages, as in the Google algorithm [4,6]). This again is an example of a random walk on a graphs (the web network) with absorbing traps (the information) and one could be interested in where to place it in order to maximise the probability of being found, while minimizing that of being obscured by another page.

It is also interesting to study the effect of partially absorbing traps, which corresponds to distracted players, who do not realize that the winning combination has appeared, or, more realistically, to customers which are not fully satisfied by a given product, which however can be taken into consideration. We shall indicate with ε the absorbency of traps ($\varepsilon = 1$ for fully absorbing traps).

Finally, one can be interested in studying the case of biased walks, which corresponds to biased coins in the Penney game, or to biased walks in towns, and weighted jumps for web crawlers. The biasing parameter will be indicated by p ($p = 1/2$ for unbiased paths).

Therefore, studying σ, we can characterize the competition in systems describable as random walk on network.

4 Random Walks, Traps and the Victory Matrix

Let us consider a random walk process on a network, which can be described as

$$\omega_i(t+1) = \sum_j \frac{A_{ij}}{k_j}\omega_j \equiv \sum_j M_{ij}\omega_j,$$

where $\omega_i(t)$ is the probability of finding a walker on node i at time t, A is the adjacency matrix, $k_j = \sum_i A_{ij}$ is the out-degree of a generic node j and M is resulting stochastic matrix of our Markovian process.

The victory matrix element V_{ij} is given by the number of walks arriving at node i without passing through node j. The simplest way to do it is to consider nodes i and j as traps, so that we are sure that, after a fairly large number of temporal steps, every possible path of the casual agent will end in one of the two nodes.

We have to keep on mind that the victory matrix also depends on the initial distribution $\omega(0)$. While for games like the Penney one it is natural to start from a uniform distribution (at beginning no coin is shown), there are variations (like the Texas Hold'em poker) in which one has a previous information. Also for walkers in city and web crawlers, it may happen that the starting point is predefined.

4.1 Victory Matrix for Fully Absorbing Traps

In order to compute V_{ij} for fully absorbing traps, it is necessary to count the number of paths ending at i with respect to all possible paths, and similarly for j. Thus we need to study the asymptotic behaviour of the probability distribution of occupation of nodes, $\tilde{\omega} = \omega(\infty)$.

This distribution evolves over time by the iteration of a modified stochastic matrix, which takes into account the presence of the absorbing nodes i and j. In formulae,

$$\tilde{\omega} = \lim_{t\to\infty} \left(M^{[i,j]}\right)^t \omega(0), \tag{4}$$

where the matrix $(M^{[i,j]})$ is obtained modifying the stochastic matrix of the graph relative to the system in analysis, M, by setting the elements of columns i and j equal to zero, except the diagonal elements $(M^{[i,j]})_{ii}$ and $(M^{[i,j]})_{jj}$, which are set equal to one so that, when the casual walker arrives on nodes i and j, he/she stays there during the following temporal steps with probability one.

We can introduce the trap matrix $T^{[ij]}$ with is zero except for $(T^{[i,j]})_{ii} = (T^{[i,j]})_{jj} = 1$, and thus

$$M^{[i,j]} = M(\mathbb{I} - T^{[i,j]}) + T^{[i,j]}.$$

If the matrix M is not singular (all nodes are connected), after a large number of multiplications, only the rows i and j of $\tilde{\omega}$ are non-zero.

In any case, we can define

$$V_{ij} = \frac{\tilde{\omega}_i}{\tilde{\omega}_i + \tilde{\omega}_j}.$$

Actually, it is possible to find an analytic expression for the victory matrix element V_{ij}, given by

$$V_{ij} = \sum_{l,k} M_{il} (\mathbb{I}_N - \mathbf{M}^{[i,j]})^{-1}_{lk} \omega_k(0). \tag{5}$$

This calculation method of V_{ij} is surely advantageous, but the benefit to use the first described one comes from the fact that is easily extensible to the case of partially absorbing traps.

4.2 Victory Matrix for Partially Absorbing Traps

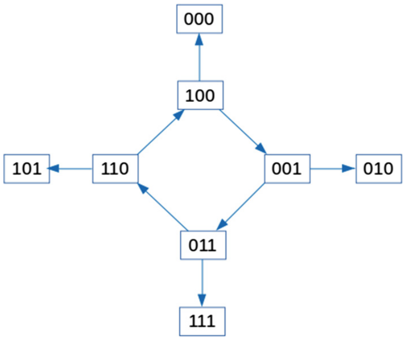
In the case of partially absorbing traps, we have to modify the algorithm, using two distributions, ω as above for the walkers that stay on the lattice (not fallen in the traps) and τ for the traps. Their evolution is given by

$$\begin{cases} \omega(t+1) & = M(\mathbb{I} - \varepsilon T^{[i,j]})\omega(t), \\ \tau(t+1) & = \tau(t) + \varepsilon T^{[i,j]}\omega(t), \end{cases}$$

where ε is the absorptivity of traps. For $\varepsilon = 1$ this expression reduces to that of Eq. 4.

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{2}{5} & \frac{2}{5} & \frac{1}{8} & \frac{5}{12} & \frac{3}{10} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{2}{3} & \frac{2}{3} & \frac{1}{4} & \frac{5}{6} & \frac{1}{2} & \frac{7}{10} \\ \frac{3}{5} & \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{3}{8} & \frac{7}{12} \\ \frac{3}{5} & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{3}{4} & \frac{7}{8} \\ \frac{7}{8} & \frac{3}{4} & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{3} & \frac{3}{5} \\ \frac{7}{12} & \frac{3}{8} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} & \frac{3}{5} \\ \frac{7}{10} & \frac{1}{2} & \frac{5}{8} & \frac{1}{4} & \frac{2}{3} & \frac{2}{3} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{10} & \frac{5}{12} & \frac{1}{8} & \frac{2}{5} & \frac{2}{5} & \frac{1}{2} & 0 \end{pmatrix}$$

(a)



(b)

Fig. 2. Victory matrix (a) and graph (b) for the Penney game with $l = 3$.

5 Numerical Results

5.1 Intransitiveness in the Penney Game

Let us consider the Penney Game for sequences of length $l \geq 3$ (Fig. 1), whose victory matrix is reported in Fig. 2.

Every column has an element bigger than 1/2, and this means that every sequence is statistically beaten by at least another one. This indicates that the Penney game for $l = 3$ is a globally intransitive system, and this is true also for larger sequence lengths.

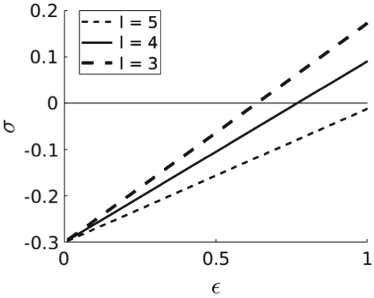
The only exception is the case $l = 2$, where $\sigma = 0$, because the sequences 10 and 01 have the same winning probability against each other.

In this case, if we consider the competition between 01 and 11 (or 10 and 00 by symmetry), the second is statistically favoured: as soon as 0 is extracted before 1, the sequence 01 has already won against 11. The sequence 11 wins just in case two 1 are extracted in the first two coin tosses, and this situation happens with probability 1/4. So, 01 has a 3/4 chance of winning against 11.

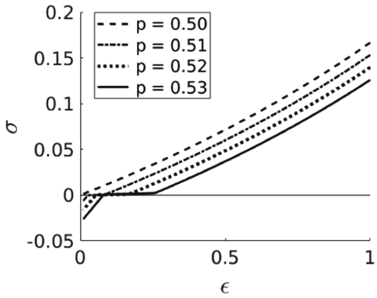
When a system is globally intransitive, we can illustrate this property using a particular graph, considering the maximum value that appears in every column of the victory matrix (therefore, for every sequence we consider the one that has the greatest chance of winning against it), Fig. 2. In this case, arrows have a precise meaning: sequences from which the arrows start are those that most likely to beat the arrival sequences. The graph so built takes the name “victory graph”, and in globally intransitive systems is always connected and it is characterized by a loop of nodes within it.

Let us now investigate how this situation modify if we consider less absorbing traps ($\varepsilon < 1$) or a biased coin $p \neq 1/2$).

Assuming the same level of distraction for the two players (the two absorbing traps have same degree of absorbency), we note how, studying the curve of



(a)



(b)

Fig. 3. (a) Index of global intransitiveness σ as function of ε, fixed $p = \frac{1}{5}$, for sequences of different length l. (b) Index of global intransitiveness σ as function of ε, fixed $l = 3$ for several values of p.

σ at varying ε and the length l of the sequences considered, we can observe dominant areas, where Penney Game is transitive and thus there is a sequence that statistically wins against all the others.

As shown in Fig. 3-a, we note that, generally, the index of global intransitiveness parameter σ is an increasing function of the absorbency parameter ε. When the system is not globally intransitive ($\sigma < 0$), as ε increases, the transitivity degree of the system grows. The length size l also plays a key role for the intransitiveness: completely transitive systems, as the Penney game with $p = \frac{1}{5}$, can show a global intransitiveness zone for high values of ε by increasing the length l of the sequences.

Fixing instead the size of the system (Fig. 3-b), studying the behaviour of σ in the neighbourhood of $\frac{1}{2}$, we observe how the system can be globally intransitive $\forall\, \varepsilon$ only when we consider the fair Penney Game, while for biased coin, departing from $p = \frac{1}{2}$, the transitivity zones are always present for small values of the traps absorbency. Clearly, when only heads or tails can come out as results of the coin tosses, the system is fully transitive and $\sigma = -\frac{1}{2}\, \forall\, \varepsilon$.

5.2 Intransitiveness and Competition on Cycles

The Penney Game victory graph of Fig. 2 shows a cycle inside its structure. It is therefore interesting to investigate the "core" of that graph, the cycle itself.

Let us investigate the competition of traps located on a cyclic graph, with walkers that move in a given direction with probability p. As in the Penney Game case, we consider the same level of absorbency for the traps. Generally, we expect competition on a directed cycle with $p = 0$ or $p = 1$ to be the globally intransitive system par excellence: for each node of the network where a trap can be located, there are other nodes (the immediately preceding neighbours, considering the direction in which the agent moves), that can "obscure" it. As an applied example, consider a rotatory square (with periodic boundary conditions) where two or more shops selling the same product compete for customers.

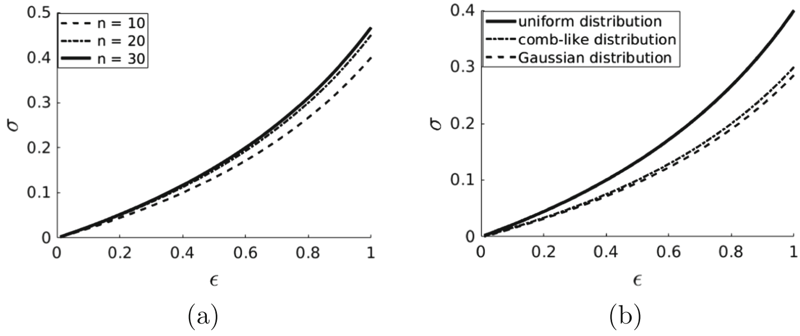


Fig. 4. (a) Behaviour of σ in function of ϵ for a cycle, for different sizes of the system and assuming an initial uniform probability distribution. (b) Behaviour of σ in function of ϵ for a cycle with 10 nodes. A different initial distribution of probability is associated to every type of line. Gaussian distribution has a standard deviation equal to 0.15.

The higher the quality and the price of goods in one shop, the higher its appeal, and consequently its absorbency ϵ.

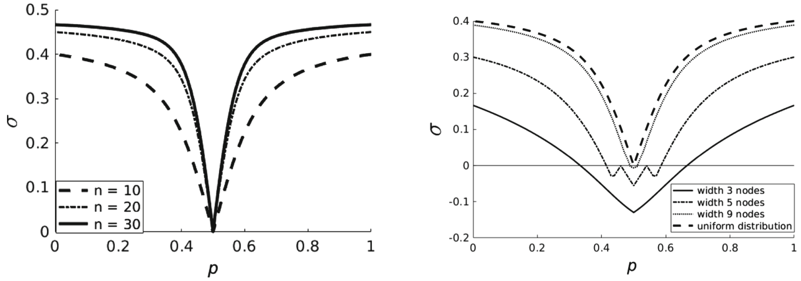
Let us analyse the competition between two equivalent (same value of ε) traps on the cycle, initially assuming a uniform probability distribution for the casual agents on the network. Similarly to the Penney Game, we notice that also the one-dimensional lattice is globally intransitive $\forall\,\varepsilon$, and that increasing the size of the network increases the intransitiveness degree, see Fig. 4-a.

Moreover, if we consider fully absorbing traps ($\varepsilon = 1$), the intransitiveness σ tends to $\frac{1}{2}$ when increasing the size of the system, which corresponds to the case where for each position of a trap on the network, there is at least another position that is able to statistically obscure it.

A similar behaviour of σ is also observed considering different initial probability distributions, assuming that a zone of the cycle is more populated in a neighbourhood of a specific node by the walkers. This can be represented by a Gaussian distribution or by a comb-like distribution, in which nodes with a constant initial density of walkers are alternated with nodes with no walkers.

For high values of ε, considering these initial distribution, we still observe positive values for σ, but smaller in amplitude (Fig. 4-b). The values taken by the index of global intransitiveness decrease because the traps can be located in nodes where the initial density of walkers is higher. For instance, let us compare the comb-like distribution with the uniform one in the case of fully absorbing traps located alternated one by one on the network.

Since $\varepsilon = 1$, considering two traps located at distance two, in two nodes where the density of walkers is identical and non-zero, the first one (in the sense of the cycle) obscures the second one almost completely, catching also all the other incoming walkers. This corresponds to the disposition of traps that allows to one trap the maximum winning probability against any other trap. The second trap in this case only catches the walkers which are already present here, who are more numerous in the case we consider a comb-like distribution than an uniform.



(a) (b)

Fig. 5. (a) Behaviour of σ in function of p, fixed the size of the system and assigned an initial uniform distribution of probability. (b) Behaviour of σ in function of p, for a cycle of 10 nodes, assigned an initial step-like distribution of probability. A different width of the step is associated to every type of line.

This causes the diminution of the amplitude of σ for the comb-like distribution, with this effect that decreases for $\varepsilon \to 0$.

The curve of σ considerably changes if all the walkers are initially placed in a single node. In this case the system is always transitive, regardless of the absorbency of the traps: if one of the latter is located in the node initially characterized by the presence of all the walkers, even if weak, it always exerts a greater attraction for them, who are less likely to be attracted by the other nodes of the network, in which the opposing trap can be located. Clearly, if the absorbency of the trap tends to one, the index of global intransitiveness of the system tends to $-\frac{1}{2}$.

Based on the previous arguments, if we have a distribution of probability equally concentrated on two nodes of the cycle at first, we always obtain $\sigma = 0$. Indeed, placing the traps in those nodes, the casual agents who move on the network are equally shared between the two traps.

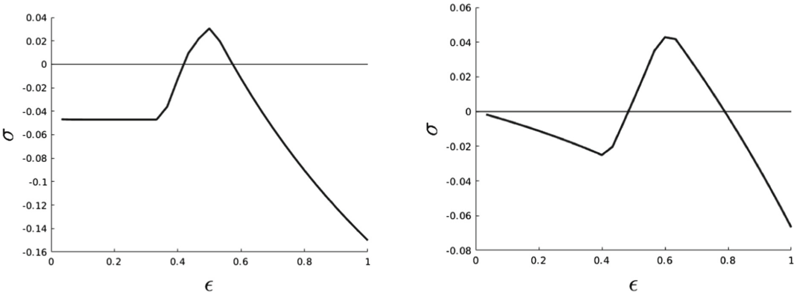
We can also study the competition on the cycle introducing two directions of travel for the walkers, with a symmetry parameter $p \in [0, 1]$. For this purpose, we modify the stochastic matrix $\mathbf{M}$ of the system in the following way,

$$M_p = (1 - p)M + pM^T. \tag{6}$$

When $p = 0$ (or $p = 1$) we still have a directed cycle only viable in one direction. If $p = \frac{1}{2}$, no direction of travel prevails in the system and $\mathbf{M}_p$ is symmetric. While values of p between zero and one half introduce a drift in the system, i.e. the casual agents statistically move clockwise (or anticlockwise).

Let us study the competition on the cycle varying the symmetry parameter p, assuming fully absorbing traps and an initial uniform distribution of probability of the walkers on the network (Fig. 5-a). We observe that, independently on its size, the system is globally intransitive $\forall\, p$.

In addition, we note that, when $p = \frac{1}{2}$, $\sigma = 0$: this result was expected, because a symmetric $\mathbf{M}$ involves that all the possible positions of the traps on



(a) (b)

Fig. 6. (a) Behaviour of σ for a scale-free network with 100 nodes as a function of the absobitiveness ε. There is an intransitive behaviour for intermediate values of this parameter. (b) Similar behaviour of σ for the subnet shown on Fig. 7

the cycle are equivalent. Then, the increase of the size of the system induces in turn an increase of intransitiveness degree: if the network allows only one direction of travel to the walkers, $\sigma \to \frac{1}{2}$, coherently with the behaviour of the index of global intransitiveness shown in Fig. 4-b.

Assuming different initial distribution of probability for the walkers, the variation of p allows to show non-trivial intransitiveness zones. This is the case for instance of an initial step-like distribution of probability (Fig. 5-b), in which the casual walkers are initially focused in a fixed number of adjacent nodes. Such as in the case of the variation of the absorbency ε of the traps, if the step extends over one or two nodes, respectively $\sigma = -\frac{1}{2}$ or $\sigma = 0\ \forall\, \eta$.

The situation becomes more interesting if the step extends at least over 3 nodes (Fig. 5-b). We observe a zone of transitivity around $p = \frac{1}{2}$, that becomes

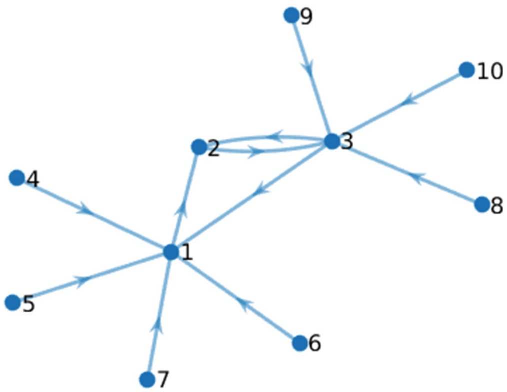


Fig. 7. A small subnet showing the behaviour of σ vs ε as in the scale-free network of Fig. 6.

smaller to the increase of the step's width, until it disappears completely when the step-like distribution becomes an uniform distribution of probability. In these cases, there is a dominant position where the trap can be placed.

In particular, we note that the system is transitive when $p = \frac{1}{2}$ and so $\mathbf{M}$ is symmetric: this fact suggests that the initial distribution of probability of the casual walkers on the network generally plays a key role for the intransitiveness of competition between absorbing traps.

5.3 Competition on the World Wide Web

A meaningful and more complicated applied example where it is important to devise optimal navigation strategies is represented by the World Wide Web. A web surfer crawling the World Wide Web (WWW) network going from one site to another one by following the directed links on the different pages is a perfect example of random walker on a complex network. Specific goals make us devise optimized searching schemes for target nodes.

Empirical observations of the Internet network have given us the knowledge of the specific topological rules that this graph undergoes. It is in particular clear that it is a directed network and that its degree distribution is scale-free [2,3]. Moreover, in real WWW networks, there is a large number of "leaves", pages with links which are not reachable since no other page contains a link to them.

Thus, in order to reproduce the web surfing process, we analysed a network of 100 nodes which has been generated by using the Barabasi-Albert algorithm [1], studying its intransitiveness at varying the absorption index ε. As shown in Fig. 6-a, in this case we have an intransitive region for intermediate values of the absorbancy index ε. By visually inspecting the scale-free network, we have found that this behaviour is given by structures like that reported in Fig. 7, where the large number of leaves may make it intransitive for a partial absorbancy of traps. This is indeed the case, as reported in Fig. 6-b.

6 Conclusions

We introduced the concept of intransitiveness for games, which is the condition for which there is no first-player winning strategy, and the second player can statistically win. Since a game can be converted into a random walk on a graph, we can extend the intransitiveness concept to Markov processes, but this property depends also on the initial distribution of probability.

In the case of random walks and Markov processes, the condition corresponding to the end of a game is that of walkers falling on a trap. Therefore, we have studied the competition among traps on random walks.

We analyse in details this problem for the Penney game (an extension of the heads ot tails game which is intransitive for sequences longer than three), for walks on a circle and for a scale-free network, reminiscent of the structure of the world wide web. We also introduce several variations: traps can be partially absorbing, the walk can be biased and the initial distribution can be arbitrary.

We found that the intransitiveness concept can be quite useful for characterizing the properties of a graph, like that of pages in the Word Wide Web, and that the consequences of the above mentioned extensions are not trivial.

References

1. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47 (2002)
2. Barabási, A.-L., Albert, R., Jeong, H.: Scale-free characteristics of random networks: the topology of the world-wide web. Phys. A Stat. Mech. Its Appl. **281**(1–4), 69–77 (2000)
3. Barabási, A.-L., Bonabeau, E.: Scale-free networks. Sci. Am. **288**(5), 60–69 (2003)
4. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: Seventh International World-Wide Web Conference (WWW), Brisbane, Australia, 14–18 April 1998 (1998)
5. Cencetti, G., Bagnoli, F., Di Patti, F., Fanelli, D.: The second will be first: competition on directed networks. Sci. Rep. **6**, 27116 (2016)
6. Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton (2011)
7. Penney, W.: Problem 95: penney-ante. J. Recreat. Math. **2**, 241 (1969)

A Self-organized Criticality Method for the Study of Color-Avoiding Percolation

Michele Giusfredi[1] and Franco Bagnoli[1,2(✉)]

[1] Department of Physics and Astronomy and CSDC, University of Florence, via G. Sansone 1, 50019 Sesto Fiorentino, Italy
franco.bagnoli@unifi.it
[2] INFN, sez. Firenze, Florence, Italy

Abstract. We study the problem of color-avoiding percolation in a network, i.e., the problem of finding a path that avoids a certain number of colors, associated to vulnerabilities of nodes or links.

We show that this problem can be formulated as a self-organized critical problem, in which the asymptotic phase space can be obtained in one simulation. By using the fragment method, we are able to obtain the phase diagram for many problems related to color-avoiding percolation, showing in particular that results obtained for scale-free networks can be recovered using the dilution of the rule on regular lattices.

1 Introduction

The percolation theory is particularly useful to analyze the flow and the diffusion of certain quantities in many complex systems that can be described as networks, and therefore it has many applications in the Internet science. Classic percolation theory can be used to determine the overall connectivity of complex networks and their resilience to the failure of some nodes, but it cannot treat the heterogeneity of its vulnerabilities. In many real-world networks, one can identify different classes of nodes that share the same vulnerabilities.

For example, one can separate internet routers that run different software versions in different classes. The routers with the same software vulnerable to a certain bug may simultaneously fail and suddenly isolate parts of the network that were connected. In communication networks, vertices can be associated to servers controlled by different companies. Some of them could be interested to eavesdrop information that passes through their servers, which therefore must be avoided for a secure communication. Even if all the companies are dangerous, a secure communication between the source and the receiver can be achieved by splitting messages in pieces by secret-sharing, and sending them through multiple paths, each one avoiding one of the vulnerable classes of nodes. Not always this methods can be applied, because sometimes between two servers this secure paths cannot be found.

The color-avoiding percolation (CAP) theory is a recent generalization of the percolation theory useful to treat these problems [4,6,7]. Every node of

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 217–226, 2019.
https://doi.org/10.1007/978-3-030-34770-3_16

the network is marked with a one or more colors, associated with the possible vulnerabilities. The goal of CAP is to find the optimal redundancy of paths between pairs of nodes that can be used to avoid all colors.

In this paper we show how CAP can be mapped into a self-organized critical (SOC) problem. We studied the critical properties of CAP in a direct lattice using different SOC methods, in particular with the fragment method [1].

2 Color Avoiding Percolation

2.1 Site Color Avoiding Percolation

Let us assign a color $c_i \in \{1, 2, \ldots, C\}$ to each vertex i of a network. The quantity C denotes the total number of colors. Assuming that the sets of nodes with a certain color are disjoint, we can define as $\rho(c)$ the probabilities that a node has the color c, such that $\sum_c \rho(c) = 1$.

We assume that nodes with the same color fail together and that those with different colors unlikely fail at the same time. Nodes carrying some colors can be considered trusted, and it is not necessary avoiding them for sending, receiving or transmitting some information.

We define the set of colors to be avoided as senders/receivers as

$$\mathcal{S} = \{c_{s_1}, c_{s_2}, \ldots, c_{s_R}\} \subseteq \{1, 2, \ldots, C\}$$

and the set of colors to be avoided as transmitters as

$$\mathcal{T} = \{c_{t_1}, c_{t_2}, \ldots, c_{t_T}\} \subseteq \{1, 2, \ldots, C\}.$$

We then define a pair of nodes as color-avoiding connected (CAC) if they have paths between them which avoid all of the untrusted node colors, with each path avoiding a different color (sometimes the same path can be used to avoid more than one color).

Clearly, not every node can be CAC with other nodes of the network. If a node, for example, is linked to only one other node carrying an untrusted color, or only with nodes that have the same vulnerable color, it cannot be part of a CAC pair, because it would remain isolated if that color was removed from the network.

We can also define the color-avoiding giant component (CAGC) as the maximal set of nodes that are mutually CAC, i.e., the largest set of nodes connected to each other and such that any pair of vertices belonging to the set is CAC. A method to extract this set from the network can be found in Ref. [6].

2.2 Bond Color Avoiding Percolation

Similarly to classic percolation, a color avoiding percolation problem can be formulated, in witch vulnerabilities, i.e., colors, are associated to links and not to nodes. In this case it is not necessary to distinguish between colors to be avoided for sending/receiving or for transmitting, because every node can be considered trusted.

Similarly to site color avoiding percolation, two nodes can be defined as pair CAC if they are connected through paths that avoid edges of a certain color, for every untrusted color $c \in \mathcal{T}$, and we can define the CAGC of the network.

2.3 Criticality in CAP

The relative size of the CAGC, i.e., the fraction of nodes of the network belonging to the CAGC, is dependent on the properties of the network under study. In the thermodynamic limit, when the number of nodes $N \to \infty$, certain control parameters must exceed their critical values to allow the formation of a CAGC with a non-zero value.

The CAGC is dependent on the topology of the network. If we consider a randomly generated network, the average connectivity $\bar{k}$ must be greater than a critical value to allow the formation of the giant component. Since CAGC is a subset of the latter, even the color avoiding percolation threshold depends on $\bar{k}$.

Clearly, the relative size of the CAGC is also dependent on the color distribution in the network. If an untrusted color has a high concentration in the network, avoiding it may become difficult. For this reason there exists a critical value ρ_{crit} above which the probability $\rho(c)$ prevents the formation of the CAGC.

3 Self Organized Criticality

In some systems, after defining a local evolution law, certain quantities show a power law, typical of critical phenomena. These are called auto-critical systems (they manifest the so-called “self-organized criticality”), to underline their ability to “go independently” to the critical point, when they are allowed to evolve in time. Known examples of auto-critic systems are the ones described in the evolution model of Bak and Sneppen [3], and the invasion percolation model [8].

Many classic percolation problems can be reformulated as SOC system [5], and the evolution of certain quantities can be used to obtain information on the criticality of the systems, such as the critical values of their control parameters.

The idea of this mapping is similar to the Invasion percolation problem [8], we shall illustrate it for a simple directed site percolation with $\bar{k} = 2$. Let us denote by $s_i(t) \in \{0, 1\}$ the state of a site at time t, where one means wet or infected and zero dry or healthy. We assume that a site can be infected independently, with probability p, by its two neighboring sites, while recovering in one step. This is therefore a SIS (susceptible-infected-susceptible) dynamics. Assuming that the evolution rule is given by the parallel application of the rule

$$s_i(t) = [r_i(t) < p](s_{i-1}(t-1) \vee s_{i+1}(t-1)), \tag{1}$$

where $r_i(t)$ is a random number uniformly distributed in the interval $[0, 1)$, independently extracted for each site and each time step, $\vee$ is the OR operation and $[\cdot]$ is the truth function which takes the value one if $\cdot$ is true, and zero otherwise.

Let us introduce the quantity $p_i(t)$, which is the minimum value of p for which the percolation cluster includes site i at time i. The quantity $s_i(t)$ can therefore be written as $[p > p_i(t)]$. Equation (1) can thus be rewritten as

$$[p > p_i(t)] = [p > r_i(t)]\big([p > p_{i-1}(t-1)] \vee [p > p_{i+1}(t-1)]\big). \tag{2}$$

Now, the expression $[p > a] \vee [p > b]$ can be rewritten as $[p > \min(a, b)]$ and $[p > a][p > b]$ is equivalent to as $[p > \max(a, b)]$, thus Eq. (2) becomes

$$[p > p_i(t)] = \Big[p > \max\big(r_i(t), \min(p_{i-1}(t-1), p_{i+1}(t-1)\big)\Big].$$

We can therefore extract the evolution rule for the $p_i(t)$:

$$p_i(t) = \max\big(r_i(t), \min(p_{i-1}(t-1), p_{i+1}(t-1)\big),$$

which has a sort of magic: automatically (or self-critically) the p_i's evolve so that all are above the critical value p_c, which can be extracted as

$$p_c = \lim_{t\to\infty} \lim_{N\to\infty} \operatorname*{MIN}_i p_i(t).$$

This method is powerful and elegant, but it can be used only if the rule can be expressed using AND and OR operation. It is possible to think to sites as segments initially marking all possible values of p, from zero to one. Since only for $p > a$ the site can be wet, the associated segment can be painted with white (dry) from zero to a, and black above it. At beginning, all segments are black. The AND operation is similar to taking the intersection of the black parts of the segments, and the OR to the union of them. Since these operations keep the black part contiguous, one has only to keep track of the lower boundary, which is exactly $p_i(t)$.

Clearly, if the rule includes also other operations like NOT or XOR, the black part of the segment becomes to fragment, and the evolution cannot be captured by only a value per site. We can however approximate the evolution of a system by iterating segments (now called fragments) sampled at many values of p, using a multi-bit technique. In this way we can compute the parallel evolution of the system for many values of p over the random field determined by the random numbers $r_i(t)$. We can do this also for several parameters at one, iterating multi-dimensional fragments [1].

3.1 The Fragment Method

The fragments in k dimensions are defined as subsets of k-dimensional unit hypercube. A fragment in 2 dimensions, for example, is a subset of the unit square $[0, 1) \times [0, 1)$.

The fragment method can be applied to any system that can be described as a probabilistic cellular automaton (PCA). In a network we can describe the evolution of the states $\sigma(t)$ of the nodes with the local evolution rule

$$\sigma(t + 1) = f(\{\sigma(t)\}; \{[r(t) < p]\}), \tag{3}$$

in which time t is a discrete variable, and f is a function of the set of states

Let's consider a probabilistic system characterized by m control parameters p_j, with $j = 1, 2, \ldots, m$. The fragment method consists in assigning a fragment in m dimensions to every node.

The important aspect of this procedure is that the evolution of the fragments is independent of the values of the control parameters p_j. These parameters are reintroduced in the problem only in the final step of the procedure, to pass from the fragment configuration to the state configuration.

The evolution of the fragments can be seen as the evolution in parallel of different copies of the original system; in particular, the fragments that are associated with the node states can be interpreted as different copies of the phase space. After a small time from the initial conditions, these fragments are very different from it, but, as time progresses, they interact with each other, auto-organize themselves to the critical point, and give with their asymptotic values a good representation of the phase diagram.

4 Applications

In this section we show some results of the SOC methods applied to different PCA problems. We can divide the problems in two cases:

- Independent colors: nodes can have at the same time more than one vulnerability/color
- Exclusive colors: every node can assume only one color (cases studied in Refs. [4,6,7])

All simulations have been made in a direct lattice, in particular in $1+1$ dimensional lattice with N sites, in witch all percolation problems were considered in their directed formulation, i.e. with random variables renewed at each time step. The sites i at time t are connected to k of those of the previous instant $\mathrm{t}-1$.

4.1 Independent Colors

In the case of the independent colors we can consider for each node i at time t a set of states $\{s_i^{(\bar{c})}(t)\}$, in which $s_i^{(\bar{c})}(t)$ is a Boolean variable associated to one of the untrusted color c. We consider $s_i^{(\bar{c})}(t) = 1$ if it is possible to arrive at node i at time t through a path that avoids the color c, and $s_i^{(\bar{c})}(t) = 0$ otherwise. If the nodes have a probability $\rho(c)$ to be colored with the color c, the evolution of the states can be written as

$$s_i^{(\bar{c})}(t+1) = [\rho(c) < r_i^{(c)}(t)] \bigvee_{j:A_{ij}=1} s_j^{(\bar{c})}(t), \tag{4}$$

where time is a discrete variable; $[\cdot]$ is the truth function that takes value 1 if $\cdot$ is true, and 0 otherwise; $r_i^{(c)}(t)$ is a random variable associated with the node i,

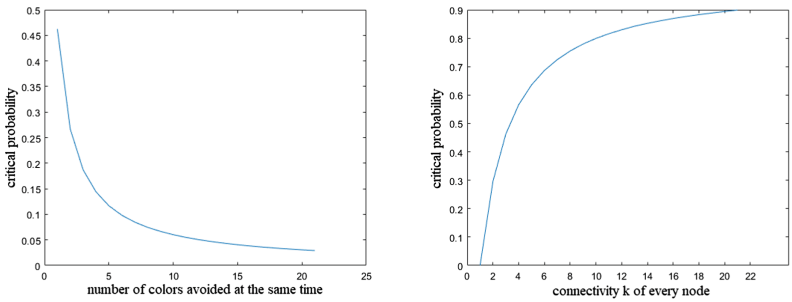


Fig. 1. Trends of the critical probability with (a) the number of colors and (b) the average connectivity $\bar{k}$ of the graph.

evenly distributed between 0 and 1; A_{ij} is the adjacency matrix; $\bigvee_{j:A_{ij}=1}$ is the logical operation OR applied on the states $s_j^{(\bar{c})}(t)$ if the nodes j are connected to node i. The term $[\rho(c) < r_i^{(c)}(t)]$ is a Boolean random variable, with probability $1 - \rho(c)$ to be 1. The logic of the previous equation is that the node i can be reached without passing through nodes of color c at time t + 1 if the node itself has not the color c, and if it's connected with at least a node reachable avoiding the color.

The equation can be rewritten considering $s_i^{(\bar{c})}(t) = [\rho(c) < p_i^{(\bar{c})}(t)]$, where $p_i^{(\bar{c})}(t)$ can be defined as the maximum probability $\rho(c)$ so that the node can be reached avoiding the color c at time t.

From Eq. (4), as shown in Ref. [2], it is possible to obtain an evolution rule for the quantities $p_i^{(\bar{c})}(t)$ as

$$p_i^{(\bar{c})}(t+1) = \min\left(r_i^{(c)}(t), \underset{j:A_{ij}=1}{\mathrm{MAX}}\, p_j^{(\bar{c})}(t)\right), \tag{5}$$

Assuming, for example, that in the initial moment all the states are equal to 1 (that is equivalent of writing $p_i^{(\bar{c})}(0) = 1 \quad \forall i$), the asymptotic values of these quantities can be used to deduce the critical value of $\rho(c)$. The maximum value of $p_i^{(\bar{c})}(t)$ among those of all the nodes, in the limit $t \to \infty$, is the maximum value that $\rho(c)$ can assume so that asymptotically there exists at least one node with state $s_i^{(\bar{c})} = 1$, and therefore reachable with a path that avoids the vulnerability.

For this reason we can say that the quantities $p_i^{(\bar{c})}(t)$ auto-organize to the critical point, and with a single simulation of their evolution they give the critical probability as

$$p_c^{(\bar{c})} = \lim_{t\to\infty} \underset{i}{\mathrm{MAX}}\, p_i^{(\bar{c})}(t). \tag{6}$$

If we search paths that avoid more than one color at the same time, assuming that all the probabilities $\rho(c)$ are equal to each other, we can define the states $s_i^{\mathcal{T}}(t)$, equal to 1 if the node i at time t can be reached avoiding colors the belonging to $\mathcal{T}$, and 0 otherwise. As in the previous case, we can find an evolution rule for the quantities $p_i^{\mathcal{T}}(t)$ as

$$p_i^{(\mathcal{T})}(t+1) = \min\left(\underset{k\in\mathcal{T}}{\mathrm{MIN}}\, r_i^{(k)}(t),\ \underset{j:A_{ij}=1}{\mathrm{MAX}}\, p_j^{(\mathcal{T})}(t)\right), \qquad (7)$$

and from their asymptotic values we can obtain the critical probability of $\rho(c)$ as in Eq. (6). Since all colors are independent from each other, the critical probability to avoid separately any of the colors, as the critical probability to avoid at the same time two or more colors, is independent of the color we are considering. In particular, as it was defined, the probability $p_c^{(\bar{c})}$ to avoid one vulnerability coincides with $1-p_c$, where p_c is the critical probability for directed percolation in the lattice.

Some of the values of the critical probabilities obtained through simulations are listed in Table 1. The values where calculated by varying the number of colors avoided at the same time C, and by varying the connectivity k of all the nodes, in a lattice with 10000 sites at the "asymptotic" time $t = 100000$.

In Fig. 1 we also show the trend of the critical probability by fixing the connectivity $k = 3$ and varying the number of avoided colors in (a), and by fixing $C = 1$ and varying k in (b).

With the fragment method it is also possible to determine the critical probabilities in the case in which the probabilities $\rho(c)$ can be different one from the other.

If every node i has connectivity $k = 3$ and there are 2 untrusted colors b and c, the local evolution rule can be written as

$$s_i^{(\bar{b}\wedge\bar{c})}(t+1) = [\rho(b) < r_i^{(b)}(t)][\rho(c) < r_i^{(c)}(t)]\left(s_{i-1}^{(\bar{b}\wedge\bar{c})}(t) \vee s_i^{(\bar{b}\wedge\bar{c})}(t) \vee s_{i+1}^{(\bar{b}\wedge\bar{c})}(t)\right). \quad (8)$$

We can study the density $m(t) = \frac{1}{N}\sum_i s_i^{(\bar{b}\wedge\bar{c})}(t)$, i.e. the fraction of sites with state $s_i^{(\bar{b}\wedge\bar{c})} = 1$ at time t.

The value that the density assumes after sufficient period of time, depends on the values of the control parameters $\rho(c)$ and $\rho(b)$. We therefore applied the fragment method to obtain the phase plane associated with m, shown in (a) of Fig. 2.

If the values of $\rho(c)$ and $\rho(b)$ are too great, the two colors cannot be avoided and at the same time in the same path, and after some time the density goes to

Table 1. Critical probabilities for given number of colors C and connectivity k.

	$C=1$	$C=2$	$C=3$	$C=4$	$C=5$
$k=2$	0.2958	0.1613	0.1107	0.0840	0.0679
$k=3$	0.4631	0.2666	0.1866	0.1438	0.1165
$k=4$	0.5655	0.3406	0.2416	0.1886	0.1534
$k=5$	0.6356	0.3966	0.2858	0.2235	0.1831

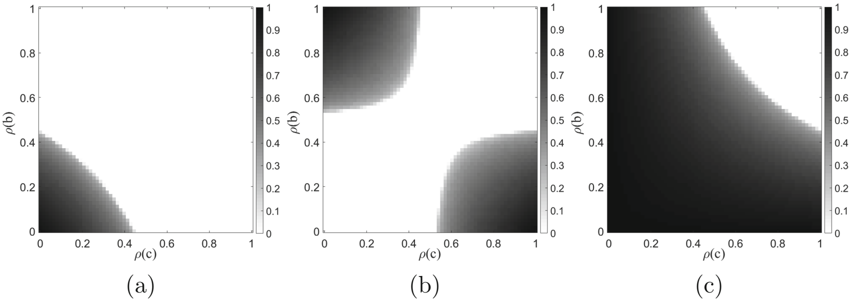


Fig. 2. (a) Phase planes of Eq. (8), (b) the same replacing the AND with a XOR and in (c) replacing the AND with the OR.

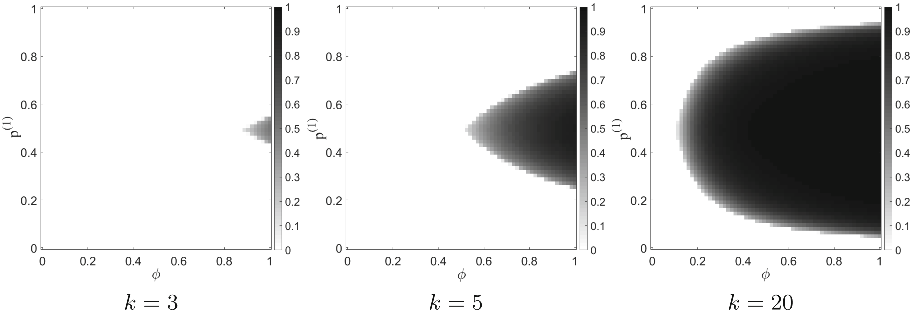


Fig. 3. Phase planes of the bond CAP for exclusive colors and different dilutions (different connectivities).

zero. If instead the concentration of the two colors in the network is lower, after a large period of time it is always possible to find sites reachable with paths that avoid the two colors, hence a non-zero density.

We have also studied other phase planes of the density m modifying the Eq. (8). In (b) is shown the phase plane obtained by replacing the multiplication between the two truth functions (that can be seen as the logical operator AND) with the logical operator XOR. In this case, the states of the nodes can be non-zero only if the two colors are not avoided at the same time. Therefore, when $\rho(c)$ and $\rho(b)$ are both small, and the colors are easy to avoid, the density goes to 0.

In (c) the multiplication was instead replaced with the logical operator OR, and the states are equal to 1 if at least one of the two vulnerabilities is avoided.

The plane phases were obtained with a single simulation of the evolution of the fragments in a lattice with N = 10000, and at the asymptotic time t = 100000.

4.2 Exclusive Colors

In the case of exclusive colors, we studied the bond CAP. We considered a network in which edges are colored with one of two colors, for example red (r) and blue (b). We also considered a dilution of the network with the parameter ϕ, the probability that an edge is not removed from the network. The local evolution rule can be written as

$$\begin{cases} s_i^{(\overline{r})}(t+1) = \bigvee_{j:A_{ij}=1} [\phi > r_{ij}^{(0)}(t)][p^{(1)} < r_{ij}^{(1)}(t)]s_j^{(\overline{r})}(t) \\ s_i^{(\overline{b})}(t+1) = \bigvee_{j:A_{ij}=1} [\phi > r_{ij}^{(0)}(t)][p^{(1)} > r_{ij}^{(1)}(t)]s_j^{(\overline{b})}(t) \end{cases} \tag{9}$$

where the states $s_i^{(\overline{r})}$ and $s_i^{(\overline{b})}$ evolve at the same time with the same random variables $r_{ij}^{(0)}(t)$ and $r_{ij}^{(1)}(t)$, associated to the edges connecting node j to node i. The dilution parameter ϕ is confronted with $r_{ij}^{(0)}(t)$ in the first truth function of the two equations to determine whether the connection is removed or not, while $p^{(1)}$ is the control parameter compared with $r_{ij}^{(1)}(t)$ to decide whether the edge is red or blue.

In this case, the product between the two states $s_i^{(\overline{r})}$ and $s_i^{(\overline{b})}$ is of particular interest. Indeed, the product can be equal to 1 only if the node i is reached through at least two paths, one that avoids the red edges, and one avoiding the blue ones, and it is not necessary to avoid the two colors in the same path. Therefore, imposing as initial condition that the states $s_i^{(\overline{r})}(0)$ and $s_i^{(\overline{b})}(0)$ are equal to 1 for all the nodes, after a sufficient amount of time the product will be non zero only for a fraction of the nodes of the network, that can be interpreted as elements of the CAGC (Fig. 3).

We have studied the value that assumes this product $\mathcal{M} = \frac{1}{N}\sum_i s_i^{(\overline{r})}(t)s_i^{(\overline{b})}(t)$ in the parameter space, as shown in Fig. 3. The phase planes were in this case obtained with a single simulation of the evolution of the fragments in a lattice with N = 10000, at the asymptotic time t = 10000, and varying the connectivity k of the nodes. If the connectivity k is lower than 3, it is not possible to find nodes such that $s_i^{(\overline{r})}(t)s_i^{(\overline{b})}(t) = 1$ at large times, regardless of the values of the control parameters. Instead, for $k \geq 3$ there are certain values of ϕ and $p^{(1)}$ for which $\mathcal{M}$ undergoes a phase transition and passes from 0 to a non null-value. Increasing the value of k, the region in the graph in which $\mathcal{M} \neq 0$ increases its size, and for example with $k = 20$, as shown in Fig. 3, it includes the majority of the phase plane.

These figures are very similar to those reported in Ref. [7], obtained using a Erdős–Rényi network, showing that indeed in these cases the relevant parameter is the average connectivity.

5 Conclusions

We investigated the problem of color-avoiding percolation on a network [4,6,7], i.e., the problem of finding a path that avoids a certain number of colors, associated to vulnerabilities of nodes or links.

We have shown that this problem can be formulated as a self-organized critical problem, in which the asymptotic phase space can be obtained in one simulation, analogous to Invasion percolation [8].

By means of the fragment method [1], we obtained the phase diagram for many problems related to color-avoiding percolation, showing in particular that results obtained for Erdős–Rényi networks can be recovered using the dilution of the rule on regular lattices.

References

1. Bagnoli, F., Palmerini, P., Rechtman, R.: Algorithmic mapping from criticalityto self-organized criticality. Phys. Rev. E **55**(4), 3970–3976 (1997). https://doi.org/10.1103/PhysRevE.55.3970. https://link.aps.org/doi/10.1103/PhysRevE.55.3970
2. Bagnoli, F., Bellini, E., Massaro, E., Rechtman, R.: Percolation and internet science. Future Internet **11**(2), 35 (2019). https://doi.org/10.3390/fi11020035
3. Bak, P., Sneppen, K.: Punctuated equilibrium and criticality in a simple modelof evolution. Phys. Rev. Lett. **71**(24), 4083–4086 (1993). https://doi.org/10.1103/PhysRevLett.71.4083. https://link.aps.org/doi/10.1103/PhysRevLett.71.4083
4. Kadović, A., Krause, S.M., Caldarelli, G., Zlatic, V.: Bond and site color-avoiding percolation in scale-free networks. Phys. Rev. E **98**(6), 062308 (2018). https://doi.org/10.1103/PhysRevE.98.062308. https://link.aps.org/doi/10.1103/PhysRevE.98.062308
5. Grassberger, P., Zhang, Y.C.: Self-organized formulation of standard percolation phenomena. Phys. A **224**(1), 169–179 (1996). https://doi.org/10.1016/0378-4371(95)00321-5. http://www.sciencedirect.com/science/article/pii/0378437195003215, dynamics of Complex Systems
6. Krause, S.M., Danziger, M.M., Zlatić, V.: Hidden connectivity in networks with vulnerable classes of nodes. Phys. Rev. X **6**(4), 041022 (2016). https://doi.org/10.1103/PhysRevX.6.041022
7. Krause, S.M., Danziger, M.M., Zlatić, V.: Color-avoiding percolation. Phys. Rev. E **96**(2), 022313 (2017). https://doi.org/10.1103/PhysRevE.96.022313
8. Wilkinson, D., Willemsen, J.F.: Invasion percolation: a new form of percolation theory. J. Phys. A: Math. Gen. **16**(14), 3365–3376 (1983). https://doi.org/10.1088/0305-4470/16/14/028

A Metric for the Fair Comparison of ISPs

Patrick Hosein(✉) and Shiva Ramoudith

The University of the West Indies, St. Augustine, Trinidad and Tobago
patrick.hosein@sta.uwi.edu, shiva@lab.tt

Abstract. Comparing Internet Service Providers (ISP) can be a daunting task because of the many options available as well as the fact that what is advertised may not actually be delivered. Instead of trying to compare a wide range of options we instead focus on the ability of an ISP to provide one service, namely internet service, at a competitive price. This requires extracting the price to be paid if only internet access was provided (i.e., no phone or TV options) and determining the throughput that the customer actually achieves. We then use this information to estimate the price that the customer would have to pay to achieve a rate of 50 Mbps. This performance metric takes into account the ISP's pricing but it also includes an adjustment based on how well the advertised bandwidth is achieved. We obtain the pricing information directly from the ISP and we obtain the achieved rate for a given advertised rate from customers of the ISP. In order to only take into account data we can trust, we also briefly present a trust framework that we are using for crowd-sourced data. In this framework new users are added to the platform and can contribute data only when at least one of the users already on the platform can vouch for them.

Keywords: Internet Service Provider · ISP pricing · Broadband pricing · Trust model

1 Introduction

The provision of Internet services by Internet Service Providers (ISP) has become quite competitive because of the low returns on investment and the increasing bandwidth demands of consumers [2,3,6,12]. Another issue is how best to handle congestion [11]. In addition, many ISPs have been complaining that content providers such as Netflix and Google are reaping financial benefits while making full use of the infrastructure provided by the ISP. This has led to the whole debate on Network Neutrality [10]. Because of the high level of competition, consumers are always seeking the most cost effective ISP for their homes. Unfortunately this is typically difficult to do because of the many options provided, bundled services such as Cable TV and Telephone [7] and finally the actual performance delivered. The latter is due to the fact that ISPs do not ensure Quality of Service guarantees (e.g., a minimum download rate) but rather specify an upper limit on the rate that the consumer will achieve.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 227–234, 2019.
https://doi.org/10.1007/978-3-030-34770-3_17

In the next section we propose a performance metric that takes into account the various issues described above. We then briefly describe a trust framework that we use to obtain truthful feedback on download rates that people actually receive since this information is used in the proposed metric. Finally we provide an example of how this metric can be used to compare ISP performance.

2 Performance Metric

The proposed metric is based on certain assumptions which we now describe. Traditionally, service providers provide Internet Service, TV Channels and Telephone service (also termed triple play plans). However, in recent years we observe two trends. Firstly, people no longer need a landline phone since they prefer to use their mobile phone. Secondly, content providers such as Netflix [1], HULU, HBO and even the traditional Cable channels such as CNN provide streaming services over the Internet negating the need for cable TV subscriptions. More people are therefore opting to forego their cable TV subscriptions. The net result is that the major focus of a consumer has become good Internet service since this can provide all of their needs. Hence we focus on the provisioning of Internet in our comparison.

The second thing to note is that the need for high upload rates is diminishing since most people now keep their content in the cloud and interact with their data using web-based applications (e.g., Google Docs). Streaming services are one-way (download) with negligible upload throughput required [5]. People do upload or synchronize information such as photos and videos but, if we consider a typical user video of a few minutes, this is quite small compared with a High Definition Netflix movie. With this in mind we focus solely on download rates and in most cases this also reflects upload subscription rates since they tend to be proportional to download subscription rates.

The final issue we address is the fact that ISPs provide subscription plans with maximum limits on the download rates as well as the upload rates. This means that they are not required to provide those rates at all times and hence during times of congestion the provided rates may be significantly less. The achieved rates are therefore what is of interest rather than the advertised rates and we can obtain this from reports from consumers. ISPs that engineer their networks to handle periods of congestion should be more desirable than those who do not.

The proposed metric can now be described. We require three pieces of information from a sample of consumers of the ISP, the name of their ISP I, the maximum download rate S of their subscription plan (even if bundled with other services), and the download rate R that they actually achieve when running a speed test to a server outside of their ISPs internal network. Our objective is to determine how much the consumer would have to pay to achieve a download rate of some fixed value, say 50 Mbps. In order to do this we need the pricing function of the ISP.

For each ISP, I, we can find the cost of each of their pure Internet service plans together with the associated maximum download rate for the plan. Next we find the best regression model fit for these plans which we denote by $P_I(s)$, which is the predicted price of a plan that offers a subscribed rate s. Finally, if we assume the same reduction factor at all speeds, then the subscription rate that would be needed to achieve a download rate of 50 Mbps would be $50S/R$. Therefore the cost required to achieve a download rate of 50 Mbps is given by $P_I(50S/R)$. This is the metric used to compare the various ISPs. Note that companies such as Ookla [8] do provide various ISP performance metrics but one must also include price information when making a comparison for consumers as we do.

3 The Trust Framework

If we can get samples of the information described in the previous section from the various ISPs we can then average over consumers of each ISP to get the average cost to obtain 50 Mbps from that ISP and then use this metric to compare ISPs. However, it is quite easy for consumers to skew the results simply by intentionally providing inaccurate data [4]. We have developed a trust framework to reduce this possibility. A person can only become a member of this trust framework when one or more of the already trusted users in the framework recommends the person. Other restrictions have been put in place to avoid other forms of abuse.

If the users of the framework can be trusted then we can assume that the data they provide is truthful. However, note that users may inadvertently enter incorrect information but the system will still be unbiased since these errors are equally likely to skew the results in either direction. We do admit that the proposed trust framework is not foolproof but our intent is to make it sufficiently difficult to beat the system so that the effort is not worth the reward.

The information required from the survey participant is minimal. We embed an application provided by OpenSpeedTest [9] on the site. The user simply clicks on this application to run the speed test and we automatically collect the required achieved rate as well as the name of the ISP directly from the application. The user must then enter their subscription download rate which is the sole piece of data that is required but which can potentially be manipulated for some personal gain. We do limit this somewhat since the user is only allowed to enter a subscription rate that is offered by their particular ISP. Note that no pricing or plan details are required. Users can resubmit information as often as they wish but we only use the most recent submission for our statistics. A sample of the information displayed from the application is provided in Fig. 1 but for an ISP that was not included in the survey (in order to protect their identities).

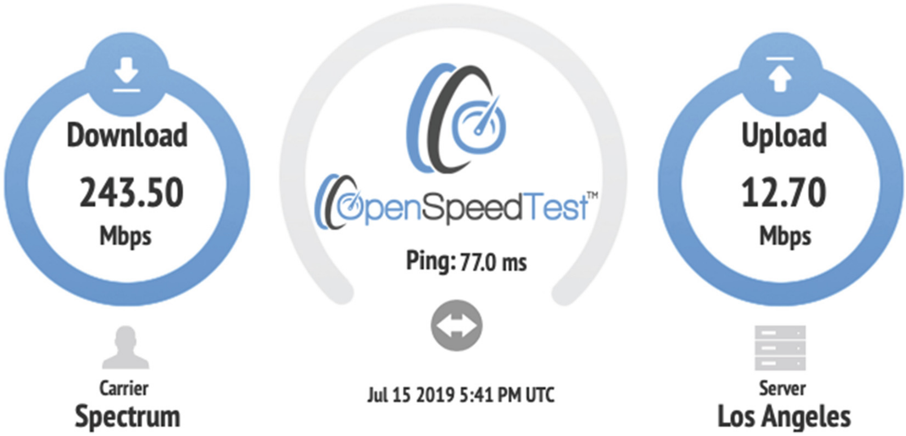


Fig. 1. Illustration of directly collected data

4 ISP Pricing Information

Typically one would expect that price would be a strictly increasing concave function of subscription rate. This is due to the fact that more resources (and hence cost) are required for larger rates but the cost per bps should decrease with increasing rate.

We compare three different ISPs in this study. These as labeled with subscripts A, D and F. For each ISP we determined the pricing for different broadband rates (without additional services such as TV and Phone). Next we determined an appropriate mathematical function to describe the relationship between price and bandwidth in order to obtain the function $P_i(s)$. These functions were found to be as follows:

$$P_A(s) = -319.6 + 177.5 \ln(s) \quad (1)$$

$$P_D(s) = 271 + 0.333s + 0.0095s^2 \quad (2)$$

$$P_F(s) = 182.5 + 1.72s - 0.000617s^2 \quad (3)$$

We plot the data for each ISP as well as the corresponding regression curve in Fig. 2. The y axis represents price (in the currency of the particular country) while the x axis represents the bandwidth or rate in Mbps. In the case of ISP A we increased the weighting of the error at $s = 50$ when doing the regression so as to increase the accuracy of the approximation at this point. We find that the functions provide a good approximation around $s = 50$. The functions for two of the ISPs are concave and strictly increasing as expected. However, the function for ISP D is convex rather than concave. We do not have an explanation as to why the ISP decided to use such a pricing model.

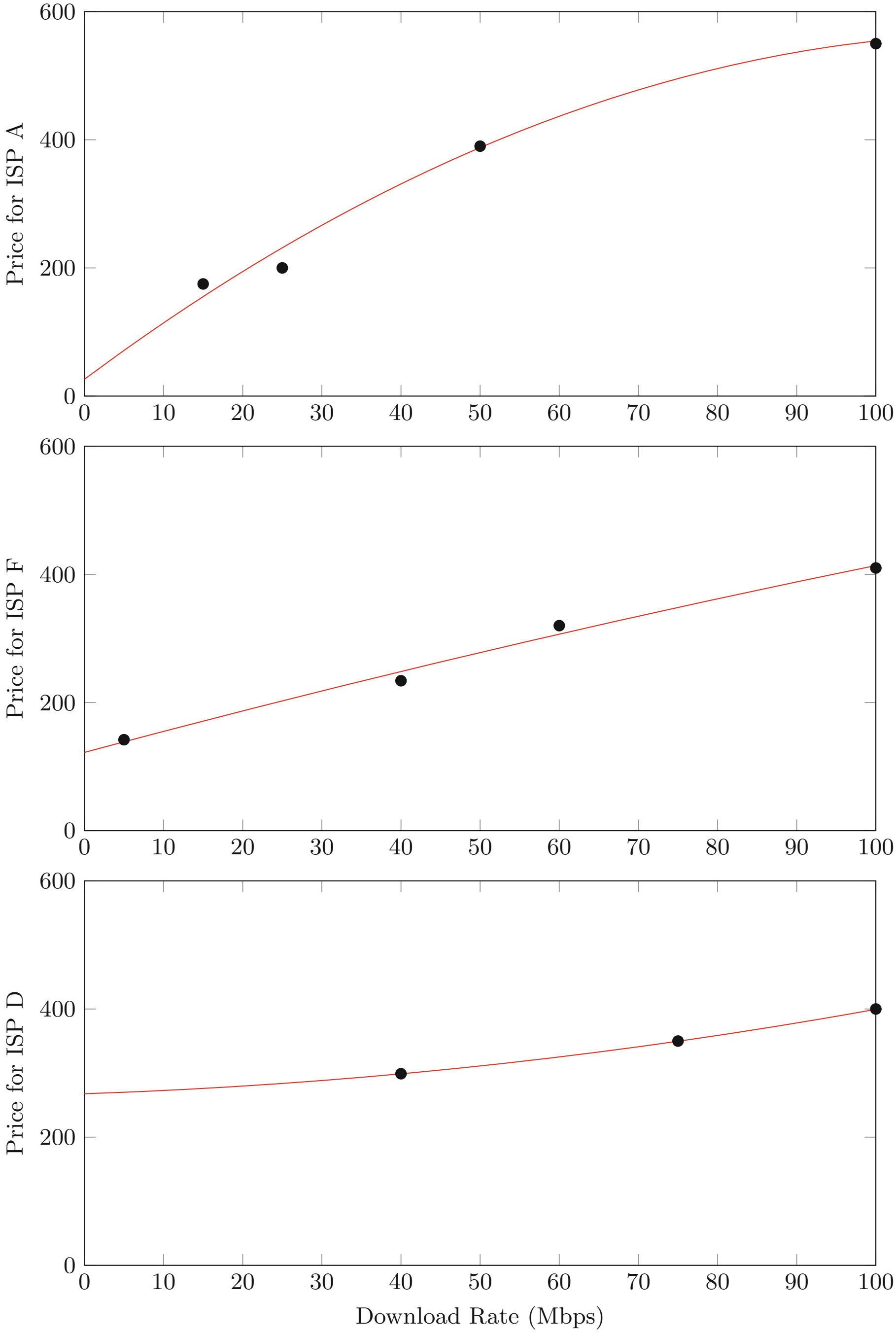


Fig. 2. ISP pricing models

5 Numerical Results

In this section we provide results from samples provided by consumers using our trust platform. Trusted users were asked to run an application on the trust platform that determines download rates and other metrics. They were then asked to provide their download subscription rate. The performance metric (estimated cost to achieve 50 Mbps) described above was computed and the average μ, over all users for each ISP was computed. The proposed metric is presented in Table 1. We also include the approximate cost of subscribing to a plan for 50 Mbps based on the subscription pricing of the ISP. We define *Fulfillment Ratio* as the ratio of the achieved rate and the subscription rate. In some cases one may actually achieve a rate higher than the subscription rate but in general it is lower and so typically the fulfillment ratio is less than 1. We computed the average fulfilment ratio to better understand the service provided. Finally, if an ISP provided the same fulfillment ratio to all consumers then all computed performance metric values would be the same for that ISP. Therefore the variation of the performance metric indicates the level of fairness provided by the ISP since it reflects the level of variation of service cost across consumers. We use the Coefficient of Variation, which is the ratio of the Standard Deviation, σ and Mean μ, of the performance metric to determine fairness. Another useful comparison metric would be the cost of achieving 50 Mbps with high probability. We do this by adding two standard deviations to the mean. If the metric had a Normal distribution then this would mean that 95% of the time the consumer would achieve at least 50 Mbps. However, since this is not the case then we cannot guarantee that this probability would be 95% but it will be high. The most desirable metric value is highlighted in bold.

Table 1. Various statistics of consumer ISP reports

	ISP A	ISP D	ISP F
Mean of performance metric (μ)	336.24	377.89	**302.35**
Cost of a 50 Mbps subscription	390	311	**278**
Average fulfillment ratio	**1.23**	0.84	0.90
Coefficient of variation (σ/μ)	**0.109**	0.379	0.137
50 Mbps with high probability ($\mu + 2\sigma$)	409.30	664.51	**385.13**
Percentage of respondents	6	43	**51**

We can note the following. According to the proposed performance metric, ISP F is the most cost effective followed by ISP A and finally ISP D. The pricing of each ISP plays a role but of even greater importance is the fulfillment ratio. ISP D has the worst fulfillment ratio and hence is not always providing what is promised and this results in a high value for their metric. By looking at the Coefficient of Variation we see that ISP D is also the most variable in terms of provided service and hence is less fair across its users. If we take this into account and determine what price would be required to ensure, with very high probability, that 50Mbs is achieved then we get the numbers in the second to last row. We find that ISP F is better than ISP A followed by ISP A. Note also that ISP F is also the most popular (more than half of the submissions). Although this is technically a better metric, we decided to use mean since more people are familiar with averages.

6 Conclusions and Future Work

We consider the issue of providing a fair comparison of Internet Service Providers in order to allow consumers to make wise choices. We addressed the major issues involved in doing such a comparison namely, bundling, the fact that the achieved rate is less than the subscribed rate and the different pricing models followed. We developed a single performance metric and then illustrated the utility of this metric with data for three ISPs. We have developed a platform, using the presented trust model, for continuously collecting and displaying consumer data.

References

1. Adhikari, V.K., et al.: Unreeling Netflix: understanding and improving Multi-CDN movie delivery. In: 2012 Proceedings IEEE INFOCOM, pp. 1620–1628. IEEE (2012)
2. Biggs, P., Kelly, T.: Broadband pricing strategies. Info **8**(6), 3–14 (2006)
3. Cao, X.R., Shen, H.X., Milito, R., Wirth, P.: Internet pricing with a game theoretical approach: concepts and examples. IEEE/ACM Trans. Netw. **10**(2), 208–216 (2002)
4. Cho, J.H., Chan, K., Adali, S.: A survey on trust modeling. ACM Comput. Surv. (CSUR) **48**(2), 28 (2015)
5. Constantiou, I.D., Courcoubetis, C.A., et al.: Information asymmetry models in the internet connectivity market. In: 4th Internet Economics Workshop, Berlin (2001)
6. He, H., Xu, K., Liu, Y.: Internet resource pricing models, mechanisms, and methods. Netw. Sci. **1**(1–4), 48–66 (2012)
7. McManus, B., Nevo, A., Nolan, Z., Williams, J.W.: Steering incentives and bundling practices in the telecommunications industry. Available at SSRN **3267060** (2018)
8. Ookla: The world standard in internet metrics. https://www.ookla.com/
9. OpenSpeedTest: Speed test - HTML5 internet speed test. http://openspeedtest.com/
10. Ramneek, Hosein, P., Seok, W., Kwak, J., Ni, A.: Emerging network technologies and network neutrality conformance. In: 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 31–36, September 2017. https://doi.org/10.1109/APNOMS.2017.8094174

11. Sen, S., Joe-Wong, C., Ha, S., Chiang, M.: A survey of smart data pricing: past proposals, current plans, and future trends. ACM Comput. Surv. (CSUR) **46**(2), 15 (2013)
12. Stiller, B., Reichl, P., Leinen, S.: Pricing and cost recovery for internet services: practical review, classification, and application of relevant models. Netnomics **3**(2), 149–171 (2001)

Network Presentation of Texts and Clustering of Messages

Andrey V. Orekhov[1(✉)], Alexander A. Kharlamov[2,3,4], and Svetlana S. Bodrunova[1]

[1] St. Petersburg State University, 7/9, Universitetskaya embankment, St. Petersburg 199034, Russian Federation
a_v_orehov@mail.ru, s.bodrunova@spbu.ru

[2] Institute of Higher Nervous Activity and Neurophysiology of Russian Academy of Sciences, 5A Butlerova Street, Moscow 117485, Russian Federation

[3] Moscow State Linguistic University, 38 Ostozhenka Street, Moscow 119034, Russian Federation

[4] Higher School of Economics, 20 Myasnitskaya Street, Moscow 101000, Russian Federation
kharlamov@analyst.ru

Abstract. For the purposes of searching for various communities on the Internet, automatic typology of text messages defined via application of methods of cluster analysis may be used. In this paper, we address one of the significant issues in text classification via cluster analysis, namely determination of the number of clusters. For clustering based on semantics, text documents are typically represented in the form of vectors within n-dimensional linear space. What we suggest as a method for determining the number of clusters is the agglomerative clustering of vectors in the linear space. In our work, statistical analysis is combined with neural network algorithms to obtain a more accurate semantic portrait of a text. Then, using the techniques of distributive semantics, mapping of the derived network structures into a vector form is constructed. A statistical criterion for the completion of the clustering process is derived, defined as a Markovian moment. By obtaining automatic partitioning into clusters, one can compare texts that are closest to the centroids with actual content samples or evaluate such texts with the help of experts. If the display of texts in a vector form is adequate, all informational messages from a fixed cluster have the same meaning and the same emotional coloring. In addition, we discuss a possibility to use vector representation of texts for sentiment detection in short texts like search engines input or tweets.

Keywords: Social network analysis · Semantic network · Neural network algorithms · Distributive semantics · Cluster analysis · Least squares method · Markov moment

In the part of method assessment, the work is supported by the Presidential grants of the Russian Federation for young Doctors of Science, grant MD-6259.2018.6.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 235–249, 2019.
https://doi.org/10.1007/978-3-030-34770-3_18

1 Introduction: Basic Definitions and Technological Background

Text typologies have become a wide area of interdisciplinary research, especially for user-generated texts from social media. In particular, it may be used to detect online communities united by discursive features, including opinion, sentiment, or just topicality. Text clustering is one of the most developed instruments for forming a text typology, but the issue of pre-setting the number of clusters remains crucial for methods as different as, e.g., topic modelling for longer [1] and shorter texts [2] and vector-based models of clusterization. In this paper, we suggest a method for calculation of the number of clusters that is based on a combination of neural network algorithms with statistical analysis.

First, we would need to operationalize the terminology that we use further in the text.

A linguistic unit (or language unit) is an element of academic abstraction or language-based modeling that exists on a certain level of a language. The fundamental property of a linguistic unit is that it is indivisible within the given level of a language but it can divide into units of lower levels or unite into units of higher levels. For our analysis, we will use two types of linguistic units, namely words and pair syntagmas.

A word is the basic unit of language and speech. The word consists of morphemes, and itself is a part of phrases and sentences [3,4].

A syntagma is a collection of several words united based on grammatical compatibility. For our paper, only two-word syntagmas will be considered [5], but the method we propose can be further developed to longer syntagmas. Words and syntagmas are met in text corpora.

A (text) corpus is a set of texts selected and processed according to specific rules to be used for the study of a language. The properties of text corpora include them being electronic (they must exist in an electronic form), representative, marked-up (with the help of tags), and pragmatically oriented (created for a specific task) [6,7].

Statistical methods for the selection of language units are diverse. One well-proven method is based on artificial neural networks. These networks are built from neural-like elements with a temporal summation of signals, to form a statistical portrait of the text. With the help of Hopfield neural networks [8,9], can make up a semantic renormalization of weights in the text [10,11]. As a result of the text analysis, the network of basic concepts and their weight characteristics is automatically extracted from it. As a semantic portrait of a text, not just a list of keywords is considered but a network of concepts, which is a set of keywords or stable phrases related to each other. Each syntagma gains some weight which reflects the significance of this syntagma in the text [12,13]. The relationship between syntagmas also casts some influence upon weight assignment. The use of these relationships, or links, allows weighting the concept of text more accurately [14,15].

This method is implemented in TextAnalyst semantic analysis technology independent from the language and subject area [12,13,16]. An essential feature of this approach is the ability to automatically establish the relationship between the identified critical concepts of the text. When identifying links, the statistics of the pairwise appearance of words in the semantic fragments of the studied text are taken into account [17]. Further, statistical indicators of keywords are recalculated into semantic weights taking into account similar characteristics of other critical concepts associated with them, as well as numerical indicators of links.

After recalculating the statistical characteristics of text into the semantic ones, fundamental concepts that are not relevant to the structure of the text gain low weight, and the most representative are endowed with high rank. The resulting semantic network allows performing various types of textual analysis. It reflects the internal structure of the text and the significance of the selected key concepts, and also shows the degree of relatedness of the concepts in the text.

2 The Binary Relation of Associativity and the Construction of Dictionaries

Semantic analysis is a stage in the sequence of actions of the automatic text analysis algorithm which consists of identifying semantic relations and forming a semantic representation of texts. In general, the semantic representation is a graph (semantic network) reflecting the binary connections between two semantic units of the text. In our case, the meaning-bearing units are words, and binary relations are carried by pair syntagmas.

We will now turn to the formal description of a homogeneous and associative semantic network [15]. Will assume that are dealing with some pragmatically oriented corpus of texts X, composed of messages $\overline{x}_i$ of some social network, where

$$X = \bigcup_{i=1} \overline{x}_i.$$

Definition 1. *A first level vocabulary G_1 (elementary dictionary) is a linearly ordered alphabetical set of words from the corpus of texts X, where each word associated with a natural number equal to the number of occurrences (frequency) of the word in the corpus X. Words from dictionary G_1 will be denoted as g_i and their frequencies as w_i.*

Consider the Cartesian product $G_1 \times G_1$.

Definition 2. *The binary relation of associativity $R \subset G_1 \times G_1$ will be the set of all paired syntagmas from the corpus of texts X. For ordered pairs $(g_i, g_j) \in R$, we will use a special notation: $< g_i, g_j >$.*

Remark 1. It's clear that if all ordered pairs of R are syntagmas, then there are no cases when the first word in the pair is the last in the previous sentence, and the second word is the first in the next sentence.

Remark 2. The binary relation of associativity has nothing to do with the algebraic property of associativity.

Any binary relation R can be represented as a directed graph [18]; in our case, each ordered pair $< g_i, g_j >$ from R is a minimal oriented subgraph with two vertices g_i and g_j. The left element of g_i of an ordered pair is called the main word of a syntagma, the right element of g_j is the word-associate, which is the semantic feature of the main word g_i. The word-associate g_j is separated from the vertex g_i by one link. The connection (edge of a directed graph) is directed from the main word to the word-associate, that is, from the vertex g_i to the vertex g_j.

Definition 3. *An asterisk $< g_i, < g_j >>$ is a directed subgraph that includes all ordered $< g_i, g_j >$ pairs where the index i for main word g_i is fixed, and the index j of words-associates g_j runs through all its possible values.*

That is, the subgraph asterisk $< g_i, < g_j >>$ consists of the main vertex g_i and all its vertices-associates g_j. The asterisk edges are directed from the main words to the words-associates. The number of edges in the asterisk subgraph is equal to the number of vertices-associates. The set of all g_j, the vertices-associates of the asterisk subgraph $< g_i, < g_j >>$, is called its periphery.

Definition 4. *The semantic network N for the corpus of texts X is the set of all asterisks $< g_i, < g_j >>$ composed of the words of the dictionary G_1. That is, a semantic network is a directed graph consisting of all possible subgraphs $< g_i, < g_j >>$, made up of words from the G_1 dictionary.*

Definition 5. *The second-level vocabulary of G_2 is the set of paired syntagmas $< g_i, g_j >$, linearly ordered. Alphabetically, it first conciders the main words g_i and then the words-associates g_j.*

The most crucial moment of building a semantic network is determining the weights of paired syntagmas from the dictionary G_2. The complexity of solving this problem is as follows. Initially, weights W_i of the first vertices of g_i for the asterisks $< g_i, < g_j >>$ is calculated as the sum of the frequencies w_j of all the words-associated with g_j, which are the periphery of this asterisk. However, some principal vertices may be associates for other principal vertices. Thereat, their frequencies w_i from the first dictionary G_1 are replaced by the newly calculated weights W_i. Therefore, the weights of the primary vertices must be recalculated several times (from 5 to 10). Moreover, only after the end of the recalculation of the weights W_i one can proceed to the determination of the weights of syntagmas from the dictionary G_2. For uniformity of the record and its brevity, we will call the frequency w_j of a word-associate its weight. If the central vertex of some asterisk is a word-associate in another asterisk, then its weight is $W_j = w_j$.

Definition 6. *The weight Ω_{ij} of the paired syntagma $< g_i, g_j >$ belonging to the dictionary G_2 is the sum of the weight w_i of the main word g_i and the product* *of the weight* w_j *of the words-associates* g_j *and the weight* v_{ij} *of the edge of the connecting vertex* g_i *and* g_j

$$\Omega_{ij} = w_i + w_j \cdot v_{ij}.$$

The weight v_{ij} *is equal to the number of occurrences of the syntagma* $< g_i, g_j >$ *in the corpus of texts* X.

The dictionary G_2 can be considered as a linearly ordered set of triples $\{g_i, g_j, \Omega_{ij}\}$, where g_i is the main word of a pair syntagma, g_j is a word-associate, Ω_{ij} is the syntagma weight of $< g_i, g_j >$. Since the dictionary G_2 is a linearly ordered and finite set, all its elements can be numbered by natural indices. Therefore, for the syntagms from the G_2 and their weights, we will use another designation: $< g_i, g_j > = < G_2 >_h$ and $\Omega_{ij} = \Omega_h$, respectively.

3 Distributive Semantics and the Vector Representation of Texts

Distributive semantics is a field of linguistics engaged in calculating the degree of semantic proximity between linguistic units by their distribution in the corpus of texts [19]. Each linguistic unit is assigned its context vector. A set of vectors forms a verbal vector space. The semantic distance between concepts expressed by linguistic units is usually calculated as the cosine distance between the vectors of the verbal space

$$\rho(\overline{a}, \overline{b}) = 1 - \frac{\sum_{i=1}^{n} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}}$$

where $\overline{a} = (a_1, a_2, \ldots, a_n)$ and $\overline{b} = (b_1, b_2, \ldots, b_n)$ are vectors between which the distance is calculated.

Within this method, we study an ordered set of universal procedures applied to the texts of a language. This set allows for identifying the primary levels of the linguistic units (phonemes, morphemes, words, or phrases), classifying them, and establishing compatibility relations between them.

As a way of representing a mathematical model, the vector space is used. The latter contains information on the distribution of linguistic units represented as multi-digit vectors that form a linear space. The vectors correspond to linguistic units at different levels of the language, and dimensions correspond to contexts. The coordinates of the vectors are numbers that indicate how many times a given word or phrase has occurred in a given context.

The distributive analysis is widely used to study semantics. Recently, a large number of various distributional-semantic methods and corresponding software have been developed. These software tools allow for automatically comparing contexts and calculating semantic distances between them.

However, when applying distributive-semantic models in real world, some problems arise. One of those is high dimensionality of the vectors that corresponds to the large amount of contexts represented in a text package.

As a consequence, there is a need to use methods that reduce the dimensionality and sparsity of the vector space. At the same time, it is necessary to save as much information as possible from the original vector space.

One of the methods for reducing the dimensionality in distributive-semantic models is the removal of some dimensions following linguistic or statistical criteria. The vector representation is a common name for various approaches to modeling and processing natural language when words or phrases from a particular dictionary are associated with a set of vectors from $\mathbb{R}^n$ [20,21].

For clustering texts from the Internet, we will construct the correspondence between the network representation of the texts and the vectors from $\mathbb{R}^n$. The orthogonal (but not normalized) basis of the linear space is the dictionary G_2. The length of each ort is equal to its weight $\Omega_{ij} = \Omega_h$.

The dimensionality of the constructed linear space is reduced by removing all syntagmas whose weight satisfies the inequality $\Omega_h < M$. The value of the constant M can be chosen experimentally depending on the dimension of the linear span by basis vectors that correspond to the syntagmas of the G_2 dictionary.

The coordinates of the vectors corresponding to the texts $\overline{x}_i$ from the X are calculated as follows. If the syntagma $<g_i, g_j> = <G_2>_h$ is present in the text at least once, then the corresponding coordinate h is numerically equal to the product of the weight of the syntagma Ω_h and the frequency of occurrence $<G_2>_h$ in $\overline{x}_i$. In all other cases, $h = 0$. The distance between the text vectors is measured using a cosine measure.

4 The Agglomerative Algorithm of "Single Linkage"

One of the particular and essential tasks of information retrieval on the Internet consists of clustering documents for the automatic typology of groups of semantically similar texts from a specific pragmatically oriented corpus. The principal difference between clustering and classification is that groups of documents are formed only by pairwise proximity of texts. At the same time, unlike for classification, neither the significant characteristics of these groups of documents nor the number of clusters are known in advance.

An arbitrary clustering algorithm is the map

$$\mathcal{A} : \begin{cases} X & \longrightarrow \mathbb{N} \\ \overline{x}_i & \longmapsto k \end{cases}$$

that maps any element $\overline{x}_i$ from the set X to the only natural number k that is the cluster number to which $\overline{x}_i$ belongs. Thus, the clustering process splits the X into a set disjoint subsets X_h called clusters:

$$X = \bigcup_{h=1}^{m} X_h,$$

where for $\forall\ h, l \mid 1 \leq h, l \leq m : X_h \cap X_l = \emptyset$.

Therefore, clustering sets an equivalence relation on X. Independent representatives of these classes are the vectors called centroids. In the n-dimensional linear space $\mathbb{R}^n$, the coordinates of the centroids are equal to the arithmetic average of the corresponding coordinates of all vectors included in the cluster. If each vector from $\mathbb{R}^n$ identifies with the material point of unit mass, then the centroids can be considered the centers of mass of the clusters [22].

For clustering of texts, a hierarchical agglomerative algorithm of single linkage has been proposed [23,24]. Elsewhere, we have re-interpreted this method as follows [22,25].

Let $X = \{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m\}$ is a sample set (corpus of texts) in which any $\overline{x}_i \in X$ belongs to the n-dimensional linear space $\mathbb{R}^n$.

In the space $\mathbb{R}^n$, a cosine metric is given.

If sample X contains m vectors, then, initially, it is assumed that X is divided into m clusters containing one element at a time. Moreover, the cluster centroids coincide with them: $X_h = \widehat{X}_h$ for each h such that $1 \leq h \leq m$.

Iterations of the $\mathcal{A}$ algorithm that implements the "single linkage" method can be described as follows. The first step of the first iteration of $\mathcal{A}_1$, for the $\mathcal{A}$ algorithm, is the construction of a diagonal matrix of cosine distances between the clusters X_h.

Then we determine the minimum distance between the clusters:

$$F_1 = \min(\rho(X_h, X_l)),$$

where $1 \leq h, l \leq m$ and $h < l$.

Thereafter, the clusters X_h and X_l for which the distance ρ is minimal, are combined into one cluster which we denote as X_1, and its centroid is $\widehat{X}_1$. The clusters X_h and X_l are replaced by the centroid $\widehat{X}_1$. Thus, after $\mathcal{A}_1$, the sample X is split into $m - 1$ clusters.

Without loss of generality, we assume that at the beginning of the g-th iteration of $\mathcal{A}_g$, the agglomerative clustering algorithm $\mathcal{A}$, sample X is divided into p clusters. The first step $\mathcal{A}_g$, is the construction of a diagonal matrix of cosine distances between clusters.

Then, just as with $\mathcal{A}_1$, the minimum distance is determined

$$F_g = \min(\rho(X_h, X_l)),$$

where $1 \leq h, l \leq p$ and $h < l$.

The clusters X_h and X_l, for which the distance ρ is minimal, are combined into a cluster which we denote as X_g. Its centroid $\widehat{X}_g$ has coordinates equal to the arithmetic mean of the corresponding coordinates of all vectors from X_h or X_l combined into X_g. At the end of $\mathcal{A}_g$, the clusters X_h and X_l are replaced by the centroid $\widehat{X}_g$. Thus, after the iteration of $\mathcal{A}_g$ is completed, the sample X is split into $p - 1$ clusters.

The main advantage of the method of "single linkage" lies in its mathematical properties. The results obtained using this method are invariant to monotonic transformations of the similarity matrix; its use is not hindered by the presence of coincident data; compared to other clustering methods, it is highly stable [24].

5 "Approximation-Estimation Test"

An essential problem of cluster analysis is determining the most preferred number of clusters. Defining the end of the clustering process is related to solving this issue. The criteria for determining the number of clusters and the criteria for completing the clustering algorithm depend on each other.

The decision on the moment of completion of the clustering algorithm (determining the preferred number of clusters) can be approached within the theory of sequential statistical analysis [26,27] and the optimal stopping rules [28,29]. If there is no criterion for the completion of the clustering process, then after the $m-1$ iteration of the "single linkage" method, the sample X will merge into one cluster, which is an absurd result.

The set of minimum distances obtained after $m-1$ iteration of the algorithm described above has the form $\{F_1, F_2, \ldots, F_{m-1}\}$. It is linearly ordered from the numerical values of elements: $0 \leq F_1 \leq F_2 \leq \ldots \leq F_{m-1}$. We use this set to derive the statistical criterion for the completion of the agglomerative clustering process that implements the "single linkage" method in the linear space $\mathbb{R}^n$.

When clusters merge or when an isolated point joins one of the clusters, a sharp jump in the numerical value of the minimum distance should occur, which, by common sense, coincides with the moment of completion of the clustering process. Moreover, the character of the increase of the ordered set of minimum distances should change from linear to nonlinear [22,25].

To determine the moment when the nature of the monotonous increase of numerical sequence changes from linear to parabolic, we use the previously constructed "approximation-estimated test" [30,31]. For this, we need to formally define the "linear increase" and "parabolic increase" of a numerical sequence. The approximation nodes for the y_n are ordered pairs (i, y_i), where i is the value of the natural argument, y_i is the corresponding value of the sequence y_n. Since the subscript of the sequence y_n uniquely determines the value of the natural argument, the node of approximation (i, y_i) will be considered identical with the element y_i.

The quadratic approximation error for the $f(x)$ function is the sum of the squares of the differences for the values of the numerical sequence at the approximation nodes and the approximating function with the corresponding argument:

$$\delta_f^2 = \sum_{i=0}^{k-1} (f(i) - y_i)^2.$$

We will distinguish the linear approximation in the class of functions of the form $l(x) = ax + b$ and the incomplete parabolic approximation (without linear term) in the class of functions $q(x) = cx^2 + d$.

Quadratic errors on k nodes for linear and incomplete parabolic approximation will, respectively, be equal to:

$$\delta_l^2(k) = \sum_{i=0}^{k-1} (a \cdot i + b - y_i)^2; \quad \delta_q^2(k) = \sum_{i=0}^{k-1} (c \cdot i^2 + d - y_i)^2. \tag{1}$$

If, in our reasoning, the number of approximation nodes is obvious from the context, then the corresponding quadratic errors will simply be denoted δ_l^2 and δ_q^2. When comparing δ_l^2 and δ_q^2, three cases are possible:

$$\delta_q^2 < \delta_l^2; \quad \delta_q^2 > \delta_l^2; \quad \delta_q^2 = \delta_l^2.$$

We say that the sequence y_n has a linear increase if y_n is monotone and the quadratic errors of the linear and incomplete parabolic approximations are related by the inequality: $\delta_q^2 > \delta_l^2$. If the inequality holds true: $\delta_q^2 < \delta_l^2$, then we say that the sequence y_n has a parabolic increase. The node y_{k-1} is called critical if for the set of approximation nodes: $y_0, y_1, \ldots, y_{k-1}$, equality is fulfilled $\delta_q^2 = \delta_l^2$.

We calculate the coefficients a, b by the least squares method for the linear function $ax + b$ and the coefficients c, d for the incomplete quadratic function $cx^2 + d$ that approximate the nodes $y_0, y_1, \ldots, y_{k-1}$.

$$a = \frac{6}{k(k^2-1)} \sum_{i=0}^{k-1} (2i + 1 - k) y_i; \qquad b = \frac{2}{k(k+1)} \sum_{i=0}^{k-1} (2k - 1 - 3i) y_i. \tag{2}$$

$$c = \frac{30}{k(k-1)(2k-1)(8k^2-3k-11)} \sum_{i=0}^{k-1} (6i^2 - (k-1)(2k-1)) y_i; \tag{3}$$

$$d = \frac{6}{k(8k^2-3k-11)} \sum_{i=0}^{k-1} (3k(k-1) - 1 - 5i^2) y_i. \tag{4}$$

In order to determine the moment when the character of increasing of a monotone sequence y_n changes from linear to parabolic, we create an "approximation-estimation test".

We assume by definition that, for the nodes approximation: $y_0, y_1, \ldots, y_{k-1}$, the criterion

$$\delta^2 = \delta^2(k_0) = \delta_l^2(k_0) - \delta_q^2(k_0)$$

fulfills. In this case, we assume that $y_0 = 0$ is always true. It is easy to achieve this condition at any approximation step using the transformation:

$$y_0 = y_j - y_j,\ y_1 = y_{j+1} - y_j,\ \ldots,\ y_{k-1} = y_{j+k-1} - y_j.$$

We calculate, using the formulas (1)–(4), coefficients for functions of the linear and incomplete parabolic approximations over four points y_0, y_1, y_2, y_3.

$$ax + b = \frac{1}{10}(-y_1 + y_2 + 3y_3)x + \frac{1}{10}(4y_1 + y_2 - 2y_3);$$

$$cx^2 + d = \frac{1}{98}((-5y_1 + y_2 + 11y_3)x^2 + \frac{1}{98}(42y_1 + 21y_2 - 14y_3).$$

Then we get an explicit expression for "approximation-estimation test":

$$\delta^2(4_0) = \frac{1}{245}(19y_1^2 - 11y_2^2 + 41y_3^2 + 12y_1y_2 - 64y_1y_3 - 46y_2y_3).$$

It can be said that, near the element y_k, the character of increase for the sequence y_n changed from linear to parabolic, if for nodes $y_0, y_1, \ldots, y_{k-1}$ linear approximation is not worse than incomplete parabolic, that is $\delta^2 = \delta_l^2 - \delta_q^2 \leq 0$, and for a set of points, $y_1, y_2, \ldots, y_k$, shifted by one step of discreteness, the inequality holds $\delta^2 = \delta_l^2 - \delta_q^2 > 0$ [30,31].

6 Applying Statistical Sequential Analysis

Let $T = \overline{1, m-1}$, be a bounded subset of the natural series containing the first $m-1$ natural numbers (we note right away that m can be arbitrarily large). Then the indexed family $\xi = \{\xi_t, t \in T\}$ of random variables $\xi_t = \xi_t(\omega)$ given for $\forall\, t \in T$ on the same probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is called a discrete random process.

Each random variable ξ_t generates a σ-algebra, which will be denoted as $\mathcal{F}_{\xi_t}$. Then the σ-algebra generated by the random process $\xi = \{\xi_t, t \in T\}$ is the smallest σ-algebra containing all $\mathcal{F}_{\xi_t}$ — that is,

$$\sigma(\xi) = \sigma\left(\bigcup_{t=1}^{m-1} \mathcal{F}_{\xi_t}\right).$$

A discrete random process $\xi = \{\xi_t, t \in T\}$ can be considered as a function of two variables $\xi = \xi(t, \omega)$, where t is a natural argument, ω is a random event. If we fix t, then, as mentioned above, we get a random variable ξ_t, but if we fix a random event ω_0, we get a function of the natural argument t, which is called the trajectory of the random process $\xi = \{\xi_t, t \in T\}$ and is a random sequence $\xi_t(\omega_0)$.

Consider the clustering of a finite set X from the linear space $\mathbb{R}^n$ as a discrete random process $\xi = \xi(t, \omega)$. Fetching the sample X from $\mathbb{R}^n$ is a random event $\omega \in \Omega$. Theoretically, any point $\overline{x} \in \mathbb{R}^n$ can belong to the sampling X; therefore, the σ-algebra from the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ contains $\mathbb{R}^n$, any finite set X from the space $\mathbb{R}^n$, all possible countable unions of such sets, and additions to them. Denote this system of sets as $\mathcal{S}(\mathbb{R}^n)$ and call it "selective σ-algebra".

$$\mathcal{F} = \mathcal{S}(\mathbb{R}^n).$$

The same reasoning is valid for any σ-algebra $\mathcal{F}_{\xi_t}$; therefore, $\sigma(\xi) = \mathcal{S}(\mathbb{R}^n)$. Note that this σ-algebra is "poorer" than the Borel σ-algebra: $\mathcal{S}(\mathbb{R}^n) \subset \mathcal{B}(\mathbb{R}^n)$. Indeed, a countable union of at most countable sets is countable; therefore, $\mathcal{S}(\mathbb{R}^n)$ does not contain intervals.

We will now assess the problem of testing statistical hypotheses H_0 and H_1. There are two hypotheses H_0: "the random sequence $\xi_t(\omega_0)$ increases as linear", and H_1: "the random sequence $\xi_t(\omega_0)$, increases nonlinearly". H_0 will be called the null hypothesis, H_1 will be called the alternative hypothesis.

It is necessary to construct a criterion as a strict mathematical rule which would prove or reject the hypothesis. Any statistical criterion is based on a random sample X. Two cases are possible. In the first one, a sample X is extracted from $\mathrm{I\!R}^n$ simultaneously and has a fixed size. In the second one, when a sample X form at the period, its size is a random variable. In the latter case, statistical sequential analysis and the construction of a stopping time are used [26,27].

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, then the indexed family of σ-algebras $\mathfrak{F} = \{\mathcal{F}_t, t \in T\}$ is called a filtration if for

$$\forall\, i, j \in T \,|\, i < j :\ \mathcal{F}_i \subset \mathcal{F}_j \subset \mathcal{F}.$$

Moreover, if for $\forall\ t \in T$ right $\mathcal{F}_t = \sigma(\xi_i,\, i < t)$, then the filtration is called natural filtration.

The random process $\xi = \{\xi_t, t \in T\}$ is said to be adapted to the filtration $\mathfrak{F}$ if for $\forall\ t \in T :\ \sigma(\xi_t) = \mathcal{F}_{\xi_t} \subset \mathcal{F}_t$. Any stochastic process $\xi = \{\xi_t, t \in T\}$ is an adapted process with respect to its natural filtration. The mapping $\tau :\ \Omega \longrightarrow T$ is called Markov moment or stopping time (with respect to the filtration $\mathfrak{F}$) if for $\forall\, t \in T$ the preimage of the set is $\{\tau \leq t\} \in \mathcal{F}_t$ [27,29].

In other words, let τ be the moment of occurrence of some event in the random process $\xi = \{\xi_t, t \in T\}$. If for $\forall\ t_0 \in T$. One can definitely say whether the event τ has happened or not, provided that the values of ξ_t are known only in the past (to the left of t_0), then τ is a Markov moment with respect to natural filtration of $\mathfrak{F}$, random process $\xi = \{\xi_t, t \in T\}$. In this case, for a random sequence $\xi_t(\omega_0)$, the natural filtration is adapted with the process that is the selective σ-algebra $\mathcal{S}\left(\mathrm{I\!R}^n\right)$.

In the case of statistical sequential analysis, one can define the Markov moment (or stopping time) of the experiment as the values of t in which the change occurred like an increase in the random sequence $\xi_t(\omega_0)$ from the linear type to the nonlinear type. As such, we reject the null hypothesis H_0 and accept the alternative hypothesis H_1.

Thus, the statistical criterion for stopping the agglomerative clustering process based on the "single linkage" method in the $\mathrm{I\!R}^n$ can be formulated as follows.

Let $\{F_1,\, F_2, \ldots,\, F_k\}$ be a linearly ordered set of minimum distances, and let the set $\{y_1,\, y_2, \ldots, y_k\}$ to be a "trend set" obtained using the transformation $y_i = F_i + q \cdot i$, where q is a "trend coefficient", i is the iteration number of the $\mathcal{A}$ agglomerative clustering algorithm.

The clustering process is considered complete at the k, iteration, if for the nodes $y_{k-3}, y_{k-2}, y_{k-1}, y_k$, inequality $\delta^2 \leq 0$ is true, but for the set of nodes $y_{k-2}, y_{k-1}, y_k, y_{k+1}$, the inequality $\delta^2 > 0$ is true. In other words, the Markov time of stopping the clustering algorithm $\mathcal{A}$ is equal to statistics:

$$\tau(F_1,\, F_2,\, \ldots,\, F_k,\, F_{k+1}) = \min\{k \mid\ \delta^2 > 0\}.$$

7 Possible Application to Short Texts from Social Media

Above, we have suggested a procedure where the Markovian moment helps define the appropriate number of clusters for vector-based text clusterization. But real-world applications of our method may see some challenges that we feel necessary to state in particular, for the studies of short, unstructured, and noisy texts such as tweets or user search queries.

The first problem is that biterm syntagmas do not work in short texts the way they do in longer texts. The well-known problems of abbreviation and non-following the grammar rules disrupt the scholarly expectations of representing the texts as efficient vector spaces and/or networks. Despite this, it has been shown in our earlier research [2,32] that biterm-based algorithms work better than others in probabilistic clustering of tweets, namely topic modelling. For text clusterization based on hard partition clustering like k-means with vector/network representation of tweets, an experimental comparison with other methods is needed. E.g., we need to prove whether the fixation of optimal cluster number works better for short user-generated texts than the bag-of-words approaches which, in theory, may come closer to the nature of the tweets by considering them small collections of lexicons non-connected grammatically.

Also, fixation of the number of clusters with the Markov moment needs to be tested on corpora of varying nature. Thus, in Twitter research, either one tweet or pooled one-user tweets may be used as a unit of analysis, and the results may vary greatly depending on the length and composition of the pools.

Third, the use of the suggested method depends on the choice of the cluster centroids. So far, in the research literature based on the datasets from social media, centroids have been worked with but differently from what we suggest in this paper. E.g., a centroid is seen as a sentence in a large document [33] (which is not suitable for short texts) or a word in a dataset with reduced text dimensionality that deprives words from their relations with other words [34]. But the problem of random centers of clusters remains, just as the problem of defining the number of clusters. To resolve the issue of random centroids, several papers have shown that centroids can be preliminarily assigned depending on the research task, e.g. popular tweets may become centroids (see [35] as an example). We seem to be the first to offer an instrument for the automated detection of centroids based on a neural network; but, for the cases when centroids are to be pre-assigned, the algorithm of calculation of centroids must be re-thought.

8 Conclusion

The task of clustering texts cannot be considered entirely resolved, because the problem of determining the preferable number of clusters is not solved. Most of the research prefers either methods using visual analysis of dendrograms or the k-means method.

Evaluating dendrograms can help in concluding about the preferred number of clusters. However, this approach is heuristic, and the essence of heuristic methods is that they are based on some plausible assumptions, and not on rigorous mathematical conclusions.

The k-means method has three significant drawbacks. First, it does not guarantee the achievement of a global minimum of the total quadratic deviation, but only that of the local minima. Second, the result of clustering depends on the choice of initial centroids, and their optimal choice is unknown. Initial centroids are chosen randomly. Third, and most important, the number of clusters must be specified in advance, which, in practice, turns clustering into a classification.

Two new techniques for automatic text classification are presented in this paper. First, the texts are initially described as associative semantic networks based on paired syntagmas. The subsequent transformation of the network model into the metric space using the methods of distributive semantics allows the implementation of a clustering algorithm, which retains all its advantages, as a tool for automatic typology. Second, an "approximation-estimation test" is derived to determine the time of completion of the clustering algorithm, which allows for the construction of statistics that calculates the Markov moment of its stop.

The suggested innovations are still to be tested against several options of data pre-processing and the choice of centroid, as well as against the results by the bag-of-words approaches, but we see the suggested combination of methods as efficient for a wide range of cases of text clusterization.

Acknowledgements. This research in the part of methods evaluation is supported by the Presidential grants of the Russian Federation for young Doctors of Science, grant MD-5962.2018.6.

References

1. Bodrunova, S.S., Koltsova, O., Koltcov, S., Nikolenko, S.: Who's Bad? Attitudes toward resettlers from the Post-Soviet South versus other nations in the Russian blogosphere. Int. J. Commun. **11**, 3242–3264 (2017)
2. Blekanov, I., Tarasov, N., Maksimov, A.: Topic modeling of conflict ad hoc discussions in social networks. In: Proceedings of the 3rd International Conference on Applications in Information Technology, pp. 122–126. ACM (2018)
3. Chodorow, M.S., Byrd, R.J., Heidorn, G.E.: Extracting semantic hierarchies from a large on-line dictionary. In: Proceedings of the 23rd Annual Meeting on Association for Computational Linguistics (ACL 1985), Stroudsburg, PA, USA, pp. 299–304. Association for Computational Linguistics (1985). https://doi.org/10.3115/981210.981247
4. Katamba, F.: English Words: Structure, History, Usage, 2nd edn. Routledge, London (2005)
5. Chandler, D.: Semiotics for Beginners: Syntagmatic Analysis. Archived from the original on May 12 (2000)
6. McEnery, T., Wilson, A.: Corpus Linguistics, 2nd edn. Edinburgh University Press, Edinburgh (2001)
7. Stubbs, M.: Text and Corpus Analysis. Computer Assisted Studies of Language and Culture. Basil Blackwell, Oxford (1996)
8. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. **79**, 2554–2558 (1982)
9. McEliece, R.J., Posner, E.C., Rodemic, E.R., Venkatesh, S.S.: The capacity of the Hopfield associative memory. IEEE Trans. Inf. Theory **33**(4), 461–482 (1987)

10. Kharlamov, A.A., Yermolenko, T.V., Zhonin, A.A.: Text understanding as interpretation of predicative structure strings of main text's sentences as result of pragmatic analysis (combination of linguistic and statistic approaches). In: Železný, M., Habernal, I., Ronzhin, A. (eds.) SPECOM 2013. LNCS (LNAI), vol. 8113, pp. 333–339. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-01931-4_44
11. Kharlamov, A.A., Yermolenko, T.V., Zhonin, A.A.: Modeling of process dynamics by sequence of homogenous semantic networks on the base of text corpus sequence analysis. In: Ronzhin, A., Potapova, R., Delic, V. (eds.) SPECOM 2014. LNCS (LNAI), vol. 8773, pp. 300–307. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11581-8_37
12. Kharlamov, A.A., Ermakov, A.E., Kuznetsov, D.M.: TextAnalyst komplexniy neurosetevoy analyzator textovoy informacii [Complex Neuronetwork Analyzer of Text Information TextAnalyst]. Vestnik moskovskogo gosudarstvennogo tekhnicheskogo universyteta im. N.E. Baumana. Seriya: Mashinostroenie [Bulletin of Moscow State Tekhnikal University by N.E. Bauman. Series: Mechanik Engeneering], no. 1, 121–126 (1998). (in Russian)
13. Kharlamov, A.A., Ermakov, A.E., Kuznetsov, D.M.: Systema dlya avtomaticheskoy smyslovoy obrabotky textov na osnove neyrostevoy nekhnologii TextAnalyst [System of Automatical Semantic Text Processing Based on Neuronetwork TextAnalyst Technology]. Informatsionniye Tekhnologii [Informational Technology], no. 10, 27–32 (1999). (in Russian)
14. Kharlamov, A.A., Yermolenko, T.V.: Neiyrosetevaya sreda (neuromophnaya associativnaya pamyat) dlya preodoleniya informatsionnoy sloshnosty. Poisk smysla dlya slabostrukturirovannykh massivakh informatsii. Chyast I. Strukturnaya obrabotka informatsii v kore [Neuronetwork Environment (Neuromorphic Associative Memory) for Information Complexity Overcoming. Searching of Sense in Weak Structured Information Arrays. Part I. Structural Information Processing in Cortex]. Informatsionniye tekhnologii [Informational Technology] **21**(11), 814–820 (2015). (in Russian)
15. Kharlamov, A.A., Yermolenko, T.V.: Neiyrosetevaya sreda (neuromophnaya associativnaya pamyat) dlya preodoleniya informatsionnoy sloshnosty. Poisk smysla d slabostrukturirovannykh massivakh informatsii. Chyast I. Obrabotka informatsii v gippokampe [Neuronetwork Environment (Neuromorphic Associative Memory) for Information Complexity Overcoming. Searching of Sense in Weak Structured Information Arrays. Part II. Information Processing in Hyppocampus]. Informatsionniye tekhnologii. [Informational Technology] **21**(11), 883–889 (2015). (in Russian)
16. Kharlamov, A.A.: Tekhnologiya avtomaticheskogo smyslovogo analiza textov TextAnalyst [Automatical Semantic Text Analysis TextAnalyst]. Vestnik moskovskogo gosudarstvennogo universiteta. Gumanitarnye nauki [Bulletin of Moscow State Linguistic University Human Sciences], no. 699, 234–244 (2014). (in Russian)
17. Geger, A.E., Geger, S.A.: Otechestvennye komputernye progremmy dlya analysa kachestvennykh dannykh: Vega I TextAnalyst [Native Computer Program for Quality Data Analysis: Vega and TextAnalyst]. Aktual nye voprosy obchestvennykh nauk: sotsyologiya, politologiya, philosophiya, istoriya [Up-to-the-minute Questions of Social Sciences: Sociology, Politology, Philosophy, History], no. 46, 34–40 (2015). (in Russian)
18. Ore, O.: Theory of Graphs. American Mathematical Society, Providence (1983)
19. Schutze, H.: Dimensions of meaning. In: Proceedings of Supercomputing 1992, pp. 787–796 (1992)

20. Clark, S.: Vector space models of lexical meaning. In: Handbook of Contemporary Semantics, 2nd edn. Blackwell, Oxford (2014)
21. Kiela, D., Clark, S.: A systematic study of semantic vector space model parameters. In: Conference: Proceedings of the 2nd Workshop on Continuous Vector Space Models and Their Compositionality (CVSC) (2014). https://doi.org/10.3115/v1/W14-1503
22. Orekhov, A.V.: Markov moment for the agglomerative method of clustering in Euclidean space. Vestn. St. Petersburg Univ. Appl. Math. Comput. Sci. Control Process. **15**(1), 76–92 (2019). https://doi.org/10.21638/11702/spbu10.2019.106
23. Everitt, B.S.: Cluster Analysis, 5th edn. Wiley, Chichester (2011)
24. Aldenderfer, M.S., Blashfield, R.K.: Cluster Analysis. Sage Publications, Inc., Beverly Hills (1984)
25. Orekhov, A.V.: Agglomerative method for texts clustering. In: Bodrunova, S.S., et al. (eds.) INSCI 2018. LNCS, vol. 11551, pp. 19–32. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17705-8_2
26. Wald, A.: Sequential Analysis. Wiley, New York (1947)
27. Sirjaev, A.N.: Statistical Sequential Analysis: Optimal Stopping Rules. American Mathematical Society, Providence (1973)
28. Chow, Y.S., Robbins, H.E., Siegmund, D.: Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin, Boston (1971)
29. Shiryaev, A.N.: Optimal Stopping Rules. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-74011-7
30. Orekhov, A.V.: Criterion for estimation of stress-deformed state of SD-materials. In: AIP Conference Proceedings, vol. 1959, p. 070028 (2018). https://doi.org/10.1063/1.5034703
31. Orekhov, A.V.: Approximation-evaluation tests for a stress-strain state of deformable solids. Vestn. St. Petersburg Univ. Appl. Math. Comput. Sci. Control Process. **14**(3), 230–242 (2018). https://doi.org/10.21638/11702/spbu10.2018.304
32. Bodrunova, S.S., Blekanov, I., Maksimov, A.: Topics in the Russian Twitter and relations between their interpretability and sentiment. Submitted to the SAMSN International Workshop, Barcelona, Spain, pp. 1–6 (2019, unpublished)
33. Radev, D.R., Blair-Goldensohn, S., Zhang, Z.: Experiments in single and multi-document summarization using MEAD. In: Proceedings of the First Document Understanding Conference, pp. 1–8 (2001)
34. Nur'Aini, K., Najahaty, I., Hidayati, L., Murfi, H., Nurrohmah, S.: Combination of singular value decomposition and K-means clustering methods for topic detection on Twitter. In: Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 123–128. IEEE (2015)
35. Aggarwal, C.C., Zhai, C.: Mining Text Data. Springer, Boston (2012). https://doi.org/10.1007/978-1-4614-3223-4

Intrusion Detection and Avoidance for a Heterogeneous Cluster of Web Sites

Darren Ramsook(✉), Patrick Hosein, and Akash Pooransingh

The University of the West Indies, St. Augustine, Trinidad
darren.ramsook@my.uwi.edu,
{patrick.hosein,akash.pooransingh}@sta.uwi.edu

Abstract. This paper outlines the design of a Hybrid Intrusion Detection System for a Web-Server hosting a heterogeneous cluster of web sites. This system was trained using the Classification and Regression Tree (CART) technique, with the Gini index as the measure of impurity, and allows for a headless operation once deployed. The model utilizes information that was mined from the Access logs of a web server. The system automatically performs pre-processing, classification and the blacklisting of those IP addresses deemed to be harmful. This model relies on the correlation between the server issued status codes, HTTP Methods, types of files being accessed, the geographical location of the client and the prospect of that being malicious. This system, which was made open source for both public use and development, achieved an accuracy score of 94.5% on the test set. This paper is aimed to Internet as a complex network Conference.

Keywords: Intrusion detection · IP Blacklist · Decision tree · Geolocation

1 Introduction

As the internet grows in both size and user-base, so does web based applications and services. These applications and services are also gradually becoming more complex and are a core part of multiple industries. This mixture of a large user base and accessibility means that the prospect of a malicious attack directed to the architecture of these applications and services is high [10].

Cybersecurity is a subset of the system architecture designed to safeguard networks, data, unauthorized access and any malicious intent. The intrusion detection system (IDS) is an essential element in a cybersecurity system and can be split into three main types: Misuse-based, anomaly based and hybrid. Misuse-based refers to systems that detect attacks based on signatures of those attacks. Anomaly-based refers to the identification of anomalies as deviation from regular usage. Anomaly-based can be seen as advantageous due to its ability to detect zero-day attacks. Hybrid systems refer to a mixture of both Misuse-based and Anomaly-based [3]. From a cybersecurity focused point of view, the fall-out of

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 250–256, 2019.
https://doi.org/10.1007/978-3-030-34770-3_19

a successful cyber attack can be detrimental with the average cost of an attack being $1.1 million USD [9].

The web server typically stores the files used to serve requests for web pages by a user. The user's browser requests the page using the HyperText Transfer Protocol (HTTP) and the server issues the corresponding data to the HTTP request. By default, most of these web servers collect log files which detail a list of actions that occur on the web server. The two main types of logs kept by these servers are Error and Access Log files, with the Error Logs storing any critical errors encountered by the server and the Access Logs storing all users that access the server's files and their method of access. It is a common practice to perform data mining techniques on these log files to classify user patterns and user classification [5,8,11,13] as well as performing anomaly detection [12].

In this paper, we describe the design of a machine learning based hybrid IDS which utilizes a decision tree classification system based on features that were data mined from the access logs of a given web server. This model was designed on the assumption that there is a correlation between the type of file being requested by the client, the country associated with the IP and the prospect of a malicious attack as it can be observed that some types of attacks utilize certain file types (e.g. Injection and Buffer Overflow attacks use `.php` files). The access logs were pre-processed and the following features for each row of data was extracted: HTTP Method (e.g. GET, POST, etc.), the type of file being requested from the server, the status code sent from the server to client, the last file visited on the server before accessing the current file and the country of the Internet Protocol (IP) address of the client. The IPs associated with malicious activity were then added to a blacklist and therefore banned from accessing the server.

The access log data-set originated from an Apache2 server and consisted of 53485 lines. It was split into both training and testing sets utilizing 75% of the data in the training and the remaining in the testing. This data-set was obtained on a server that hosts over 100 school sites. Each school site is managed by system administrators of the school. They use a wide range of content management systems and the school system administrators are typically not trained on protecting their sites. Therefore our objective is to use a CMS agnostic approach to protect these school sites which are constantly under attack.

2 The Decision Tree Model

2.1 Log File Processing

Data was obtained from the Trinidad & Tobago Network Centre (TTNIC) [1]. TTNIC offers a pro-bono service to educational institutes in Trinidad & Tobago, whereby they are granted a domain name and web server space. The educational institutes are made aware that they are responsible for the security of their web servers. Traditionally, these institutes do not have the resources and expertise to establish a robust level of protection, which in turn led to the creation of this system for use by TTNIC to add protection over these servers.

The data supplied was stored in a string format with an array of classification types for each line. The classification types were of three categories: Safe, Unknown and Malicious with each being represented numerically as 1, 2, 3 respectively. Pre-processing was needed in order to convert this data into a relational model and also to extract essential features.

The access logs reside in the server and reports on all requests that are processed by the server. The access log data-set was supplied in the custom log format `"%h %l %u %t '%r' %s %b '%Referer' '%User-agent'"` with the features being described in Table 1. An example of a line from the access logs with the IP address redacted is:

```
xxx.xxx.xxx.xxx - - [31/Jan/2019:04:28:04 +0000] "GET /xmlrpc.php
HTTP/1.1" 200 42 "-" "Mozilla/5.0 (compatible; MSIE 9.0; Windows
NT 6.1; Trident/5.0)"
```

To properly convert and extract the desired features, a regular expression for the log string was developed and applied on the entire dataset. The country location for each IP address was found through the use of the IPInfoDB API [6]. This was then coded into numerical categorical values for training. For example, a sample entry of one row of the final uncoded data-set used for training is "Get, .html, 200, edu.tt, US, 1" corresponding to the "HTTP Method, File Type, Status, Referer, Country and Classification" fields.

Table 1. Custom log format

Log features	Description
`%h`	IP address of client that made a request to the server
`%l`	Remote logname
`%u`	User ID of the person requesting the file
`%t`	Timestamp of the log line
`%r`	The request from the client to server and the HTTP method used
`%s`	Status code issued by the server
`%b`	The size of the object returned to the client from the server
`%Referer`	The location that the client was referred from
`%User-agent`	Information about the client browser

2.2 The Training Process

The non-parametric classification and regression tree (CART) technique [2] was used in the training of this classification model. This decision tree was formed as follows. Assume the training vectors $x_i \in R^n$, where $i = 1 \dots l$ and the label vector $y \in R^l$, where l is the number of sub collections in the set. Let the data given at node m be defined as D. Partition the data into $D_{left}(\theta)$ and $D_{right}(\theta)$

by evaluating possible candidate split $\theta = (k, t_m)$, which utilizes feature k and threshold value t_m:

$$D_{left}(\theta) = (x, y)|x_j <= t_m, \qquad D_{right}(\theta) = D \setminus D_{left}(\theta) \tag{1}$$

The impurity at node m was given through the use of the Gini impurity function, $Gini(\theta)$ defined as

$$G(D, \theta) = \frac{|D_{left}|}{|D_m|} Gini(D_{left}(\theta)) + \frac{|D_{right}|}{|D_m|} Gini(D_{right}(\theta)) \tag{2}$$

The parameters are selected such that the impurity is minimized:

$$\theta^* = argmin_\theta G(D, \theta) \tag{3}$$

When the output is a classification for node m, which represents a region R_m and N_m observations, the proportion of class j observations in node m is given by

$$p_{mj} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = j), \tag{4}$$

with the measure of Gini impurity being

$$Gini(X_m) = \sum_j p_{mj}(1 - p_{mj}). \tag{5}$$

This process is repeated until the maximum depth is satisfied.

3 Classification and Access Control

The decision rules obtained from the previous section were then used to detect fraudulent IP addresses and then blacklist them. The model was scheduled with a time-based scheduler for Unix operating systems (a cronjob) on the web server. This scheduled the model to run at user set intervals performing data pre-processing on the logs, followed by classification and addition of the flagged IPs to a blacklist. The interaction of the scheduling and blacklisting processes are shown in Fig. 1.

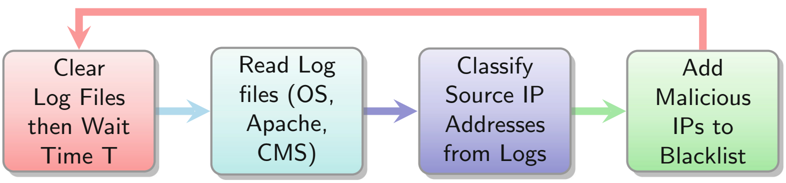


Fig. 1. Overview of the detection and blacklisting processes

4 Numerical Results

Several combinations of the features outlined previously were used in the generation of various decision tree models. The accuracy of each combination is compared in Table 2. These results were determined through the use of the Jaccard Similarity Score [7]. The combination of the features "Country, Status Issued, HTTP Method & File Type Requested" resulted in the highest classification accuracy of 94.5% and is referred to as Model Opt for the remainder of this paper. Note that the model trained on the feature "Country" alone resulted in a relatively high classification accuracy of 82.6%. The decision rule based on country was simply that access from local IPs were deemed safe while access from foreign IPs were deemed unsafe. In other words, we found that a large majority of attacks were generated outside the home country of the sites.

Table 2. Accuracy comparison of decision tree models with different features

Features used in decision tree model	Accuracy
Country	82.6%
Country & Status Issued	72.8%
Country, Status Issued & HTTP Method	94.3%
Country, Status Issued, HTTP Method & File Type Requested	94.5%
Country, Status Issued, HTTP Method, File Type Requested & Referer	94.3%

There are various performance metrics that are typically considered fro comparisons including `Precision`, `Recall` and `F1 Score`. These metrics are defined by

$$\texttt{Precision} = \frac{TruePositives}{TruePositives + FalsePositives} \tag{6}$$

$$\texttt{Recall} = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{7}$$

$$\texttt{F1 Score} = 2 * \frac{\texttt{Recall} * \texttt{Precision}}{\texttt{Recall} + \texttt{Precision}} \tag{8}$$

and are provided in Table 3. The time taken for the model to pre-process, classify and populate the IP Blacklist was 2005.8 s for a data-set of 53485 rows with 3748 unique IP addresses. The pre-processing without the call to the IPinfoDB API was estimated to be 6.78 s, and 1880.76 s with the call to the API.

5 Monitoring Effectiveness of Approach

The system was scheduled to run every 12 h over a 120 h time period. Figure 2 shows the number of new IP addresses added to the blacklist every 12 h. The three countries with the greatest contributions to the list are color coded

Table 3. Performance metrics for model opt

Class	Precision	Recall	F1 score
Safe	0.94	0.93	0.93
Unknown	0.42	0.22	0.29
Malicious	0.96	0.99	0.97
Weighted average	0.93	0.94	0.94

and identified. After the 120th h, the blacklist grew at an average of 135 new IPs every 12 h. After the first 12 h the largest number of IP addresses were added. This rate decreases over time as the system may have prevented attacks originating from already blacklisted IP addresses.

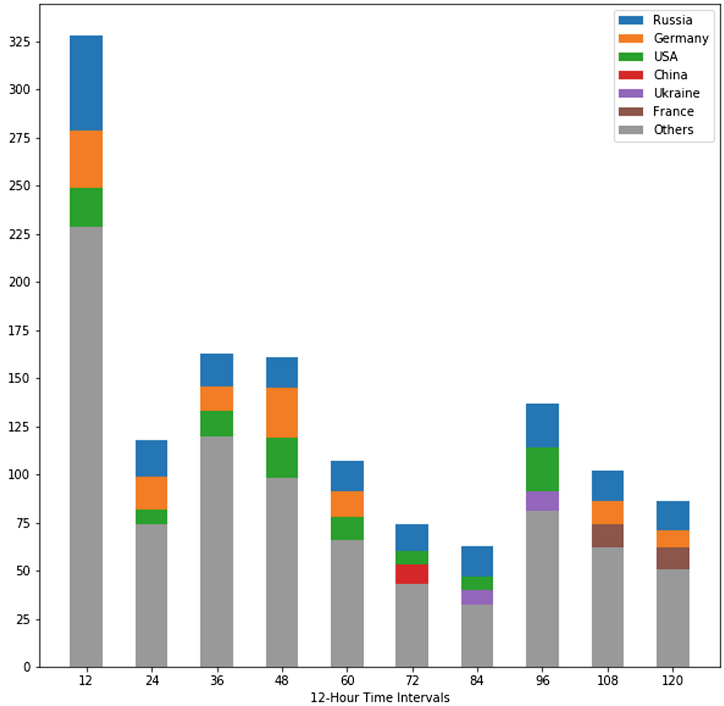


Fig. 2. Number of new IP Addresses blacklisted over 120 h at 12 h intervals

6 Conclusions and Future Work

We investigated a simple IDS for protection of web sites. Through the measure of impurity by using the Gini Index, decision tree models were generated utilizing different feature combinations.

The model trained with the combination of features “Country, Status Issued, HTTP Method & File Type Requested” resulted in the highest accuracy of approximately 94.5%. This classification model had a high accuracy of detecting malicious actions in terms of the F1-Score (0.97). Classifying safe actions also yielded a relatively high F1-Score of 0.93 while unknown actions had an F1-Score of 0.29.

The time taken to pre-process records for new IPs are relatively high due to the bottle-necking at the IPinfoDB API call. Future improvements in run-time can include use of a local IP database for geolocation services. Various ways of verifying the origin of the IP addresses should also be considered as the possibility exists that the user’s true IP addresses are being masked.

This platform has been deployed for the protection of a web server hosting over 100 school sites and has already proved to be very effective in reducing attacks as well as reduction of resources (bandwidth, CPU, memory) utilized by these attackers. The program files and models were made open source after development and can be found on github [4].

References

1. About us. https://www.nic.tt/
2. Breiman, L.: Classification and Regression Trees. Routledge, New York (2017)
3. Buczak, A.L., Guven, E.: A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun. Surv. Tutor. **18**(2), 1153–1176 (2016). https://doi.org/10.1109/COMST.2015.2494502
4. DarrenR96: Darrenr96/ai-apache-server-intrusion-detection-and-avoidance, May 2019. https://github.com/DarrenR96/AI-Apache-Server-Intrusion-Detection-and-Avoidance
5. Grace, L., Maheswari, V., Nagamalai, D.: Analysis of web logs and web user in web mining. arXiv preprint: arXiv:1101.5668 (2011)
6. IPInfoDB: IP info, IP geolocation tools and API. https://ipinfodb.com/
7. Jaccard, P.: Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. **44**, 223–270 (1908)
8. Pamnani, R., Chawan, P.: Web usage mining: a research area in web mining. In: Proceedings of ISCET, pp. 73–77 (2010)
9. Radware: Radware’s 2018-2019 global application & network security report (2018). https://www.radware.com/ert-report-2018. Accessed 20 Apr 2019
10. Singh, N., Jain, A., Raw, R.S., Raman, R.: Detection of web-based attacks by analyzing web server log files. In: Mohapatra, D.P., Patnaik, S. (eds.) Intelligent Computing, Networking, and Informatics. AISC, vol. 243, pp. 101–109. Springer, New Delhi (2014). https://doi.org/10.1007/978-81-322-1665-0_10
11. Suneetha, K., Krishnamoorthi, R.: Identifying user behavior by analyzing web server access log file. IJCSNS Int. J. Comput. Sci. Netw. Secur. **9**(4), 327–332 (2009)
12. Tharshini, M., Ragavinodini, M., Senthilkumar, R.: Access log anomaly detection. In: 2017 Ninth International Conference on Advanced Computing (ICoAC), pp. 375–381. IEEE (2017)
13. Thuraisingham, B., Khan, L., Masud, M.M., Hamlen, K.W.: Data mining for security applications. In: 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, vol. 2, pp. 585–589. IEEE (2008)

Improving User Experience by Browser Extensions: A New Role of Public Service Media?

Urbano Reviglio(✉)

LAST-JD International Joint Doctorate in Law, Science and Technology, University of Bologna, Bologna, Italy
urbanoreviglio@hotmail.com

Abstract. The paper questions the role of public service media in the digital era. The Internet has in fact disrupted previous patterns of production, distribution and consumption of information. Concerns arose on social media effects on well-being and how mainstream platforms design affects information consumption. The paper is an interdisciplinary contribution structured as follows. Firstly, it critically analyses the risks resulting from social media's usage, with a special focus on personalization practices. Then, it explores the development of Public Service Broadcasting and questions the role that Public Service Media (PSM) has to sustain media quality, information diversity and, more generally, its traditional values. Thus, arguments in favor of a renovated and proactive role of Public Service Media are provided. In particular, an agonistic approach to social media, an 'architecture for serendipity' and the role of attention management are advocated. Finally, drawing from information architecture and nudging theory, the paper introduces the concept of 'meta-design' as the ability to re-shape a digital environment by browser extensions in order to change design choices as well as to inform and educate users. The conclusion is that improving user experience by meta-design can actually represent a novel experimental role for PSM and, eventually, a soft regulatory tool for sustaining individuals and the general public interest.

Keywords: Public service media · Social media · Design · Nudging · Browsers

1 Introduction

Online communication is at a critical juncture that entails global phenomena such as polarized news coverage, hate speech, misinformation and, more generally, market concentration and the crisis of journalism [1–3]. In particular, there are several negative effects of social media on well-being [4]. Moreover, personalization of media content can limit the access and the exposure to diverse information [4, 5] and, as a result, weakening democracy [6]. Also, there are increasing concerns about the potential usage of persuasive and 'hyper-nudging' techniques that threaten the autonomy of users [7, 8]. As a reaction to this emerging media environment, policy-makers, designers, engineers and ethicists are discussing potential solutions. Yet, for many reasons the

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 257–271, 2019.
https://doi.org/10.1007/978-3-030-34770-3_20

risks and consequences of the Internet and its usage are very hard to prove [9] and, eventually, to counteract, especially regarding algorithms, personalization and so-called filter bubbles [10].

Much has been discussed about the role of public service media in the digital era [11–15]. Apart from relatively few examples, however, Public Service Broadcasting (PSB) and its digital version Public Service Media (PSM) have never been at the center of media reform activities. Despite the urgency to find new strategies, PSM would need to take shape in a concrete manner. So far, such inability to evolve has actually represented an impending ‘crisis of imagination’ [16]. Thus, the article aims to address the following key issues; what might be the role of PSM in the emerging media landscape? Can it represent an institutional mediator able to proactively help solve media crisis and, more generally, reinforce democracy and human rights? How can PSM concretely counteract the detrimental principles that shape social media today such as superficiality and hedonism, while sustaining media pluralism and PSB traditional values, such as diversity, universality, publicness and quality?

The paper is a theoretical and interdisciplinary contribution structured as follows. In Sect. 2, the paper analyses the risks of social media with a special focus on media pluralism and personalization. Contrary to recent research minimizing phenomenon such as filter bubbles, arguments in favor of a precautionary approach are indeed provided. In Sect. 3, after a close examination of its evolution, the paper questions the role that PSM has in the current social media landscape. Approaches to favor a renovated and proactive role of PSM are then discussed in Sect. 4. Subsequently, in Sect. 5, drawing from insights on information architecture and nudge theory, paradigmatic browser extensions and aggregators are briefly presented to eventually frame the concept of *meta-design.* To conclude, limitations are critically explored and final considerations are made in Sect. 6.

2 The Internet, Social Media and ‘Diversity Exposure’

The Internet constitutes a significant challenge to established media policies and the role of PSB. With large media providers no longer serving a gatekeeping function, the consumption of information turned on the choices of individual users and algorithms. Existing models of information and news production, distribution and consumption have thus radically changed. Social media, in particular, is nowadays at a critical juncture that entails global phenomena such as polarized news coverage, hate speech, misinformation and, more generally, users’ data exploitation, market concentration and the crisis of journalism [1–3]. More broadly, there are increasing concerns that social media can have several detrimental effects on well-being [7].

Above all, in order to provide relevant information to users, the epochal transition from information scarcity to information abundance brought the need for a balance between a pull and a push approach [17], between explicit and implicit personalization [18]. In other words, between autonomy of choice and algorithmic/platform delegation. Online personalization indeed performs a fundamental role of knowledge management to restrain information overload, reduce complexity and satisfy individuals as well as preserve their time and attention. It is, however, an imperfect tool subject to political and commercial manipulations. The landscape of today's social media is indeed characterized by an oligopolistic market in which users' attention is treated as a commodity in what has been described as 'surveillance capitalism' [1]. This small group of platforms act as the ultimate gatekeepers for billions of users worldwide. They provide personalized experiences (mostly implicit) but very little control over the information filtering processes [17, 18]. As such, short-term pleasure, soft news, homophily, entertainment and advertisement occupy a privileged place.

Academic literature describes how online content personalization can limit diversity, and create filter bubbles at individual level [4] and echo chambers at group level [3]. The major concern is that users may end up in a self-reinforcing cycle of opinions, hardly pushed to discover alternative and challenging standpoints in the so-called 'marketplace of ideas'. As a result, individuals could reduce opportunities to self-determination as well as their ability to build productive social capital [4]. At the collective level, media pluralism could be weakened as well. Thus, the audience become increasingly fragmented and people – especially the less skilled and literate – become more politically polarized and vulnerable to censorship and propaganda [3]. Media pluralism and autonomy of choice can indeed be threatened, and so can democracy.

Currently, however, research is contradicting, ambiguous and, to some extent, unreliable [9–11, 20]. The risks of social media usage [9] and personalization [10, 21], in fact, are very hard to prove and, eventually, to counteract. Most research is often inconsistent and inconclusive, mainly because it is based on small or unsatisfactory samples. In the light of the rapidly changing media landscape many studies quickly become out-dated. Platforms are indeed constantly changing their algorithms and business models. There is a crisis on the study of algorithms [10]. Significant gaps in research remain as well as in a consensus on a common set of definitions [20]. While insights on the main causes and risks are currently known [9] little is known about the extent of these consequences and the potential socio-technical solutions. Of course, many proposals to counteract such risks have been discussed. While some have focused on improving diversity by design [13], others focused on users behavior and media literacy [19]. In any case, there has not been sufficient evidence to justify any strong legal intervention.

The concept of media pluralism and diversity are intertwined and well-established democratic principles, especially in Western Europe's media research and policy since the 1960s. Indeed, the Council of Europe [21] and European Commission [22] explicitly recognized diversity as a policy goal that "embraces a number of aspects, such as diversity of ownership, variety in the sources of information and in the range of contents available" (p. 5). Yet media policy has mainly aimed at organizing the supply-side of pluralism through various sources that focus on content diversity [13]. The practical implications of this policy have long been associated with PSB. Given the mutated media landscape, the paper specifically questions the role of PSM. There are still discussions on how PSB and its digital counterpart PSM can adapt and, eventually, steer its development, in particular to preserve (or even enlarge) an audience that is fragmenting and to ultimately sustain end-users to make sense of information in the digital era. Fundamental questions need to be answered so far: how can PSB and PSM still survive and maintain a relevant role in shaping the social media landscape?

In particular, in what ways might PSM sustain the consumption of diverse information and a more fruitful public discussion based on the principles that guided PSBs in the past? In other words, what are the media instruments with which PSM might intervene more effectively?

3 The Role of Public Service Media in the Digital Age

The notion of public service in relationship to media was developed in the early 20th century under a specific set of political, technological and social conditions. Perspectives on this relationship changed across time and space. Yet there remained 'an overlapping consensus on certain core normative criteria' that can be categorized into three main principles, where the latter two are derived from the first [23, 24]: (1) Enhancing, developing and serving social, political and cultural citizenship; (2) Universality; and (3) Quality of services and output. Van Cuilenburg and McQuail [25] identified three phases in Western communication policy-making paradigms that show the evolution of these principles: (1) emerging communications industry policy in the period before the Second World War (WWII), (2) public service media policy after the War until the 1980s, and (3) new communications policy since the end of PSB monopolies.

The first phase was characterized by public monopolies with the limited goals of protecting the interests of government and nation, promote innovation and provide an efficient public service, whereas the second phase was characterized more by normative and sociopolitical rather than technological and economic considerations. Media policy started to promote diversity of ownership and content, limit monopoly and deal more effectively with the press. Such policies went much further in Europe than in the US, in particular regarding PSB, despite significant differences from one country to another. In the last phase, the rise of commercial television, which, from the 1980s, led to a majority of private stations by the mid-1990s, gradually reversed the monopoly PBSs used to have. Even if it retains some legitimacy and popularity, the normative policy paradigm of the postwar era declined in authority and scope and changed in the means for achieving its goals. This was not much due to technological changes but mostly to economic and political forces legitimized by the ambitions of media companies and governments. As soon as PSB became the exception rather than the rule, its activities and funding, if not its very existence, began to be considered as a disturbance of market relations [11]. PBSs indeed changed their programs to better adapt to audiences and advertisers while taking over certain orientations of private competitors, such as competition and cost awareness. As a consequence, PSB core values began to be eroded [25].

Despite differences that relate to the national context of PSBs, it is still widely believed that basic functions, such as a low-cost and universally available reliable provision of information, education and culture, and the catering for minority tastes and interests, cannot or will not be sufficiently served by the commercial market [11]. Yet, it is believed that PSB should not be restricted to remedying 'market failures' but to respond to the needs of citizens – which are indeed different from those of consumers. Thus, not only should it provide a counterweight to the commercial media but it is expected to set quality standards for the whole media landscape. Anyhow, fierce competition and new consumption patterns in the television landscape made it increasingly difficult to attract and sustain audiences.

At the beginning of the digital revolution, the discussion about the future of PSB revolved around two competing visions: those who were critics of the further commercialization of PSB and those who simply wanted to save PSB by adapting it to the system (most of the PBSs themselves and policy-makers). In particular, the European Broadcasting Union (EBU)[1] followed the second path. The EBU Digital Strategy Group [26] considers PSB as an 'island of trust' amidst multimedia companies offering linear broadcasts and online programs. Public broadcasters have learned to adapt to predominantly commercial media markets, by choosing a middle way between popularization and purification in their program strategy.

As a consequence of the new emerging landscape, the transition from PSB into PSM was debated.[2] In general, this implied the extension of public services beyond radio and television to encompass the full specter of the Internet. As a consequence, public broadcasters were to redefine their relation to the public. One of the main challenge remained the gradually diminishing reach of PSB among minorities and 'problematic groups' such as younger generations, migrants and the less educated [14]. To overcome this, many PBSs have started experimenting with new ways of legitimization, accountability and transparency towards citizens and society. PSBs initially started using social media to reach and engage new audiences. Such turn to the digital environment also stimulated PSBs in Europe to experiment with a public version of "social TV". Online participation has been considered a key-strategy to regain position in national arenas. Such commitments, though, do not appear particularly effective in shaping information consumption, and therefore are often considered insufficient.

4 Towards a New Role of PSM

Most scholars agree that PSB as well as PSM require new forms of justification, not simply to save its role as a public institution but to educate a generation grown up in a global multi-platform world. A new media policy paradigm is indeed needed, and the three main paradigms discussed earlier highlight the deterministic influence not much of technology but of social values.

Some scholars advocated alternative and radical proposals such as a 'Public Service Internet' [28] or 'Public Service Algorithms' [15, 29]. Burri [15] also advocates the

[1] EBU is an alliance of public service media (PSM) organizations, established on 12 February 1950, made up of 72 members in 56 countries, yet unrelated to European Institutions. Its main objective is to assist its Members in this period of unprecedented technological changes.

[2] Splichal [27] gives a concise definition of PSM: *"In normative terms, public service media must be a service of the public, by the public, and for the public. It is a service of the public because it is financed by it and should be owned by it. It ought to be a service by the public – not only financed and controlled, but also produced by it. It must be a service for the public – but also for the government and other powers acting in the public sphere. In sum, public service media ought to become 'a cornerstone of democracy"* (p. 255).

idea of a “Public Service Navigator” as a mechanism for influencing the conditions of access to content, particularly its visibility, discoverability, and usability. Yet, other scholars have been skeptical of the possibility to reform the public service system as such. The debate has thus mainly revolved on platforms’ regulation versus self-regulation [30]. Generally, it is thought that existing tools such as consumer data protection and antitrust regulation can be applied to the these new challenges and that, more generally, media should be left to self-regulation. In the European media policy it is in fact assumed that in most cases soft law promotes self-regulation, even if doubts concerning its effectiveness and the transparency of certain tools remain. A paradigmatic example is in fact illustrated by EBU which is currently seeking to support its members as they struggle with both the opportunities and challenges of Big Data through the *EBU Big Data initiative*, an interdisciplinary network launched in 2015 that aims to guide PSM in the implementation of data-driven strategies. Yet, more recently, media policy is moving towards a *cooperative responsibility*, that is, the result of the dynamic interaction between platforms, users and public institutions such as PSM [31].

4.1 Legitimizing a New Role of PSM

To legitimize any radical policy intervention as well as to guarantee an online media environment aligned with the most profound individual and societal needs, PSM should especially focus on the individual cognitive factors that challenge the experience of diversity online [32]. This means to employ a user-centric perspective and to especially focus on media literacy. To achieve this, we introduce three radical perspectives. Firstly, we argue the need to acknowledge the significance of power relations, that is, applying an agonistic approach to media pluralism. This could result in providing tools for contestation and for proactively seeking – but also be exposed to – challenging and minoritarian information. Relatedly, we advocate the need for an ‘architecture for serendipity’, that is, increase by design the access and exposure to alternative perspectives that intersects a user profile. Finally, we introduce an often underestimated element to improve users consumption: attention – the main commodity of the Internet [1] – which can be translated into ‘attention self-management’ to actually counteract the risks of addictive and superficial usage of social media and, more generally, the Internet itself, so as to foster a slow and more constructive consumption of media [33].

To begin with, media pluralism as a normative principle is vague and under-theorized, and it is not a reliable indicator of a society’s level of freedom and it can also create only the illusion of content diversity [34]. Moreover, in the digital age it is becoming less clear in which sense it is meaningful to speak of media pluralism, if the consumption is potentially characterized by limitless choice. As such, Karpinnen [ibid] advocates the idea of *‘agonistic pluralism’* to the context of media politics. The starting point is that media pluralism cannot be conceived only in terms of heterogeneity and a diversification of options but it needs to be analyzed in connection with the structural relations of power that define the criteria that guide information selection and limit available choices. Such perspective helps to defend concrete institutional arrangements in media policy. PSM can indeed be seen as a key tool in creating a plurality of power structures that is more open to democratic contestation, that resists the hegemonic tendencies of the market, and that reduces semantic inequalities. By analyzing digital tools that help to burst filter bubbles and weaken echo chambers, Bozdag and Van den Hoven [6] concluded that not all democratic models are represented in these tools, and that agonistic elements should also be included so that the needs/voices of minorities can be heard.

To achieve media diversity, there is also an increasing recognition that the design for serendipity in personalization has the potential to prevent the threats of filter bubbles and echo chambers [35]. In the last couple of years, emerging literature on how to research and cultivate serendipity in digital environments stimulated the academic debate. Serendipity is indeed an abstract but stable analytical category that can be used to detect and define public interests, so as to be inscribed as a design principle in digital environments and, at the same time, be used to build a sustainable strategic policy approach towards the media field. In general terms, its design implies diversification of information and interactive control. As such, it complements the value of diversity with paternalism and users' empowerment. From a theoretical perspective, instead, it recognizes the intrinsic limitations of personalization and profiling practices and, therefore, represents the attempts to overcome them.

Further, research on serendipity also investigates it as a capability of self-learning. The increasing role of accidental news consumption in the formation of individual and public opinion has been acknowledged, also called Incidental Exposure to Online News (IEON) [36, see also 37]. IEON is a promising behavior to encourage in the online environment as it may have a positive role in informing citizens and facilitating people's political participation. The majority of these studies are indeed based on the concept of *incidental learning*. Of course, there are still many limitations in this new field of study. Because information behavior varies so much across people, situations, and objects of interest research is not easily generalizable. Yet it informs how to experiment and, eventually, implement more 'serendipitous architectures' for social media.

Finally, to regain a relevant role in shaping users' information consumption and sustain knowledge acquisition, a special focus must be given to attention and its management. Attention management is indeed considered one of the most important skills for the 21st century, even replacing time management [38]. The ability to control distractions and stay focused is indeed essential to produce higher quality results in all aspects of life, particularly in hyperconnected societies [39]. Yet, the Internet is nowadays mostly based on the so-called 'attention economy' and related surveillance practices [1]. Mainstream platforms, in fact, conceive the attention of each user as a commodity that is then used to identify patterns of consumer behaviours, preferences, tendencies, etc. The attention of the users is, therefore, exploited, as it was – to a much more limited extent – in the pre-digital era. As such, users hold in a subordinate position. It is therefore imperative for PSM to counteract systems that are designed to capture users' attention and to provide tools that give them the autonomy to opt-out from certain design choices and better manage their own attention – also called attentive user interfaces. Psychological factors and media literacy are indeed paramount to overcome the challenges to experience diversity online and, more generally, for public discourse [32, 40].

5 Improving User Experience by ‘Meta-design’

In this paper, it is argued that a renovated PSM could provide affordances and tools for the Internet – and especially in social media – according to its traditional values, yet with updated goals. Drawing from insights on information architecture [41] and nudge theory [42], a review of current examples of browser extensions, aggregators and tool that empower users is done[3]. On the one hand, in fact, information architecture is a primary contributor to the shaping of the dialogue between the digital and the real and, thus, it is an ideal starting point to frame the concept of meta-design. On the other hand, nudge theory provide insights on its ethical framework. It is indeed possible to nudge users without introducing manipulative measures [43]. For example, Floridi [44] advocates for what he defines as ‘pro-ethical design’ which aims to modify the level of abstraction of the choice architecture by educating users to make their own critical choices and to assume explicit responsibilities.

As such, the concept of meta-design is preliminary framed as the re-design of a digital environment in order to change design choices as well as to inform and educate users; more broadly, as the act to reshape and enrich a website’s or platform’s information architecture and design throughout user-friendly tools such as aggregators and browser extensions so as to primarily influence and improve user experience as well as to sustain media literacy. In such endeavor, a user-centric perspective is implicitly applied to reflect on ways to extract value from users’ profiles meaningfully, to illustrate blind-spots, and to ultimately sustain the experience of diversity. Importantly, the paper also shows that most of these architectural modifications can be done without necessarily dealing with social media platforms. In fact, it is possible to help users to navigate the complexities of the new and fast-changing media landscape by exploiting the potential of ‘meta-design’. The following analysis provides insights on a potential multi-layered meta-design based on personalization and discovery tools on the one hand, and design choices and user interfaces on the other.

5.1 Personalization and Discovery Tools

A majority of PSM in Europe are currently moving in the direction of digital and algorithmic personalization [45], assuming the role of a public service navigator [13]. The encouragement of content discovery to disrupt filter bubbles has been already stressed by EBU [46]. Actual specifications for recommender systems (RSs) are, however, still to be developed. Due to fears of filter bubbles and echo chambers, the attention of EBU in the attempt to steer such development, for instance, was focused exclusively on exposure to diversity—at the cost of considering the need for common arenas of discourse, thus the values of universality and publicness, and how

[3] This selection was not an easy task. Unfortunately, there is no specific tag to collect those that might have been interesting to analyze. There are already hundreds of browser extensions inherent education in browser stores. As a consequence, the choice was to take most of them from an article review role on online media consumption. Others, instead, were collected from personal experience and online search. As such, the selection is limited yet sufficient to show the main features of the tools.

broadcasting and interactivity/choice can reinforce each other [14]. Moreover, many institutions with similar histories and comparable media system frameworks are taking up different positions. Some consider the possibility to reach the value of universality through personalization, while other consider personalization to work against it. Also, some privilege implicit personalization over explicit one, with very different outcomes on PSM goals.

In practice, it could be possible to design personalization to balance conflicting values throughout information architectures that favor exposure, exploration and awareness. In this respect, the idea to create an architecture for serendipity has been advocated [35]. This would eventually translates into four major goals; on the one hand, to design RSs that (1) provide a certain amount of non-personalized content, (2) provide content that intersects users profiles, so that a user might discover new ideas and interests outside one's filter bubble, and (3) to provide content that is politically and ethically challenging in order to temper the effects of echo chambers. On the other hand, design for serendipity is more generally intended (4) to provide empowering tools to increase information findability and discoverability. Also Sunstein [3] stressed how an architecture of serendipity as such would sustain 'chance encounters and shared experiences', which are considered preconditions for a well-functioning democracy. For example, he proposed a 'serendipity button' during political elections in order to be exposed to cross-ideological content. Importantly, serendipity may work against the accuracy of recommendations and, therefore, it can threaten PSM's commercial needs.

An interesting example of an exploratory and serendipitous tool is provided by the MIT project Gobo, a social media news aggregator with sliders that users can control to filter information.[4] This project sheds light not only on the possibility to provide sliders to filter media content based on certain democratic principles but, more interestingly, the possibility to aggregate and re-filter the information a user might want to consume in social media. Hence, it shows the possibility to bypass the algorithmic curation of information intermediaries. This, however, applies only to public posts. Yet, there may be implemented specific privacy settings to permit them to opt-in so that their posts may be included in the re-filtered feeds. Similar aggregators can be designed also for PSBs' websites and video sharing platforms like Youtube in order to increase information discoverability and the experience of diversity.

5.2 Design Choices, User Interface and Dashboards

A paradigmatic example of meta-design in social media is Social Fixer, an extension that improves user experience in Facebook and allows several design choices changes. Interestingly, it can set Facebook's NewsFeed to switch to "Recent Stories First" – the chronological feed, which is not to algorithmically curated. By default, in fact, Facebook only shows you the top stories rather than every post made by your friends. Yet anytime one logs out of the website, the algorithmic curation reset by default the

[4] Other similar news aggregators are *Newsbird, Allside, News 360, Digg and Flipboard* [see 18].

NewsFeed. Such simple design choice would permanently free users from the algorithmic influence of Facebook. Another example is the possibility to hide any metrics' posts in order not to get captured into so-called 'trendism', that is, be influenced by comparative quantifications (e.g. likes and shares). This could help users to self-control more effectively their attention. Yet, not only this project highlights the potential for reshaping certain design choices but also the potential to communicate with messages and, thus, inform and educate users during their navigation.

Another significant mean to gain more influence over information consumption is by being aware of the production and consumption of information (*what* and *why* information is filtered and *how* one consumes it) [see 17]. For example, it is possible to encourage blind-spots explorations. In this respect, a notable example is the visualization tool developed by Nagulendra and Vassileva [47] which displays to users their filter bubbles, showing them which categories and friends are in their bubble and which ones are not, allowing them to control the algorithm by manipulating the visualization to escape the bubble, by adding or removing friends of a certain category to the filters. To our knowledge, this is the only attempt to show filter bubbles to users. Yet the quality of the design can certainly be improved and its scope expanded. While data collection may prove to be limited, users may opt-in to provide their personal data.

Further, extensions like Data Selfie provide insights on what algorithms think about you, your inferred personality and your vulnerabilities. This might be helpful given that persuasive technologies have different outcomes on diverse personality types, and the awareness of one's own vulnerabilities could truly help users in their self-management. Even simple design choices can be extremely effective. For instance, Social Fixer affords the option to disable the auto-loading of posts (similarly to the auto-play function for videos which is usually set by default). Hence, one can decide to scroll n posts anytime one logins, without being persuaded in an endless stream of content. Equally important, notifications management could be useful too to avoid distraction. Eventually, so-called information nutrition labels [48] can also help to manage your attention as well as increase awareness. For example, they can show the time reading, style and reliability of any article by text analysis.

6 Discussion

The above browser extensions, visualization tools and aggregators provide insights on the potential role of PSM in improving users' experience according to its values. It is by no means a ready-made framework any PSM could operationalize. Yet it highlights how certain extensions and tools are bringing innovations that PSM might replicate and eventually ameliorate. Despite technical challenges, meta-design can actually help PSM to collect data about users to better understand them, then inform, educate and nudge them towards its values. If such potential is employed, it might help to set the standards on a more mature information consumption, afford novel conditions to participate and benefit from social media usage and, eventually, contribute to solve the main challenges to experience diversity online [see 32]. At the same time, it would sustain the four modalities of expressing voice[5] [see 18]. Ultimately, it aims to preserve not only media pluralism and users' experience of diversity but, above all, their awareness and autonomy of choice. In addition, it has the potential to help to overcome some of the major identified causes of the crisis of PBSs (such as outdated values, harsh competition and government control, see [11]).

One may thus wonder why such tools have not been institutionally employed so far. To the knowledge of the author, in fact, there are in fact no noteworthy experiences of public efforts towards the creation of similar news aggregators, nudging by design or visualization tools. One explanation may be the widespread assumption that users are fully autonomous players able to manage the complexities of online information consumption. And although user-centred solutions may better foster the transparency of RSs, for example, they have significant shortcomings: they shift the responsibility and accountability for the protection of rights and utility to the users, and this may result in inefficiency such as insufficient privacy protection [49]. This is also partially true for media consumption.

Further, there is mainly a lack of scientific evidence, both about the risks of Internet usage and the effectiveness of these tools. Also, political challenges may be paramount. Citizens may not want to use these tools, while platforms proved to be difficult stakeholders to engage with whenever their revenues are threatened. It is evident the potential opposition of the platforms in which these tools may be experimented, so that they may be easily sabotaged. Somewhat paradoxically, another significant element is likely to be that such public intervention would more likely account for an irrelevant share of investments given to PBS and PSM. Clearly, these tools are extremely cheap compared to the content produced by PSB. Thus, economic incentives in this respect appear to be very weak for an institution that is especially struggling with funding. However, considering the rising awareness in the public opinion of the risks of social media and the educative potential of browser extensions, there is room to expect their employment in the near future.

Despite obvious political challenges, there are certainly several other limitations concerning the outlined preliminary framework. First of all, as said, its effectiveness, in particular concerning personalization and the attempt to provide users with exposure to challenging information. Interactivity might indeed threaten the principles of publicness and universality and not fulfil the goal of maintaining a common sense and belonging. The political consequences of exposure to dissimilar views, then, have long been a subject of polarization research, but findings are largely mixed. For example, polarized individuals may polarize further while exposed to challenging information (so-called *backfire effect*). Another related-risk is that users do not exploit such tools – they do not always want to be in control (e.g. the *paradox of control*). As a consequence, inequality may be even strengthened. These issues certainly require more understanding of users' information behaviors as media landscape changes rapidly and steadily. It is not possible to make harsh generalizations: the heterogeneity and nuances

[5] Building on the work of sociologist Hirshman, Harambam et al. [18] intend 'voice' as "the possibility to exert control over the data-driven processes that shape news provision".

of users must be better comprehended. These challenges actually represent additional reasons to begin experimenting with meta-design in order to understand and eventually exploit its opportunities.

7 Conclusions

Public Service Media (PSM) is striving to adapt to the digital revolution. There are still many opportunities to explore and challenges to face. As such, in Sect. 2 the article provided an updated and extensive review of research on the risks of social media usage. By doing this, a precautionary approach has been advocated, in particular due to the heterogeneity of the audience, intrinsic research limitations, the oligopolistic social media landscape, and the rise of increasingly sophisticated and subtle techniques of behavioral modification. In Sect. 3, a critical analysis of the role and the evolution of PSM was done. A renovated role of PSM is advocated, especially regarding asymmetrical power relations between platforms and users, the opportunities to improve user experience and the rising of PSMs' recommender systems.

In Sect. 4, it is argued that traditional PSMs' values such as universality, publicness, interactivity and diversity can be mediated and reinforced by an architecture for serendipity and an agonistic approach to design, that is, empowering users to sustain their motivations, raise their awareness and increase their skills to experience diversity. To achieve this, a core focus of media pluralism should be media literacy, through interactive education and, particularly, attention management. Further, in order to understand how to actually increase the experience of diversity as well as to sustain media pluralism and, more generally, the traditional values of PSM, a brief review of paradigmatic browser extensions and aggregators was also done in Sect. 5. It is argued, in fact, that these tools represent an undervalued, if not neglected, mean to experiment with digital environments, especially to better understand emerging audiences. No experience of institutionally-mediated browser extensions has been detected so far. Thus, an analysis on significant private and academic experiences was conducted and, eventually, the concept of meta-design was initially framed. In short, this is intended as the act to by-pass the design of websites and extract value from users' information consumption – primarily by browser extensions – so as to raise awareness and agency of worldwide users without having to legally deal with platforms themselves.

The outlined preliminary framework certainly needs empirical research. Despite outlined challenges, there is room to design architectures and tools able to nudge users according to PSM values. The role of meta-design might radically inform, educate, empower and nudge users. Of course, operationalizing such an endeavor is not a trivial task. In this paper, only an initial attempt to advocate for such a proactive role of PSM has been made. Although faced with many technical, social and political limitations, the presented concept of meta-design can represent a novel mean of soft and/or co-regulation and audiences' analysis for PSM institutions, particularly in Europe where the regulation tradition of media pluralism and PSB could favor the legitimization of such innovations.

References

1. Zuboff, S.: Big other: surveillance capitalism and the prospects of an information civilization. J. Inf. Technol. **30**(1), 75–89 (2015)
2. Bárd, P., Bayer, J.: A Comparative Analysis of Media Freedom and Pluralism in the EU Member States (2016)
3. Sunstein, C.R.: #Republic: Divided Democracy in the Age of Social Media. Princeton University Press, Princeton (2017)
4. Pariser, E.: The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think. Penguin, New York (2011)
5. Bozdag, E., van den Hoven, J.: Breaking the filter bubble: democracy and design. Ethics Inf. Technol. **17**(4), 249–265 (2015)
6. Yeung, K.: 'Hypernudge': big data as a mode of regulation by design. Inf. Commun. Soc. **20** (1), 118–136 (2017)
7. Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., Kross, E.: Do social network sites enhance or undermine subjective well-being? A critical review. Soc. Issues Policy Rev. **11** (1), 274–302 (2017)
8. Zuiderveen Borgesius, F.J., Trilling, D., Moeller, J., Bodó, B., de Vreese, C.H., Helberger, N.: Should we worry about filter bubbles? Internet Policy Rev. **5**, 1–16 (2016). https://doi.org/10.14763/2016.1.401
9. Bodo, B., et al.: Tackling the algorithmic control crisis-the technical, legal, and ethical challenges of research into algorithmic agents. Yale JL & Tech. **19**, 133–182 (2017)
10. Bardoel, J., d'Haenens, L.: Reinventing public service broadcasting in Europe: prospects, promises and problems. Media Cult. Soc. **30**(3), 337–355 (2008)
11. Gulyás, Á., Hammer, F. (eds.): Public Service Media in the Digital Age: International Perspectives. Cambridge Scholars Publishing, Cambridge (2014)
12. Helberger, N.: Public service media, merely facilitating or actively stimulating diverse media choices? Public service media at the crossroad. Int. J. Commun. **9**, 17 (2015)
13. Horowitz, M.A., Nieminen, H.: Diversity and rights. Connecting media reform and public service media. IC-Rev. Científica Inf. Comun. **14**, 99–119 (2017)
14. Burri, M.: Public Service Broadcasting 3.0: Legal Design for the Digital Present. Routledge, London (2015)
15. van Es, K.F.: An Impending Crisis of Imagination: Data-Driven Personalization in Public Service Broadcasters. Media@ LSE Working Paper Series (43), pp. 1–18 (2017)
16. Pentina, I., Tarafdar, M.: From "information" to "knowing": exploring the role of social media in contemporary news consumption. Comput. Hum. Behav. **35**, 211–223 (2014)
17. Thurman, N., Schifferes, S.: The future of personalization at news websites: lessons from a longitudinal study. J. Stud. **13**(5–6), 775–790 (2012)
18. Harambam, J., Helberger, N., van Hoboken, J.: Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem. Phil. Trans. R. Soc. A **376**(2133), 20180088 (2018)
19. Tucker, J., et al.: Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature (2018)
20. Dubois, E., Blank, G.: The echo chamber is overstated: the moderating effect of political interest and diverse media. Inf. Commun. Soc. **21**(5), 729–745 (2018)
21. Recommendation CM/Rec(2018)1[1] of the Committee of Ministers to member States on media pluralism and transparency of media ownership (2018)
22. European Commission: In Media pluralism in the member states of the European Union, Brussels, Belgium, pp. 1–91 (2007)

23. Born, G., Prosser, T.: Culture and consumerism: citizenship, public service broadcasting and the BBC's fair trading obligations. Mod. Law Rev. **64**(5), 657–687 (2001)
24. Picard, G., Pickard, V.: Essential Principles for Contemporary Media and Communications Policymaking. Reuters Institute for the Study of Journalism. University of Oxford (2017)
25. Van Cuilenburg, J., McQuail, D.: Media policy paradigm shifts: towards a new communications policy paradigm. Eur. J. Commun. **18**(2), 181–207 (2003)
26. EBU Digital Strategy Group: Media with a Purpose: Public Service Broadcasting in the Digital Age. EBU, Geneva (2002)
27. Splichal, S.: Does history matter? Grasping the idea of public service at its roots. In: Lowe, G.F., Bardoel, J. (eds.) From Public Service Broadcasting to Public Service Media. RIPE@2007, pp. 237–256. Nordicom, Gothenburg (2007)
28. Fuchs, C.: Towards the public service internet as alternative to the commercial internet. ORF Texte **20**, 43–50 (2017)
29. Bennett, J.: Public service algorithms. In: Freedman, D., Goblet, V. (eds.) A Future for Public Service TV, pp. 112–120. University of Goldsmith Press, London (2018)
30. Napoli, P.M.: Social media and the public interest: governance of news platforms in the realm of individual and algorithmic gatekeepers. Telecommun. Policy **39**(9), 751–760 (2015)
31. Helberger, N., Pierson, J., Poell, T.: Governing online platforms: from contested to cooperative responsibility. Inf. Soc. **34**(1), 1–14 (2018)
32. Hoffmann, C.P., Lutz, C., Meckel, M., Ranzini, G.: Diversity by choice: applying a social cognitive perspective to the role of public service media in the digital age. Int. J. Commun. **9** (1), 1360–1381 (2015)
33. Rauch, J.: Slow Media: Why Slow is Satisfying, Sustainable, and Smart. Oxford University Press, Oxford (2018)
34. Karppinen, K.: Media and the paradoxes of pluralism. In: The Media and Social Theory, pp. 27–42 (2008)
35. Reviglio, U.: Serendipity as an emerging design principle of the infosphere: challenges and opportunities. Ethics Inf. Technol. **21**(2), 151–166 (2019)
36. Yadamsuren, B., Erdelez, S.: Incidental exposure to online news. Synth. Lect. Inf. Concepts Retr. Serv. **8**(5), i73 (2016)
37. Fletcher, R., Nielsen, R.K.: Are people incidentally exposed to news on social media? A comparative analysis. New Media Soc. **20**(7), 2450–2468 (2018)
38. Davenport, T.H., Beck, J.C.: The Attention Economy: Understanding the New Currency of Business. Harvard Business Press, Boston (2001)
39. Floridi, L. (ed.): The Onlife Manifesto: Being Human in a Hyperconnected Era. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-04093-6
40. Carson, A.B.: Public discourse in the age of personalization: psychological explanations and political implications of search engine bias and the filter bubble. J. Sci. Policy Gov. **7**(1), 1–13 (2015)
41. Resmini, A. (ed.): Reframing Information Architecture. HCIS. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06492-5
42. Thaler, R.H., Sunstein, C.R.: Nudge: Improving Decisions About Health, Wealth, and Happiness. Penguin, New York (2009)
43. Helberger, N., Karppinen, K., D'Acunto, L.: Exposure diversity as a design principle for recommender systems. Inf. Commun. Soc. **21**(2), 191–207 (2016)
44. Floridi, L.: Tolerant paternalism: pro-ethical design as a resolution of the dilemma of toleration. Sci. Eng. Ethics **22**(6), 1669–1688 (2016)

45. Van den Bulck, H., Moe, H.: Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. Media Cult. Soc. **40**(6), 875–892 (2017)
46. EBU: Big Data Initiative: Time to Invest. Le Grand-Saconnex (2017)
47. Nagulendra, S., Vassileva, J.: Providing awareness, explanation and control of personalized filtering in a social networking site. Inf. Syst. Front. **18**(1), 145–158 (2016)
48. Gollub, T., Potthast, M., Stein, B.: Shaping the information nutrition label. In: Proceedings of the NewsIR 2018 Workshop at ECIR (2018)
49. Milano, S., Taddeo, M., Floridi, L.: Recommender Systems and their Ethical Challenges (2019). Available at SSRN 3378581

Data Gathering, Management and Protection

The International Register of Ideas and Innovations. A Visionary Social Network to Develop Innovation and Protect IP Using Blockchain and Proof-of-Originality Algorithm

Giovanni Rinaldi[1(✉)] and Marco Manca[2(✉)]

[1] LINK Associati, Via Tiarini 22, 40129 Bologna, Italy
giovanni.rinaldi@ematchidea.com

[2] Board of Directors, SCImPULSE Foundation,
Johan Ruytersdreef 30, 6132TJ Sittard, The Netherlands
marco.manca@cern.ch

Abstract. The International Register of the Ideas and Innovations (IRI) is a "visionary" and at present an experimental project that wants to protect all over the world the intellectual work leveraging on the blockchain paradigm (BC). IRI protects all the intellectual work not only the traditional ones (patents, trademarks, copyright, industrial design, geographical indication) but also those not protected by traditional offices such as ideas, projects, models, services, artefacts. The starting point is how to generate trust between many and different subjects engaged in the innovation value chain in order to create, build and collaborate in the developing of innovation. For this reason BC is also used to certify the actions, activities and relationships between researchers, inventors, students, innovators, funders, industries, user communities during the creation, construction and marketing of innovation. Detaching from the main BC platforms that need the consensus of the transaction based on a mechanism of proofing (especially regarding the financial area like bitcoin or Ethereum) the core of the project is the Proof-of-Originality algorithm (PoOr) that detects and compares the originality of the proposal. The challenge is to verify how the publication on IRI of an idea, project, …, assumes a "de-facto" protection. If an innovator can protect the idea and if the actions and the relationships that allowed the realization of the innovation are certified, he can share it and this allows the co-construction, the development and at the same time this generates trust for the selling or the search for funding.

1 Introduction

Recent market changes have shown the urgency to find new paradigms, new solutions for growing new opportunities: we need much greater collaboration among idea creators, funders, R&D, and policy makers, to create more effective and faster pathways to innovation and to create value from human brain. The convergence of technologies, the use of the network and the grouping of minds and intelligences in virtual communities, have created a "free innovation", driven by the need to resolve problems. According von Hippel, "free innovation" has today achieved important successes. In a recent

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 275–282, 2019.
https://doi.org/10.1007/978-3-030-34770-3_21

research [1] that focuses on 6 countries (USA, UK, Korea, Japan, Canada and Norway) he shows that there are more than 25 million innovators, many of which are not recorded by official statistics. They produce innovations aimed at responding to the needs of the community to which they belong. Chesbrough shows [2, 3] that an innovative model that creates new business and new revenues is based on open innovation; value is generated by the ability to enhance external expertise that can be integrated into the value chain. Otherwise entrepreneurial innovation goes off over time. Making is connecting; you have to connect things (material and ideas) to create something; new ideas, information, resources, opportunities and trust move and transform through networks of relationships without necessarily adhere to what the company organization chart says [4]. The relationship between the production and the networking features the added value of contemporary economy, and there is also a large number of people which is not only the audience, but produces things standing in the network. As the result of the changing of the paradigm in which new ideas have social origins (within or outside organizations), innovation and new opportunities are based on the exchange of ideas, on the circulation and sharing of information and the necessity to identify skill gap between idea and the market. Innovation is not “an aspect” of the company, reduced to just research and development or new technology. It’s a dimension that affects every area of the enterprise and the nature of the business. The hint of “bet” on the intelligence and, in general, human capital, is making its way as an alternative to the social, economic and financial impoverishment. Taking up the challenge to put together different users’ communities in the whole value chain of innovation should enable the community of policy makers and educators to discover directly on the field new methodologies, drive new design processes, identify new business models. The paths leading to business in a collaborative perspective can also provide useful outcomes for social actions. Collaboration and social partnership, getting individuals actively included, provides new opportunities in business along all the value chain needed for the design and production of goods, products and services.

2 The International Register of Ideas and Innovations (IRI)

We think we can give an answer to the needs expressed in the previous chapter. We start from the concept that innovation can be driven by an ecosystem in which all the users interested in the innovation value chain are grouped in the same social network. At present we assist to a proliferation of social networks but each of them is based on the needs of a portion of interested people. Research, market, science, nurturing and composition of ideas are aspects of the innovation and innovation needs of all these components. This requires that different people are engaged in the same value chain. To compose a new idea people needs methods to compose and nurturing the cue in order to became the idea; this needs collaboration, science, research, to bring the idea to innovation; but how to acquire funds, according the potential market is the important step. When the idea is born in the brain and needs collaboration to nurture and grow must be bring to the market. How the engagement of different people can create the ecosystem? Creating trust. This is the main challenge from a social point of view. But people are together if they get something of value; and to get something of value it is necessary trust: trust in order to create collaboration, trust in order to find funds, trust in order to sell or acquire innovations, ideas, projects. The platform proposed is a provocation. It wants to follow the innovation from the birth to the adjustment up to the search for financing and the market. For these reasons must be based on trust. The challenge is how trust can be created by BC. At last, how to get trust in exchanging currency is the base point of the bitcoin “experiment” proposed by Satoshi Nakamoto (SN) [5]. Traditionally to exchange currency I trust a bank, but what if the third party (the bank) is excluded by the transaction? We need a mechanism that supplies the trust of the bank: this is the base concept of bitcoin. IRI is a platform in which user can publish and protect an idea, an innovation, a project, a research, an artifact, …, all the types of intellectual work. Moreover the platform certify any action necessary for the implementation of the proposal. The need to create, collaborate, research and build is based on trust. But how to build trust on the internet? We create a technical platform based on BC paradigm that answers these questions. Why should I accept collaborators for improving my project if I haven’t trust? Why should entrepreneurs finance innovations or ideas or why should I share data and information without trust? And how can I govern the value chain of innovation? Trust, governance, protection and collaboration are the face of the same medal. To create innovation, collaboration is needed. To collaborate, find funds or sell ideas, trust is needed. The publication on IRI of the products of the human brain certifies a protection “de-facto”. The certification of the actions done in order to build the proposal creates trust. We want to demonstrate how the new paradigm BC can be useful to create trust between the people engaged in the innovation value chain. This comes in two ways: IP protection of innovations and ideas and each product of human brain and the certification of the actions and activities using Smart Contract (SC); the interaction between users and the transactions are controlled by SC.

3 The Blockchain Paradigm

The BC paradigm has these fundamental features: it is a distributed (but it is more correct to say “replicated”), shared, decentralized and encrypted dataset; it ensures immutability and incorruptibility of all information; it is transparent and allows everyone (if permissioned) to access information; based on the consent for which can be modified only with the consent of all the participants to the network. BC paradigm was born as solution to cope with problems created by the exchange of cryptocurrency in order to ensure transaction correctness, avoid double-spending, remove the third party, maintaining privacy and trust. For reaching these results SN used and composed the technical features in order to build a shared, decentralized and encrypted dataset using specifics algorithms that guaranteed the goal. It is interesting to note that BC characteristics, theoretically, are required for different domains - in some cases highlighting some to detriment of others, making some less stringent or applying them only to part of the information managed – but certainly with the necessary architectural changes. But often the architectures developed afterwards have not proved completely suitable to resolve the problems for other domains of interest. These architectures has been used tout-court or with many changes producing the illusion to use them as the answer to different needs for different domains. Bitcoin introduced the consensus mechanism (transaction correctness and 3-party exclusion), but in a cooperative and permissioned environment it assumes a different meaning. If in the first case the need to incentive the nodes to confirm the transaction has originated the Proof-of-Work consensus [6], in the case of permissioned network (as Hyperledger) [7] probably the incentive could be not necessary (for example in the healthcare environment). Different ways for managing consensus are born, in some cases the concept is not the mechanism of reaching consensus but (as in Ethereum [8]) the concept of verification of the transaction. The technology changes up to consider only a restricted numbers of validators the delegated in the DPoS [9] used in closed networks or up to Proof of Authority (PoA) in which an authority is delegated to the validation re-introducing the 3-party agency excluded by SN. Moreover, the initial idea of the BC paradigm created by SN predicted a ledger shared, copied (once consent has been reached) on all the nodes, all the information necessary for the transaction is recorded on it. But whereas for the cryptocurrency exchange the necessary information is light, in other domains the object of the transaction could be very large and complex for example the descriptions of the object for which the intellectual propriety is claimed (as in IRI), or the Electronic Health Record of a patient in healthcare domain [10]. On the other hand the object of the transaction must be explicit because it cannot risk being modified. This problem requires to re-consider the ledger as a pure transactional database, but the reference to the object must be considered and this must be subjected to the same conditions of immutability and incorruptibility. Smart contract have been introduced also partially to resolve this problem, and this has increased the complexity of the architecture. IRI it is a challenge that wants to verify the usage of the BC paradigm in the innovation ecosystem that needs collaboration between different subjects and for this reason the transactions must be based on trust. The platform is mainly based on two workflows. Each actor inside the innovation ecosystem can build, develop and adjust an innovative project searching collaboration in compete trust because each action is certified by BC and SC. Eventually actors can publish an idea, innovation, model, project, research … on the IRI, using the BC. The publication establishes a "de facto" status protecting IP of the product of the human brain.

4 The blockchain (BC) for the IRI

We will start from the BC paradigm in order to understand as these information features (distribution/replication, sharing, decentralization and encryption) can be useful (at what form and in which grade). This central feature of BC results in obvious challenges to implementing solutions that share sensitive data, where only a restricted number of recipients should be given access to a piece of data, or a cryptographic artefact that can unlock a piece of data stored off the BC. Innovation value chain has the need to maintain reservation in searching collaboration in order to complete the idea or in order to sell it; and protection when the idea is published into the register and everyone can consult it without changing it. The BC implementation responds to 2 needs: maintaining security and reservation during the phases of completion and adjustment of the idea/research/artefact and the guarantee of transparency, reservation, not replicability during the publication into the IRI. These 2 features ensure trust in the whole innovation value chain from the birth of the idea up to the selling or funding request. During the idea construction each piece of data has one user (owner) who can share a piece of data they own with other users or groups at varying levels of access (summary versus full data). To limit full access, and instead enable summary access, each piece of data consists of a descriptor, viewable to anyone on the BC network, a summary, and content, which are stratified at different access levels. Therefore, having summary access gives the receiver only access to the descriptor and the summary, whereas full access provides all the components. Data sharing between users is modelled by a system where users can share data with other users and groups, as well as receive data requests from other users at any access level. If a user responds to a request by granting data access, a cryptographic artefact is exposed to the receiver in a way that allows only that receiver to view data at the specified access level. Given the complexity of the information that describes the innovative idea or the artefact the system ensures that sensitive information (including all documents) is never exposed on the BC. BC allows the link to the information describing the idea and ensures immutability when the idea is published. BC technology is not usually designed for large transaction data loads, so building a data sharing system, we prefer to store data in a separate software solution that can provide a global reference to uniquely identify the idea/research/artefact to publish. Because data is stored within an external storage solution, the BC component is responsible for executing Smart Contracts that, refer to idea information and provide info on how data is owned, retrieved, and decrypted. Moreover, since BC is not built for data queries based on a set of user criteria, the platform includes a component that enables users to build indices to enable this capability. Application services act as the public endpoint in the system and cover all higher order business logic which leads to the creation of -and interaction with- smart contracts and data retrieval within the BC infrastructure. The high-level abstraction layer allows the interaction with SC that turns as the workflow for the actions on the information managed by BC. SC are responsible for requesting and accepting collaboration; for publishing ideas and innovation into the register and for accessing the idea in order to be financed or bought. Innovation is composed by the description of idea based also on economic values; it depends by the grade of originality; in order to evaluate the idea it is necessary to describe the steps are been needed to build it; who are the people involved and what was their engagement. These information generate trust to sell the innovation or to seek funds, or to complete the idea searching experts and ask for collaboration. We build these information in a sort of chain of blocks. Publication on IRI generates a time stamp (indicating the date of publication) and the stuff published can no longer be modified. The mechanism ensures not modifiability, transparency, the uniqueness of the “work of intellect” against plagiarism and copying, certifying a state “de facto”. The architecture is composed by a set of software tools that are composed and adapted according the real needs in the innovation ecosystem. The components will be grouped in the following sets. (1) The access and participation to the network: permissioned in a public – private environment. (2) The traditional consensus method (the information has to be verified by at least half of the participants, and will only be approved if this half reaches a consensus), fundamental in financial area is not necessary for the domain we intend to cope. We introduce an algorithm the PoOr (Proof-of-Originality) that detects and compares the originality of the innovation published on IRI. (3) How the object value is embedded or linked to the mechanism of block creation. (the valuable information we intend to treat could be complex and not directly embedded to the chain.) (4) In this context the concept of "ledger" as defined by SN as list of transactions is out of scope. This implies a revision of the traditional ledger as the dataset that describes the features of the idea the innovator wants to publish. In our case we need to certify the immutability of the proposal described in the valuable information treated. On the other hand, the portfolio built with information blocks tend to be quite transparent. All participants in a distributed ledger system can see the transaction, which further helps to reduce fraud. The BC proposed, at core, is a shared database that has entries that must be confirmed and encrypted when being filled. The best way to think of them is as a very secure document. This document composition is the idea/innovation that is published on IRI. Any document entry should have some form of a logical link to the other predecessors when being keyed into the BC. The document entries are supposed to be added as blocks to the resulting transaction chain. Additionally, they also have to implement cryptographic signatures, commonly known as a hash, to make the information secure and eliminate the ability to edit it. (5) Smart Contract (SC) development as mechanism to ensure the activities done and the transactions. Transaction in this case is not just referred to cryptocurrency but to each modification of the state of the chain referred to a valuable information. Once a transaction is processed, the files are time stamped within the portfolio, and each is given a cryptographic signature. The material published -consisting of a chain of blocks- describes the publication. In it will be recorded: the grade of originality of the idea; the steps that were necessary for the implementation of the proposal, the CV of the proponents and any transactions. The grade of originality will be provided by an algorithm: "Proof-of-Originality". The description of the steps that were necessary to achieve the final state of the proposal highlights the complexity of the activity done. These actions will be added as immutable blocks and certified by third parties when requested (Fig. 1).

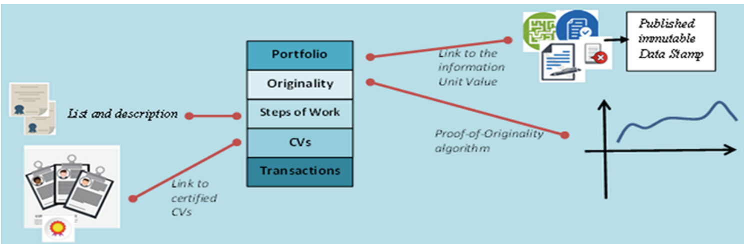


Fig. 1. The description of innovation: the portfolio schema

(6) The construction of a "descriptive" ledger. Up to now the ledger is the container of the transaction referred to the value object transacted. In our case the ledger is a portfolio: the data set in which are described the features of the idea, project, solution that I want to protect. It will contain the following blocks. The link to the project description (in order to maintain the portfolio light). It generates a new chain of block

based on immutable document describing the innovation signed by hashing algorithm. The PoOr algorithm result (the measure of originality). And the information for the trust: the certified steps which allowed the realization of the project; the certified actions to reach the goal; the certified relevant information about the presenters of the proposals. These information could be reserved, but if declared will allow the increasing trust in the project. And at last the information will be certified because connected to a SC that will describe the action done immutably. Our BC implementation ensures:

Idea identification: Unique innovation identifier is provided with contributed information validated as a side chain;
Idea tracing: Allows makers, developers, consultants, producers, collaborators, distributors to provide tracing information in shared ledger with automatic verification of important information;
Idea verification: Creates system and open solution to verify product identifier and other contributed information;
Information requirement: Can create shared ledger of product and transaction information including verification of licensure information.

5 The Proof-of-Originality Algorithm (PoOr)

Innovation is not a bulb that lights suddenly. Innovation is the result of work, study and improvement of other proposals: generally the grade of innovation depends on the improvement. The time of publication is the time-stamp from which the originality is calculated. Originality is not an absolute concept but depend on the time. The algorithm "Proof-of-Originality" analyses the similar proposal structured in a set of sections (including input, outcomes, details) defining the difference about the concepts through a text analysis. Proposals must be described according the same types of sections, in order to make the extraction of concepts comparable. These observables are the features of the proposal and are joined together as bound states. Innovation is the measure of how much an object changes depending on the observables. Text search algorithms will extract the concepts embedded into the observables. It will be evaluated the number of occurrence defined as similitudes, cues, improvements. The value of originality tends to 0 for ideas similar and acquire high value if originality growth. The algorithm is based on the exploration of occurrences between proposals to compare. The first step is the utilization of text mining algorithms, it explores and analyses large amounts of unstructured text data in each topic and identify concepts, patterns, topics, keywords and other attributes in the data extracted. A learning algorithm learns from the proposals analysed. Occurrences could be declared by the authors, found equal or similar by the extraction algorithm. The number of occurrences found is an estimation of originality and describes as the proposal derives from other studies. The distance between the concepts is evaluated in a neighbourhood and it is represented as a cloud of similarity. Originality is defined from a sum of functions described by each cloud and the resulting function can have peaks or valleys, diverging or nullifying. The cloud is evaluated through the difference between the 2 functions: if the density of the cloud increases with decreasing neighbour originality decreases whereas the lower the density the higher the originality. The functions define the state of a system based on the observables. The cloud calculated in the neighbour is a measure of uncertainty. The measure of originality as a part of the chain of blocks describing the proposal is unchangeable and signed with a hash function.

6 Conclusions

The project was born as provocation and as a method to implement collaborative pathways so that the ideas, projects, solutions, and knowledge are transformed into socio-economic value. The starting point is to create a trusted certified network of ideas, projects, proposals, that based on the measure of originality and the trust visibility will give the bases for future collaborations in research and industry, allowing business. The proposal (ideas or innovations) is published as a chain of blocks immutable signed by hash functions. It acquires a measure of originality through the PoOr algorithm. IRI certifies the publication and the activity done during the development of the idea or innovation, based on BC and SC as guarantee for lenders and business angels who want to fund the idea. IRI associates through BC, the ownership of the intellectual work with the sale of the "object" ensuring transparency and correctness. In recent years the certification of a state "de facto" has assumed a sort of acquired intellectual property for jurisprudence. Trust will be generated by the evidence of the network of relationship of the members of the working group. Sharing the portfolio allows all the actors to check the network of relationships which allowed the realization of the idea. This "certification" create trust for who accepts to join the working group or to buy the exploitation of the idea/products created. The challenge we are facing regards the test how BC technology and Smart Contracts can provide trust for all the actors of the innovation ecosystem and generate collaboration and business.

References

1. von Hippel, E.: Free Innovation. The MIT Press, Cambridge (2017). ISBN 9780262035217
2. The Era of Open Innovation. MIT Loan Management Review, vol. 44, no. 3. Spring (2003)
3. Open Innovation: A New Paradigm for Understanding Industrial Innovation. Oxford University Press (2006)
4. Gauntlett, D.: Making is Connecting: The Social Meaning of Creativity, from DIY and Knitting to YouTube and Web 2.0. Paperback – 4 March 2011
5. Satoshi Nakamoto Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf
6. https://en.bitcoin.it/wiki/Proof_of_work
7. https://www.hyperledger.org/
8. https://www.ethereum.org/
9. https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ
10. Bell, L., Buchanan, W.J., Cameron, J., Lo, O.: Application of Blockchain within Healthcare. Blockchain in Healthcare Today, ISSN 2573-8240. https://doi.org/10.30953/bhty.v1.8

Towards a Trusted Virtual Smart Cities Operation Center Using the Blockchain Mirror Model

Emanuele Bellini[1,4(✉)], Alessandro Bellini[1], Franco Pirri[2], and Laura Coconea[3]

[1] Mathema s.r.l., Florence, Italy
emanuele.bellini@mathema.com
[2] Department of Information Engineering, University of Florence, Florence, Italy
[3] SWARCO, Turin, Italy
[4] LOGOS Research and Innovation, Florence, Italy

Abstract. The article proposes a new vision for the smart city operation center combining blockchain and virtual reality. The objective is to achieve high synchronization among the different operators, an enhanced operator performance and the creation of a trustworthy communication layer to record all the decisions taken in the virtual environment and to share data among the operators securely.

Keywords: Blockchain · Internet of Everything · Trust · Smart city operation center · Virtual reality

1 Introduction

A Smart City is a socio-technical system of systems because they unite technology, such as automated processes or sensors, and social and organizational elements, which are embedded in the system and interact each other through complex relations [2]. In fact, we consider a socio-technical system complex if the interactions between single elements result in non-linear outcomes that are hard to predict [3, 4]. This non-linearity often results in “intractability” [5]: even if all involved elements and their individual behaviors are known, the system’s performance is difficult, if not impossible to predict, even for experts [3]. Furthermore, such systems are open to influences from the environment. In complex socio-technical systems, failure can even arise when all involved components are working as expected [6]. These systems are exposed to constant pressures emanating from both internal and external factors, and to which they have to adapt continuously [7]. To maintain the system functional, its human operators need to deal with the uncertainty created by its unpredictable and intractable behavior [7]. Failure-focused safety-approaches are considered insufficient to deal with unexpected disruptions that are likely to happen in such systems. Resilience engineering aims at enabling human operators to keep the system functioning by providing them with the means to make decisions under uncertainty and coping with the local impacts of overall system variability [8, 13].

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 283–291, 2019.
https://doi.org/10.1007/978-3-030-34770-3_22

Public transport, healthcare, public security, and so forth, represent systems of a complex socio-technical SoS (…) highly interconnected and interdependent. A failure of only one of these systems could have severe and unexpected consequences (cascading effects), which is why they should be capable of sustaining operation under both known and unknown conditions.

Enhancing the resilience of urban systems, with particular focus on the response to disruptive events, such as natural catastrophes or acts of terrorism, requires improved cooperation and coordination between multiple stakeholders that operate across urban critical infrastructures and services (urban police, fire department, civil protection, health care, and the like). To address this need a new class of Command and Control tools capable to exchange real-time information among locally distributed decision makers is expected., the growing presence of technology at every system level, pushed by the Smart City trend, which is in turn supported by the Internet of Everything (IoE), makes Knowledge-Information-Data (KID) generated by the city, one of the most critical resources and therefore, one of the key factors to be considered when addressing urban operation [1, 14]. The IoE considers People, Things, Data and Process as a part of a whole ecosystem composed of heterogeneous connected devices and sensors. It establishes an end-to-end ecosystem of connectivity that enables a significant enhancement of the monitoring and control capability in the Smart City, by improving the granularity and breadth of knowledge and awareness about the system status and dynamics, continuously collecting Big Data from heterogeneous data sources/streams. These new possibilities reflect the growing interest in creating the so-called Smart City Operation Center (or Command & Control), where city operators can monitor the status of the city in real time and take more informed decisions during many scenarios, then ever.

However, the data generated by the Smart City is becoming overwhelming. Operators must deal with a considerable number of heterogeneous data simultaneously to manage even more complex situations. Thus, in order to avoid the constant risk of information overload and achieve high operational coordination and resource allocation optimization among existing OCs, they require a trustworthy and wider (potentially unlimited) informative space to share, sense-make, manipulate KIDs and take decisions (actuation), that existing physical OCs, cannot allow. In synthesis there is an urgent need for effectiveness and trust in all Smart City operating phases. This is where the combination of Virtual Reality and permissioned blockchain technology come to help.

2 Trusted Virtual Operation Center

As complex socio-technical systems are confronted by constant pressure to adapt [12], monitoring and decision making are not exceptions but rather a permanent necessity that more and more assumes an iterative nature across multiple stakeholders. A key aspect of the complexity that makes a case for the resilience approach resides in the fact that in socio-technical systems such as cities, people nowadays work "across organizational, geographical, cultural and temporal boundaries" [12]. Standards also recognize that achieving resilience requires organizations to integrate and coordinate their activities also across boundaries with other stakeholders (*BS 65000)* and that information need to be shared in a timely manner with such stakeholders (*ISO/DIS 22316:2016*.). Thus, the growing complexity of socio-technical systems requires a shift from centralized decision making to distributed and collaborative decision making [9]. Leveson and colleagues argued that this requires "top-down systems thinking that recognizes safety as an emergent system property rather than a bottom-up, summation of reliable components and actions" [6]. This means that instead of just focusing on running their own functions as smooth as possible, operators need an updated view on the bigger picture to develop an understanding of how, on one hand, their function contributes, under the given circumstances, to the maintenance of system performance, and on the other hand, their function relies on the performance of other coupled functions with the supply chain or process flows.

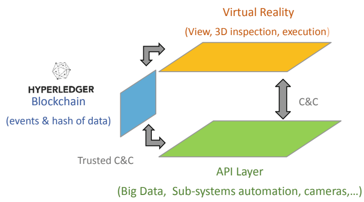


Fig. 1. Trusted virtual operation center concept

Hence, the main reasons to move towards a Trusted Virtual OC (T-VR-OC) are: (1) decision makers require access to reliable multisource data, allowing analyzing the specific system they are dealing with, (2) decision makers needs to reach a better situation awareness while reducing mental workload and information overload and (3) decision makers need to collaborate across organizational borders in a partial trustless condition. The requirement (1) is addressed by the blockchain (blue layer in Fig. 1) that acts as a secure channel of data exchange and event recording. The requirement (2) is addressed by the Virtual Reality (yellow layer in Fig. 1) where high density and heterogeneity data/information are properly represented in order to take well-informed decisions. The requirement (3) is addressed by the replication of the VR based OC in other organizations and the implementation of a consortium based blockchain (Fig. 2).

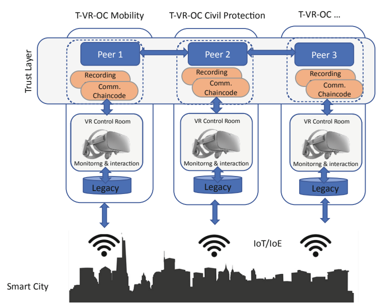


Fig. 2. Scenario of usage of trusted virtual reality operation centers

In [10], for instance, is claimed that the "presentation of alerts should lead the operator through the perceptual (initial detection), decision making and psycho-motor (action required) processes concerned with handling that alert". In classical OC design, alerts are designed for the one specific actor who is capable of manipulating system variables, such as the performance level of sub-systems. For example, a pressure alert is displayed to the person in charge of handling respective valves. This is different in a VR based collaborative OC: usually, a direct relationship does not exist between the alert and what most recipients of the alert may do about it. As an example, consider a road disruption blocking a major street in the city, causing a traffic jam. One operator of the collaborative system could be devoted to adapt nearby traffic lights to redirect incoming traffic while another one might send urban police and ambulance services to take care of the situation. Others might simply need to be informed about the event and the response measures adopted to advise travelers on public transport. Thus, the responses to a critical event in a city should not be as direct as in single-organization control rooms, because of the cascade effects of such an event require the management of high number of informative inputs so that the alerts need to be translated into complex responses. On the other hand, the information on which the decisions are taken should be reliable. Moreover, the decisions taken by the operators need to be recorded in a trustworthy manner for eventual ex-post investigation on misbehavior or missing actions.

2.1 Permissioned Blockchain Based Trustworthy in OC

We consider an OC trustworthy if all incoming and outcoming data are registered in a blockchain ledger and all software – applications or system software in devices (e.g. 3D devices) – can be monitored to detect an eventual infraction or intrusion. Moreover, the KID sharing among organizations can be entrusted using Smart Contract (or chain-code). In fact, in this way it is possible to encode, in a script formal procedure, constraints for inter-organizational communication of sensible information, speeding up operations. Differently from the permissionless one, with the permissioned block-chain (e.g. Hyperledger [22]), it is possible to share only the information that an organization actually wants to share, securing data privacy among the organizations involved in the network (channel feature).

All these activities developed to ensure trusted data and trusted procedures may be designed following the guidance of the theoretical model named Mirror Model (MM). The Mirror Model [18] is a simple formal model that provides the general and formal conditions to yield trusted data and trusted procedures (that in the MM are mathematical functions) i.e. to guarantee by design that a process is safe event in the case of a complex process such that involved in a Smart City Operation Center. In MM, are identified three domains: Reality, Representation and Trust. In particular, "*Reality is where things objectively exist, independently of any efforts made by humans to design or control the Reality, Representation is a model conceived by architects or engineers or scientists to depict the aspects of interest or concern in the Reality and, finally, Trust is a structure of validation of the Representation of Reality whose aim is to make the Representation and its data dependable*" [18]. Through the Trust and Representation domains we can act indirectly on the Reality to make it more reliable and secure for the benefit of humans and the environment in which they live. Thus, according to the Mirror Model for blockchain adoption on real life scenario, the T-OC can be entrusted as depicted in Fig. 3.

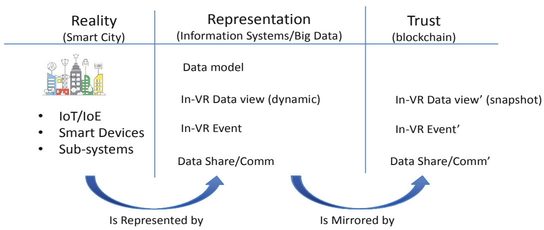


Fig. 3. Mirror model based T-OC design

It is worth to notice that the blockchain is not used to mirror the data model used at the Representation Level, but only to record the event and the snapshot of the data related to it. In fact, the blockchain should not be considered as a general-purpose storage like a database with some security features. Its application should be considered only when evidences related to a designed event needs to be shared and preserved in an untampered way [20, 26]. In our context, the events related to user interaction occurred in the VR that can be intercepted are: clicks, widget open/closed, widget relocations, points of view change, 6-Axis body movement, etc.

The technology adopted in the experiment is Hyperledger Fabric [22], a permissioned blockchain that allows peers to interact after a X.509 certificate-based authentication. Moreover, it supports chaincode implementation, a program written in a generic programming language (e.g. Go, Node.js or Java) that is executed on a distributed system composed by the network of the peer nodes. Hyperledger is configured according to the consortium dimension and it is distributed with virtualizing docker technology.

A chaincode is connected to the virtual environment and is invoked each time an event occurs registering type of the event, time, user, and any other information useful to identify the action and who has performed. It is also important to record the status of the system at the time of the event t_e. Thus, at t_e, a snapshot of the data available/displayed in the VR is cooperatively saved with Interplanetary File system (IPFS) [21] within the blockchain consortium. This scenario is depicted in Fig. 4.

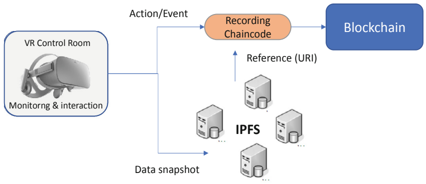


Fig. 4. Event recording

Another critical action on OC that can be entrusted by the introduction of permissioned blockchain, is the cross-organization data sharing and communication. According to MM, the data and communication can be mirrored into the blockchain and shared exploiting the blockchain secure ledger replication, so that the organization belonging to the consortium can directly access to the data from its own node instead of investing resources to avoid sniffing and other attacks on the communication channel (Fig. 5).

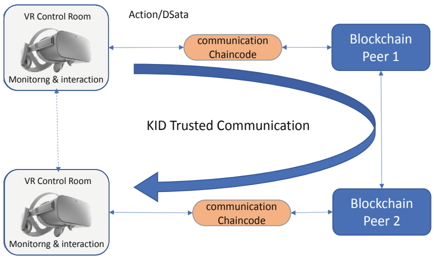


Fig. 5. Blockchain-based KID trusted communication

The data sharing is supported by a chaincode deployed on the Hyperledger peers that allows data sharing along the organizations involved in the consortium (running a peer mode) and belonging to a specific channel.

2.2 Virtual Reality for OC

The usage of VR provides a complete control over the wearer's visual and auditory experience as they interact within a completely synthetic environment. This control over the environment can provide virtual experiences of either subdued or amplified versions of reality [17]. In this respect, a VR based OC extends the physical environment and devices capability in two dimensions: interactions and space. The interaction capability is enhanced including a number of new classes of action based on spatial gestures, movements and sounds. The possibility to manipulate and inspect artefact in a 3D fashion allows users to be more effective in managing information complexity. Moreover, the capability of the VR of being customized, portable, cost effective and easier to create mental models of the system's functioning [11], makes this technology attractive in different domains such as nuclear plant [16] or medicine [17].

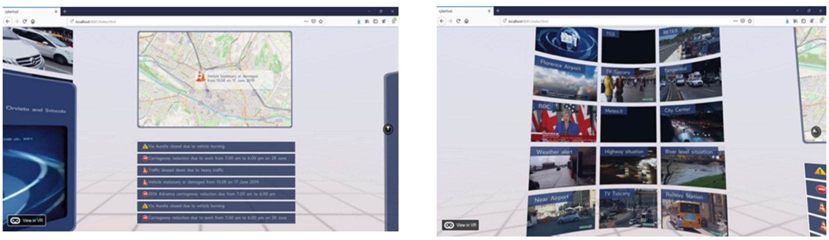


Fig. 6. VR OC - view

The technology adopted in the experiment is based on head mounted OCULUS Rift hardware [23], the open source framework (REACT 360) to develop the virtual environment and a PC with a dedicated 3D NVIDIA graphic board. The decision to use OCULUS is related to the availability of software libraries, while the decision of exploiting open source framework respect to the well-known solution as Unity, is derived by some restrictive condition of usage that could be an obstacle in case of business distribution of the tool. In Fig. 6 is represented a multi-widget video wall with cross-widget interaction that represents streaming videos, maps, tables, chart, etc. into a semi-circled environment. In this way the operators are involved in an environment with high information complexity but with the possibility of manipulating and organizing the space according to his way of thinking. The widgets/objects can be opened, closed, moved and positioned in any part of the 3D space (up, down, in front, behind, left, right). In order to enhance the interaction capability of the operators, have been considered the exploit gestures and body movements as a source for triggering actions in the VR. Body movement is characterized with respect to the sagittal, vertical and transverse axis of the body. Its detection requires sensors able to implement at least six degree of freedom (DOF). Secondly, the tracking of body movement, requires sensors capable to acquire samples at the appropriate rate: for instance, the frequency range of human body motion is around 10 Hz [24]. In order to implement a body motion solution able to track the intended variables, has been selected the open source UDOO Go board with 9 axes[1], a microchip that integrates accelerometer, gyroscope and magnetometer. The implemented 3 chips mounted on one arm to track its movement. This solution has been combined with the sensors embedded in Oculus that already tracks head movement for spatial navigation. Then, a number of gestures/movements have been associated to events in VR. For instance, each time the operator extends his arm frontally, the selected object in VR is moved far away and vice versa. The gesture mapped in the study are: arm extension and rotation. Each of them is associated to specific operation as virtual keyboard activation-deactivation and widget movement/positioning. The interactive system, to be properly used, requires a training phase for the operator as well as a calibration phase in order to enhance the precision of the interaction.

3 Conclusions

This article presents the first result of a work-in-progress industrial research and development activities on Trusted Virtual Reality Operation Center for Smart City integrating VR and Blockchain technology together. The proposed system can be evaluated according the framework proposed in [19]. In particular the solution is *non-regressive* since the validity of the approach is not affected by the introduction of new features (e.g. recording more events in the control room). It also may trigger *rebound* effect creating additional demands (new features) thanks to the increment of trust and interaction capability [19]. The results show that the usage of VR to remotely operate in a most effective and cost-reduced way, combined with the capability of permissioned

[1] http://www.udoo.org/download/files/datasheetsdatasheet_udoo_go.pdf.

blockchain of supporting trusted events recording and data sharing among different organizations, is promising. Finally, the effort needed to obtain the same functionalities in the classical operation center is strongly reduced (1/100).

According to this evaluation, the usage of virtual reality (VR) headsets to remotely operate in a most effective and cost-reduced way combined with the capability of permissioned blockchain of supporting trusted decisions recording and data sharing among different organizations, is promising for the future of complex system command and control.

Acknowledgment. The present work is partially supported by e-Kinesys project funded by Fondazione Cassa di Risparmio di Firenze.

References

1. Bellini, E., Nesi, P.: Exploiting smart technologies to build Smart Resilient Cities. In: Gardoni, P. (ed.) Smart Cities and the Role of Information and Communication Technology on achieving Sustainability and Resilience Book. Univ. Illinois - Taylor & Francis (2018). https://doi.org/10.4324/9781315142074-35
2. Tschiersch, I., Schael, T.: Concepts of human-centred systems. In: Brandt, D. (ed.) Human-Centred System Design. FIRST: PEOPLE, Second: Organization, third: Technology; 20 Case Reports, Aachener Reihe Mensch und Technik 42, pp. 97–107. Wissenschaftsverlag Mainz, Aachen (2003)
3. Leveson, N.: A new accident model for engineering safer systems. Saf. Sci. **42**(4), 237–270 (2004)
4. Mansfield, J.: The nature of change or the law of unintended consequences. An introductory text to designing complex systems and managing change. Imperial College Pr, London (2010)
5. Hollnagel, E.: Prologue: the scope of resilience engineering. In: Hollnagel, E., Pariès, J., Woods, D.D., Wreathall, J. (eds.) Resilience Engineering in Practice. A Guidebook. Ashgate Studies in Resilience Engineering, pp. xxix–xxxix. Ashgate, Aldershot (2011)
6. Leveson, N., Dulac, N., Marais, K., Carroll, J.: Moving beyond normal accidents and high reliability organizations: A systems approach to safety in complex systems. Organ. Stud. **30**(2–3), 227–249 (2009)
7. Jackson, S.: Architecting Resilient Systems. Wiley, Hoboken (2009)
8. Hollnagel, E.: Resilience – the challenge of the unstable. In: Hollnagel, E., Woods, D.D., Leveson, N. (eds.) Resilience engineering. Concepts and precepts, pp. 9–17. CRC Press, Ashgate, Aldershot (2006)
9. Schneeweiss, C.: Distributed Decision Making. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-24724-1
10. Crampin, T.: Human Factors in Control room Design: A Practical Guide for Project Managers and Senior Engineers. Wiley, Hoboken (2017)
11. Bockelmann, M., Nachreiner, F., Nickel, P.: Bildschirmarbeit in Leitwarten. Handlungshilfen zur ergonomischen Gestaltung von Arbeitsplätzen nach der Bildschirmarbeitsverordnung; Forschung Projekt F 2249. Bundesanst. für Arbeitsschutz und Arbeitsmedizin, Dortmund, Berlin, Dresden (2012)
12. Carayon, P.: Human factors of complex sociotechnical systems. Appl. Ergon. **37**(4), 525–535 (2006)

13. Bellini, E., et al.: Towards resilience operationalization in Urban Transport System: the RESOLUTE project approach. In: Risk, Reliability and Safety: Innovating Theory and Practice - Proceedings of the 26th European Safety and Reliability Conference, ESREL 2016, p. 345 (2017)
14. Bellini, E., Ceravolo, P., Nesi, P.: Quantify resilience enhancement of UTS through exploiting connected community and internet of everything emerging technologies. ACM Trans. Internet Technol **18**(1), Article 7 (2017). 34 pages. 27. https://doi.org/10.1145/3137572
15. Bergroth, J.D., Koskinen, H.M.K., Laarni, J.O.: Use of immersive 3-D virtual reality environments in control room validations. Nuclear Technol. **202**(2–3), 278–289 (2018). https://doi.org/10.1080/00295450.2017.1420335
16. Kotek, L., Tuma, Z., Blecha, P., Nemcova, Z., Habada, P.: Evaluation of human error of response to auditory and visual signals in the virtual reality
17. Silva, J.N.A., Southworth, M., Raptis, C., Silva, J.: Emerging applications of virtual reality in cardiovascular medicine. JACC Basic Transl. Sci. **3**(3), 420–430 (2018)
18. Bellini, A., Bellini, E., Gherardelli, M., Pirri, F.: Enhancing IoT data dependability through a blockchain mirror model. Future Internet **11**, 117 (2019)
19. Bellini, E., Coconea, L., Nesi, P.: A functional resonance analysis method driven resilience quantification for socio-technical systems. IEEE Syst. J. (2019). https://doi.org/10.1109/JSYST.2019.2905713
20. Brotsis, S., et al.: Blockchain solutions for forensic evidence preservation in IoT environments. In: 2019 IEEE Conference on Network Softwarization (IEEE NetSoft) (2019)
21. Interplanetary File System (IPFS). https://ipfs.io/
22. Hyperledger Fabric. https://www.hyperledger.org/projects/fabric
23. Oculus Rift. https://www.oculus.com/?locale=en_US
24. Zeng, H., Zhao, Y.: Sensing Movement: microsensors for body motion measurement. Sens. (Basel) **11**(1), 638–660 (2011)
25. Bellini, E., Nesi, P., Pantaleo, G., Venturi, A.: Functional resonance analysis method based-decision support tool for urban transport system resilience management. In: Proceedings of IEEE 2nd International Smart Cities Conference: Improving the Citizens Quality of Life, ISC2 2016 (2016)
26. Shala, B., Trick, U., Lehmann, A., Ghita, B., Shiaeles, S.: Blockchain-based trust communities for decentralized M2M application services. In: Xhafa, F., Leu, F.-Y., Ficco, M., Yang, C.-T. (eds.) 3PGCIC 2018. LNDECT, vol. 24, pp. 62–73. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-02607-3_6
27. Bendiab, K., Kolokotronis, N., Shiaeles, S., Boucherkha, S.: WiP: a novel blockchain-based trust model for cloud identity management. In: 2018 IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/Cyber SciTech) (2018). https://doi.org/10.1109/dasc/picom/datacom/cyberscitec.2018.00126

Monitoring an Environment Using Wireless Sensor Network

Rayanne Maria Cunha Silveira[1(✉)], Thalyson Danilo Rocha Dutra[2], Francisco Paulo Roberto Sampaio Alves[1], Allyx Fontaine[3], Thierry Desjardins[4], Raimundo C. S. Freire[5], Allan K. D. Barros[1], and Ewaldo E. C. Santana[6]

[1] Departament of Electrical Engineering, Federal University of Maranhão, São Luís, Brazil
rayanne.silveira@aluno.ecp.ufma.br
[2] Departament of Computer Engineering, Federal University of Maranhão, São Luís, Brazil
[3] Université de Guyane, UMR Espace-Dev, Cayenne, France
[4] Institut de Recherche pour le développement, Marseille, France
[5] Departament of Electrical Engineering, Federal University of Campina Grande, Campina Grande, Brazil
[6] Departament of Mathematics, State University of Maranhão, São Luís, Brazil

Abstract. In the past few years, Internet of Things has become of common use in various areas, and especially in agriculture. There are many ways to increase the production using technology on agriculture. One of the most used is monitoring systems to follow characteristic properties and, in some cases, to control them by triggering a process that would correct an unwanted situation. Therefore, knowing the usefulness of this type of system, the present work proposes the monitoring of an environment using Internet of Things (IoT): Wireless Sensor Networks (WSN) are placed in strategic location in order to collect and send data to a web service. Currently, two topologies were tested to monitor the temperature and soil moisture of a soccer field. The samples were stored in a database and the next step is to integrate the system within the web application that we have already developed.

Keywords: IoT · Monitoring · Wireless Sensor Network · Agriculture

1 Introduction

Agriculture is an extremely important activity because of the source of various elements of human food it provides, from vegetables to grains. With the increasing technological development, this sector has received several tools that contribute to the increase of production and quality control of the same. In many countries this activity also plays a fundamental role in the economy, as in Brazil where it is one of the main activities.

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 292–301, 2019.
https://doi.org/10.1007/978-3-030-34770-3_23

In this context, Brazil has invested in technologies to improve the sector. In terms of output growth, Brazil generally opts for productivity gains from the application of improvements in factors of production [7]. One way to increase this production is automating the environment, using sensors and physical parameters to control factors of the environment, or simply monitor the environment to follow these physical parameters.

Therefore, observing environment is an alternative that gives clues to have gains in the agriculture production. One of the most used technologies to build such system is Wireless Sensor Networks (WSN). According to [4] the use of WSN in conjunction with Internet of Things (IoT) has a promising future, because they allow the global interconnection of heterogeneous physical objects.

In this context this paper proposes a WSN to monitor an environment collecting soil information to help the analysis of this area. The proposed tool is part of an international project whose aim is to collect automatically and frequently data in order to see the impact of soil features in the environment. Currently, two topologies were tested to monitor the temperature and soil moisture of a soccer field. The samples obtained from the experiments were stored in a database and the next step is to integrate the tool within the web application developed in [6].

The paper is organised as follow. In Sect. 2, we present related works. Section 3 describes the proposed tool: the hardware and the environment used to do the experiments. We give the results in Sect. 4 before concluding in Sect. 5.

2 Related Work

Monitoring an environment using different technologies is a recurrent alternative to solve many agriculture problems, and it has become a topic of interest unmissable those last years. We present in this section some of the existing systems used to monitor an environment and different applications of those tools in agriculture.

There are different technologies applied to build a monitoring system in the area of farming. In [1], the authors propose a web platform, called CLUeFARM, to monitor farms from a distance through this web application. In [5] the same platform was used to monitor exclusively the soil, and the results were positive, showing that this kind of application contributes to the management of farms and agricultural activities.

Dasgupta et al. [2] has developed, beside the monitoring system, a mechanism to control the irrigation of large- and small-scale plantation. This mechanism has two operating modes, one manual and the other autonomous. The autonomous mode uses a data collection provided by a monitoring system to define the rate of irrigation according to the present physical parameters. In this system, two sensors were used: the DHT11 for the ambient atmospheric temperature and humidity and a resistive humidity sensor for the soil humidity. They use the Arduino as a controller.

Yue et al. [8] has developed a monitoring system to feed an Artificial Neural Network used to measure the progression of the monitored parameters in an infrastructure and to additionally create correlations among them. The experimentation were led during two months and consisted in monitoring the water content of soil and some other parameters. Unfortunately, the hardware (type of sensors) used is not explicit in the paper and we do not have a basis of comparison to our work.

In [6], a WSN is proposed to monitor the soil with the main objective to build a big database. To give a friendly access to the collected data obtained from the WSN a Web-Service has been provided. The WSN was made using Xbee to send data and two sensors: LM35 for temperature, and the same we use for humidity. The authors used the NodeMCU ESP8266 to send the collected data to the server.

Our tool is based on sensors and a controller chosen according to some environmental constraint and to the cost. Indeed, the aim of the tool is to cover an area of hundred of kilometers, therefore a big amount of sensors will be deployed on an hostile area for electronic components (very humid). In fact, a huge constraint in Brazil is corrosion and the capacitive humidity sensor SEN0193 of DFRobot we use has a special protection to prevent it. The temperature sensor used is water proof, unlike LM35, then it is more appropriate to deal with humid environment. Our choice for the controller, ESP32, is based of the WiFi connection it provides. No other device to complement it is therefore necessary. Furthermore, the ESP32 is cheaper than Xbee, given that a large number of nodes is necessary for the covering of the area. All the chosen technologies are compatible with the web-service developed in [6]. The final objective is to implement our tool within this platform.

3 Monitoring an Environment: An Implementation Using WSN

As mentioned, the use of technology to increase the agriculture production has grown in the last few years. Environmental monitoring is one of the most popular and efficient tools since it provides the farmer with a profile of this environment, generating data that can be used as a set point for the activation of mechanisms or even to point out unwanted scenarios.

In agriculture, two of the most important properties to monitor are the soil temperature and humidity. According to [8] they perform an important role on the crops, having impact in the growth, quantity and quality of the production.

3.1 Wireless Sensor Network

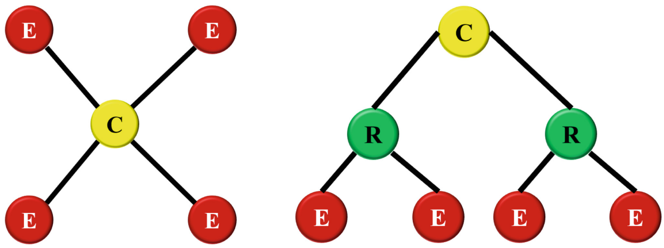
We have used a Wireless Sensor Network to monitor the selected environment. In this section, we describe the technology that has been used, the topologies that we have tested in our experimentation as well as the quantity of devices.

The components of the network, called nodes, are devoted to different functions. They are classified according to three types: end device, router and coordinator. The end device is responsible for collecting samples via its sensors.

The router sends the packets containing the data of the samples until they reach the coordinator. The coordinator receives and organizes all collected data and makes them available to the final user [3], in our case the Web Service. The technology used to build the node of our WSN is the ESP32 devKit. We describe it properly in Sect. 3.2.

Another important aspect to be defined is the topology, that is how will be all the devices linked to each other. We have tested and compared two different topologies: the star topology and the tree topology (Fig. 1). In the star topology, there is a central node (the coordinator) and all the other devices are connected to it. In the tree topology, the root is the coordinator, the leaves are the end devices and the internal nodes are the rooters.

The number of devices used in our experimentation was low. Currently, we have just developed a prototype that will be adapted to the final application on an area of around hundred kilometers. In the star topology we have used four end devices and one coordinator, as showed in Fig. 1a. And, in the tree topology was used four end devices, two routers and one coordinator, as illustrated in Fig. 1b. In each end device was connected one temperature sensor, one humidity sensor and one 450 mAh battery.



(a) Star Topology (b) Tree Topology

Fig. 1. Topologies of Wireless Sensor Network

In both topologies the data collected by the end devices are organized by the coordinator node to be published in the web platform of the Guyamazon project SenCSoil. The aim is to have all data collected online to facilitate the access to all the information. Therefore, the coordinator also works as a gateway and sends the data to the cloud. This setting is only possible because the ESP32 has a WiFi connection and does not need any other device to play this role.

3.2 Hardware

As mentioned before, the main device used to build the WSN is the ESP32 devKit, shown in Fig. 2. This device has a microcontroller, an interface for many types of sensors and is able to send packets using wireless connection.

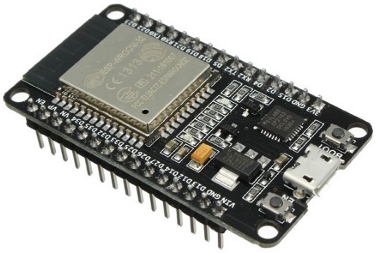


Fig. 2. ESP32 devKit

The communication protocol of the WSN was implemented in the ESP32 using the Arduino IDE. In star topology two different algorithms were embedded, one for the coordinator (Algorithm 1) and the other for all end devices (Algorithm 2). In the first one, the device listens to a defined channel, waiting for a packet, when a packet is found it receives and organizes the data. The second one makes one read using the sensors in a defined interval of time. The collected values are organized and sent to the coordinator, after the reading step, the device sleeps until the next read, to save energy. In the tree topology the end device sends to the router node instead of the coordinator, the router device has a different algorithm (Algorithm 3): listening to the channel, receiving the packet and sending to the coordinator. The coordinator algorithm is the same used as in the star topology.

Algorithm 1. Coordinator node Algorithm

```
Coordinator(){
    DefineChannel();
    if(SomethingOnChannel()){
    ReceivePacket();
    }
    else{
    stillListen();
    }
ReceivePacket(){
    ReadMessage();
    SavetheData();
}
```

In addition, this device has a lower cost than many technologies, such as Xbee. Note that, contrary to the ESP32 devKit, a device such as Xbee, in addition to its higher price, needs an extra microcontroller as an Arduino for instance.

The ESP32 also provides a low consumption of energy, by using for the end devices the sleep mode to save energy when it does not take a measure.

Algorithm 2. End Device Algorithm

```
EndDevice(){
    DefineChannel();
    ReadSensors();
    BuildPacket();
    SendPacket(Coordinator or Router);
}
```

Algorithm 3. Router Node Algorithm

```
Router(){
    ListenChannel();
    if(SomethingOnChannel()){
        ReceivePacket();
        SendPackettoCoord();
    }
    else{
    stillListen();
    }
```

The complete node is equipped with connectors to add sensors, a 9 V battery and a voltage regulator - to ensure that the right voltage will be sent to it.

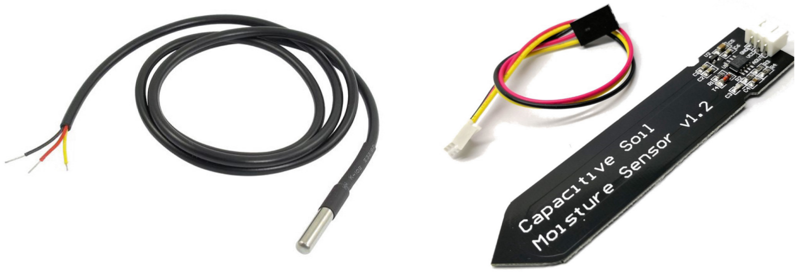
In this experimentation, we have used two kind of sensors: a temperature sensor and a humidity sensor. Both sensors will be fixed on the soil. Each read has been launched within an interval of 2 h leading to 25 samples per end device.

Temperature Sensor. The temperature sensor used was the DS18B20, a water proof, low cost sensor and 1-wire Bus connection. The sensor, shown in Fig. 3a, fits perfectly to the application, because of its size and probe format that facilitates soil fixation.

The DS18B20 can give the temperature value in Celsius, Kelvin or Fahrenheit, but in this experiment all values are in Celsius. This device has a low cost (less than one dollar) and operates with a voltage compatible with ESP32 (3.3 V–5 V).

Humidity Sensor. The humidity sensor used was the DFROBOT SEN0193, shown in Fig. 3b. This sensor is a capacitive sensor that, compared to the resistive sensor, provides greater accuracy. Also contrary to resistive sensors, this sensor does not have problems with the corrosion.

To make the data acquisition, the sensor was threaded into the ground respecting the limitation given by the manufacturer. Also a calibration was made to define the equivalence between the read value and the soil humidity. To get the upper and lower bounds, various reads were made with the sensor fixed in



(a) DS18B20 Temperature Sensor (b) Capacitive Soil Humidity Sensor

Fig. 3. Used sensors

a dry soil, fixed in water and outside the soil. The humidity was defined in the following scale: value on water - 100% humidity, value outside the soil - 0% humidity.

3.3 Environment

Our experimentation has been led in a soccer field with a area of 1225 m^2 (see Fig. 4) in a inner city of Maranhão, Brazil.

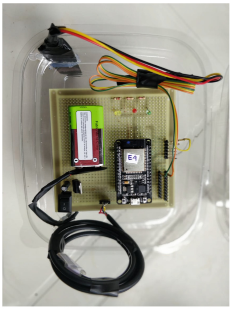
In the future, our prototype will be used to collect information on a huge area with various kind of vegetation in order to determine the soil characteristics for research studies, but also many other applications to irrigate the field, follow the growth of the plants, etc.

Fig. 4. Observed soccer field

As mentioned earlier, just a few nodes, 4 for the star topology and 6 for the tree topology, were used to build the proposed WSN. Therefore in the star topology, just half of the field was monitored. The distance between each node was 2 m, and the more distant node was 15 m from the coordinator. In the tree topology, it was possible to locate the nodes less close from each other, however the coverage area is still the same because of the quantity of end devices.

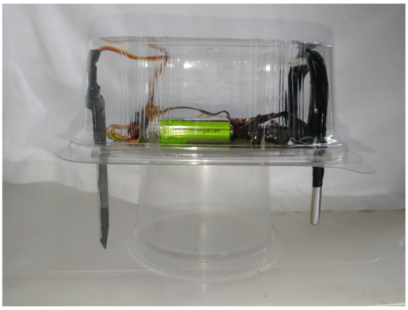
4 Results and Discussion

All devices has been fixed on arbitrary locations: as mentioned previously they are 2 meters distant from each other in both topologies. To carry out the experiment a prototype was developed, see Fig. 5, using a plastic protection to prevent natural events, such as rain, that may reach the device or damage the board.



(a) Board of End Device Node

(b) Device Fixed on Soil



(c) End Device Node Lateral View

Fig. 5. Developed prototype

The 9 V batteries were sufficient to manage the experimentation of 2 h each one. The batteries were used just in end devices, the routers were with constant energy supply. The coordinator was connected to a computer to have constant energy supply and to send the data to a database.

The used sensors have a good behavior, resisting to big changes of humidity and temperature, and its shape was ideal for the application, since both were fixed to the soil without any problem. Future tests will be made in order to integrate this WSN into the Web Application used to publish the data, proposed in [6] and with a bigger interval of reads.

In each experiment, illustred in Fig. 6, using star topology the devices were fixed in the red X position and using the tree topology the devices were fixed in the blue X position and the routers are represented by the R. We have taken as a constraint to use the same number of end devices for the two topologies. A total of 25 samples from each end device has been collected, in an interval of 2 h. Each sample has one value of identification of the device, one temperature value and one humidity value. All collected temperature and humidity values are plotted in the graph in Fig. 7.

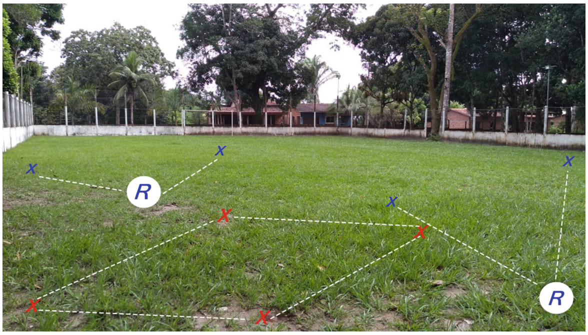
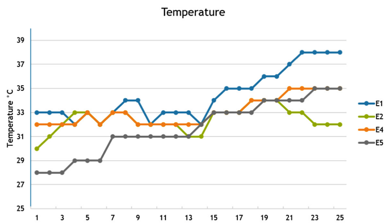
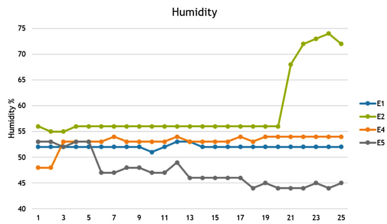


Fig. 6. Proposed configuration of WSN

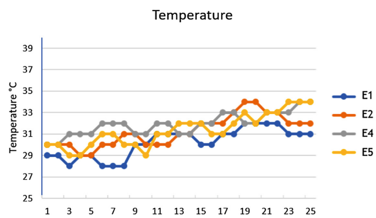
In the same Fig. 6, it is possible to notice the main advantage of the tree topology compared to the star topology, because with the first the range of the network is bigger. For the star topology, it is necessary that all end devices reaches the coordinator, so the range is smaller. The use of router devices on



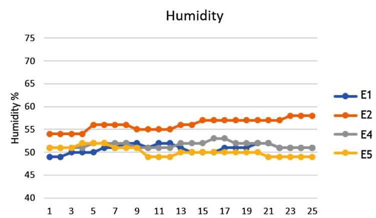
(a) Temperature Values Star Topology



(b) Humidity Values Star Topology



(c) Temperature Values Tree Topology



(d) Humidity Values Tree Topology

Fig. 7. Plot of collected data

tree topology allows us to reach the other side of the field, that is to say to obtain a good covering of the field.

In Fig. 7b one can see that high values were received, this occurred due to the contact of a small quantity of water with the soil in this region. In the others figures one can see that there are no notable changes during each 5 min.

5 Conclusion

The use of WSN to monitor areas is very common and more and more popular. However, different technologies to build those networks are constantly studied to see what type of network fits better with a defined scenario. In this paper a brief study was made using the ESP32 as node, DS18B20 sensor to read the soil temperature and a capacitive humidity sensor to collect the soil moisture.

An experiment was made and it has shown that the WSN is able to collect the desired data. In the future the data collected will be available in a Web Application, to make more easy its visualization by researchers and farmers. Another work is to test other sensors and to use different intervals, according to the desired features we want to measure.

References

1. Colezea, M., et al.: CLUeFARM: integrated web-service platform for smart farms. Comput. Electron. Agric. **154**, 134–154 (2018)
2. Dasgupta, A., Daruka, A., Pandey, A., Bose, A., Mukherjee, S., Saha, S.: Smart irrigation: IOT-based irrigation monitoring system. In: Chakraborty, M., Chakrabarti, S., Balas, V.E., Mandal, J.K. (eds.) Proceedings of International Ethical Hacking Conference 2018. AISC, vol. 811, pp. 395–403. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1544-2_32
3. Faludi, R.: Building Wireless Sensor Networks: With ZigBee, XBee, Arduino, and Processing. O'Reilly Media Inc., Sebastopol (2010)
4. Garcia, G.T., et al.: Wireless sensor network for monitoring physical variables applied to green technology (IoT green technology). Eur. J. Electri. Eng. Comput. Sci. **2**(2), 1–7 (2018)
5. Serrouch, A., Mocanu, M., Pop, F.: Soil management services in CLUeFARM. In: 2015 14th International Symposium on Parallel and Distributed Computing (ISPDC), pp. 204–209. IEEE (2015)
6. Silveira, R.M.C., Alves, F.P.R.S., Fontaine, A., Santana, E.E.C.: A web-service to monitor a wireless sensor network. In: Miller, J., Stroulia, E., Lee, K., Zhang, L.-J. (eds.) ICWS 2019. LNCS, vol. 11512, pp. 126–146. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23499-7_9
7. Von Zuben, M., Velini, E.D., de Barros, A.M., Cristina, T., Waak, W.: Fórum inovação para sustentabilidade na agricultura. AgroANALYSIS **38**(7), 31–37 (2019)
8. Yue, R., et al.: Soil temperature and humidity monitoring system design for farm land based on ZigBee communication technology. DEStech Trans. Environ. Energy Earth Sci. (ICEPE) 191–195 (2019)

Life and Death of Data in Data Lakes: Preserving Data Usability and Responsible Governance

Marzieh Derakhshannia[1], Carmen Gervet[2], Hicham Hajj-Hassan[3], Anne Laurent[1](✉), and Arnaud Martin[4]

[1] LIRMM, Univ Montpellier, CNRS, Montpellier, France
dm.derakhshannia@gmail.com, anne.laurent@umontpellier.fr

[2] Espace Dev, Univ Montpellier, IRD, Univ. La Runion, Univ de Guyane, Univ des Antilles, Montpellier, France
carmen.gervet@umontpellier.fr

[3] CNRS-L, Beirut, Lebanon
hishamhh@cnrs.edu.lb

[4] CEFE, Univ Montpellier, CNRS, Montpellier, France
arnaud.martin@umontpellier.fr

Abstract. Data crossing seeks the extraction of novel knowledge through correlations and dependencies among heterogeneous data, and is considered a key process in sustainable science to push back the current frontiers of knowledge, especially to address challenges such as the socio-economic impacts of climate change. To tackle such complex challenges, interdisciplinary approaches and data sharing methodologies are ubiquitous, with a strong focus on data openness and ensuring that the fair principles hold. Data lakes are data repositories, recently developed to store such big heterogeneous data that are then available for crossing and be exploited without a priori objectives regarding their usage (unlike data warehouses). Such data lakes can then be used to populate Open and Linked Open Data in a central location regardless of its source or format. In this context of no prior knowledge regarding its usage, it may be tempting to store and share all the available data. However, this comes with two main disadvantages: (1) overwhelming amount of data that could prevent end users from exploiting the data, (2) and environmental reasons (energy consumption of data storage). Moreover, data of poor quality may deserve the lake usability and be deleted. We thus claim in this position paper that a data life cycle must be designed so as to integrate data death for some of the data. The choice of the data to be stored regarding the ones to forget is then of crucial importance in data lakes. We propose here some first positions for this aspect of data governance.

Keywords: Data lakes · Web of data · Data life cycle and data governance · Sustainability

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 302–309, 2019.
https://doi.org/10.1007/978-3-030-34770-3_24

1 Context

Data crossing is often considered as a key for discovering new knowledge and developing advanced models for phenomena understanding, as for instance climate change and its impacts on societal impacts For this reason, data sharing, especially data opening through Internet and the Web of data is a crucial challenge.

Data lakes [9] have been designed for allowing data retention before their usage. In such data repositories, data are made available from various data sources and can be exploited by data analysts to serve end users' needs with the idea that *the larger the data, the better the results*.

However in this paper, we state the conjecture that data perennity is no longer an option for a sustainable science despite the availability of massive data storage capacities. This raises the challenge of the lifespan of stored data, while maintaining its ability to generate the knowledge necessary for the users over time (consumers). How to make decisions regarding the lifetime and death of data when you only have the present knowledge of its usage, without further planning? How to ensure that the decisions taken regarding each data lifespan, with not entail a potential future loss of knowledge?

We consider mainly the case where the laws regarding the life cycle of the data is not imposed (ex: GDPR that forces the destruction or confidentiality of personal data).

We might think that a strong element to guide such decisions is to account for the data producer and consumers profiles. However, there is no guarantee that current behavior might dictate properly the 'forgetting function' definition. Thus in this position paper, we wish to distinguish both aspects, and not assume that they are necessarily interdependent over a certain time interval. This brings about the question of determining the criteria underlying the formal definition of the forgetting function (most probably time dependent), and thus the death concept.

2 Background

Data bases have been studied for the last 50 years from the very beginning of computer science. When dealing with data, their management is a key topic has been intensively addressed. The way they are collected, stored and exploited is managed through the so-called data governance process [7]. Regarding scientific data in particular, the concept of data life cycle[1], described by Fig. 1 is often referred to as a key principle. In such a process, the question of deleting data and purging data bases has been discussed.

In particular, the question of big data management imposes to consider the infrastructures to be deployed. Some works have pointed out the necessity to build large data repositories in order to keep them sustainable [6].

[1] https://www.dataone.org/data-life-cycle.

In the context of big data, data lakes are one of the emerging models to manage such huge repositories. In [9], data lakes have been defined as *a logical view of all data sources or data set, in their raw format, available and accessible by data scientist or statistician to find new insight.*

- *A data lake is governed by a metadata sources index to guarantee the data quality.*
- *A data lake is controlled by rules, tools and processes to guarantee the data governance.*
- *A data lake is limited to data scientist or data statistician access to guarantee data security, data privacy and compliance.*
- *A data lake access all type of data.*
- *A data lake has a logical and physical design.*

A data lake is a key element of the information architecture and a new step on its evolution.

Data lakes must be distinguished from data warehouses as data warehouses are rather dedicated to answer a priori known key performance indicators regarding several analysis dimensions, which impose a fixed multidimensional model, data sources and data preprocessing while data lakes are repositories for raw data whose further usages are not a priori known.

Indeed, [2] and [3] consider data lakes as a data storage management placement populated by all data sources in a raw format or as-is data. Data governance is pointed as being crucial and the information architecture is discussed. In the IBM redbook [5], the notion of data reservoir is associated to the data lake term and used as same meaning. The concept a metadata catalogue guarantees the respect of data governance to prevent the data lake to transform into a data swamp.

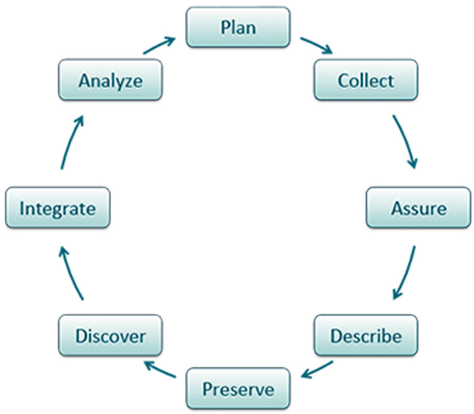


Fig. 1. Data life cycle.

Internet of things supports such approaches since sensors produce huge volumes of data that are not all consumed but that are usually considered as *potentially interesting*. It should be noted that data lakes store raw data, which excludes preprocessing. However, new models are emerging with fog computing that consider that some treatments can be embedded within the sensors.

3 Method

In this paper, we claim that some data must be deleted from (or not even put in) data lakes by the organization in charge of the lake. The questions are then:

- can some data be retained for a (given) period, never put into the data lake or even deleted from the lake at some period? For example, we may think about an analogy with the withdrawal of a medicinal product that threatens or could involve risk patients.
- when data is put into the data lake, to which spatial and temporal scale must the data be considered?

We report and discuss below some first research positions:

- providing solutions to extend naive data aggregation and sampling (e.g., choosing data distributions);
- providing decision support systems to help end users deciding on where to point the cursor between storing all data as raw data or preparing and sampling data;
- relying on and copying supply chain principles to decide what data and when putting into the data lake.

3.1 Forgetting Functions

When dealing with large data with the idea of saving storage memory consumption, it may be interesting to compress the data. This is even the case before storage when data are collected as sensors are designed for measuring at some time periods.

[1] has proposed so-called *forgetting functions* in data warehouses with two main principles, either sampling the data or aggregating.

We propose to extend this approach so as to be used in data lakes. The main challenge is then that it is difficult to evaluate the *utility* of the data regarding their usage as data lakes are not meant to serve *immediate* needs but rather store data in case it may be useful *in the future*.

Moreover, if the data are aggregated in order to be compressed, we claim that distributions must be used rather than simple functions as *mean, median, min, max*. This will allow to describe and model the uncertainty, imprecision and variability of the data. Formalisms as fuzzy logic or possibility distributions may be used for this purpose.

3.2 Providing Decision Support Systems

The goal here is to design and build decision support systems that will drive the management of available data. As previously mentioned, storing and managing all the data fed to the lake is antinomic with a sustainable science. Thus some data must either be removed from the lake or idled, compressed and left aside. The criteria that will guide the decision making need to be defined. A common approach can be to define criteria through usage, frequency, etc. We state that other properties of the data can be relevant to maintain a sustainable data lake, or data lakes, as a dynamic resource. Considering the features of interest is one venue that might hold over a certain period but might be reconsidered further on. Part of the research is to seek functions that will determine criteria for managing the life span of the data.

3.3 Imitating Natural Data Lakes

Natural lakes are ecosystems with many organisms which are subjugated (governed) by chance and necessity. Chance is gene mixing, mutations, etc. Life reproduces itself with a prolixity far superior to the level of acceptance of the system. There are therefore regulations that are carried out through the mechanism of natural selection (only the fittest ones survive and reproduce). Natural selection is the constraint imposed on living things. Chance does not produce information; it only produces complexity in the sense of Kolmogoroff [8]. Necessity is the one that produces information. In the nature, information is what has managed to evolve [4].

Complexity in a data lake can be illustrated by the addition of new data, the heterogeneity of the data but also by the product of data crossover (relationship search for example). The need is to retain the data that will provide useful information and eliminate data that will never be used. So we have to imagine a system that retains the potentially useful data.

To complicate the comparison, it must also be noted that nature sometimes keeps useless things (for example, there are DNA sequences that are never coded into proteins and therefore seem useless). So nature does not really delete all the data, it probably puts it to sleep. DNA that is not translated into protein can be altered (mutation) without, on the face of it, any consequence for the system (organism, ecosystem). The homeostasis of the system therefore depends on both a sorting by natural selection and a preservation of elements apparently not used.

This can be compared to data lakes (DNA being comparable to raw data) for which we thus claim for the necessity to delete data without eliminating everything.

3.4 Imitating Supply Chain Management

The supply chain is a set of integrated corporations and processes which is included sourcing the raw materials, manufacturing the products and transferring the finished products to the customers [11]. Supply chain management is a manner to coordinate and organize all processes and activities according to the supply chain goals [12]. Since data lakes collect all types of data and transform them for the final user, we can consider data as a product and data lakes and data governance as supply chain management.

There are some supply chain strategies that reduce or eliminate the useless, non-valuable or destructive products or activities, as for example lean management and green supply chain. Lean management is beneficial strategy that improves quality and profit by emphasizing on wast reduction. The main purpose of lean strategy is to wipe out all processes and products which do not create value for the chain [10].

In recent years, successful supply chains do not only try to gain more profit, improve product quality, increase service level and reduce final price regardless of ecological affairs, but also they attempt to pay more attention to the environmental consideration and issues which influence on economic and social systems [13], the firms and companies are pressured from ecologists and government to reduce or eliminate the pollutant processes and materials throughout the supply chain which have destructive effects on the environment.

Green supply chain management has become popular and strategic to improve environmental performance by green procurement and production throughout the entire Stages of a products life cycle. The green vision encompasses all decision-making process of the supply chain, as for example Eco-design, considering environmental criteria for supplier evaluation, green production, green transportation, green purchasing and revers logistics (disposal) [14].

There are a lot of standards and criteria to build a green supply chain. According to these standards, all of the products and activities that have ecological and environmental standards remain in chain otherwise they do not have permission to be present in supply chain.

We can define the proper criteria and standards for the products that are allowed to enter in supply chain or remain in it up to the last stage of the chain and if the products or activities do not have appropriate standards and requirements they will not be allowed to enter into the chain and if they were entered, they would be removed form it.

Data in data lakes act like products in supply chain. Therefore, if some data are poor and do not meet the criterion will not get into the data lakes, from our point of view, data in data lake could be never put into or deleted from data lake when they threat the veracity of data lake just like the harmful or non-valuable products in green supply chain.

4 Conclusion

Efficient data governance could be a practical solution to preserve data lakes from huge amount of useless data. Therefore, we can easily answer to the questions of this issue, some data can be retrained for a period or be removed form the lake if they increase the risk of data lake's authenticity.

In this work, we propose preliminary directions based on comparisons with supply chain management and natural lakes on the one hand, and on forgetting functions and decision support systems on the other hand. Criteria and solutions will be developed in our future work and assessed on real data lakes from both scientific and business data.

Acknowledgments. Supported by PHC CEDRE 42415YJ, French Ministry of European and Foreign Affairs (MEAE), French Ministry of Higher Education, Research and Innovation (MESRI) and Lebanese Ministry of Education and Higher Education (MEHE).

References

1. Boly, A., Hébrail, G.: Forgetting data intelligently in data warehouses. In: 2007 IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies, RIVF 2007, Hanoi, Vietnam, 5–9 March 2007, pp. 220–227. IEEE (2007). https://doi.org/10.1109/RIVF.2007.369160
2. Fang, H.: Managing data lakes in big data era: what's a data lake and why has it became popular in data management ecosystem. In: International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 820–824. IEEE (2015)
3. Gartner: Gartner Says Beware of the Data Lake Fallacy (2014). http://www.gartner.com/newsroom/id/2809117
4. Gaucherel, C., Gouyon, P., Dessalles, J.L.: Information. The Hidden Side of Life. ISTE (2019)
5. IBM: Governing and Managing Big Data for Analytics and Decision Makers (2014). http://www.redbooks.ibm.com/abstracts/redp5120.html?Open
6. Jones, N.: How to stop data centres from gobbling up the worlds electricity. Nature **561**, 163–166 (2018). https://doi.org/10.1038/d41586-018-06610-y
7. Khatri, V., Brown, C.V.: Designing data governance. Commun. ACM **53**(1), 148–152 (2010). https://doi.org/10.1145/1629175.1629210
8. Li, M., Vitnyi, P.M.: An Introduction to Kolmogorov Complexity and Its Applications, 3rd edn. Springer, New York (2008). https://doi.org/10.1007/978-0-387-49820-1
9. Madera, C., Laurent, A.: The next information architecture evolution: the data lake wave. In: Chbeir, R., Agrawal, R., Biskri, I. (eds.) Proceedings of the 8th International Conference on Management of Digital EcoSystems, MEDES 2016, Biarritz, France, 1–4 November 2016. pp. 174–180. ACM (2016). https://doi.org/10.1145/3012071
10. Martnez-Jurado, P.J., Moyano-Fuentes, J.: Lean management, supply chain management and sustainability: a literature review. J. Clean. Prod. **85**, 134–150 (2014). https://doi.org/10.1016/j.jclepro.2013.09.042
11. Mentzer, J., Witt, W.D., et al.: Defining supply chain (SC) management. J. Bus. Logist. **22**(2) (2001). https://doi.org/10.1002/j.2158-1592.2001.tb00001.x
12. Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E.: Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies. McGraw-Hill Publishing, New York (2003)

13. Zhu, Q., Sarkis, J.: Relationships between operational practices and performance among early adopters of green supply chain management practices in chinese manufacturing enterprises. J. Oper. Manag. **22**, 256–289 (2004)
14. Zhu, Q., Sarkis, J., Lai, K.: Green supply chain management implications for closing the loop. J. Transp. Res. Part E **44**, 1–18 (2008)

Mapping Heterogeneous Textual Data: A Multidimensional Approach Based on Spatiality and Theme

Jacques Fize[1,2](✉), Mathieu Roche[1,2], and Maguelonne Teisseire[2]

[1] CIRAD, UMR TETIS, 34398 Montpellier, France
{jacques.fize,mathieu.roche}@cirad.fr
[2] TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France
maguelonne.teisseire@irstea.fr

Abstract. In this paper, we propose a multidimensional mapping approach for heterogeneous textual data that exploits firstly the spatial dimension and secondly the thematic dimension. Based on the Spatial Textual Representation (STR) as well as the Geodict geographic database, the contribution presented in this paper integrates the thematic dimension of documents. To support our proposal on mapping textual documents, we evaluate the different aspects of the process using two real corpora, including one corpus that is highly heterogeneous.

Keywords: Text mining · Spatial and thematic dimensions · Heterogeneous data

1 Introduction

Over the last few decades, the improvement of data collection and storage techniques have raised new issues around the Big Data area. Big Data is characterized by three V's: Volume, Velocity and Variety. In this work, we focus on the heterogeneous dimension of data, the Variety, and more particularly on the mapping of textual data. The heterogeneity of textual data is characterized by the variety of structures (e.g. narrative, data table, etc.), formats (e.g. txt, xls, pdf, etc.), sources (newspapers, social networks, etc.), language or vocabulary used. To identify matching links in such data, new representation models and appropriate similarity measures are required. We thus propose a multidimensional mapping approach for heterogeneous textual data that exploits the spatial dimension and then the thematic dimension. For the spatial mapping of textual data, we have proposed a dedicated representation: the Spatial Textual Representation (STR) [6]. The STR is a graph representation that is composed of connected spatial entities (i.e. vertices) according to their spatial relationships (i.e. edges).

In this paper, we focus on the integration of the thematic dimension of documents and present a new contribution to enhance textual mapping documents

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 310–317, 2019.
https://doi.org/10.1007/978-3-030-34770-3_25

on the spatial dimension. Generally, the thematic dimension of a document is associated with a set of keywords (or phrases). These keywords can be obtained automatically or by using resources built by experts (ontologies, thesaurus, dictionaries). We propose to use different collections of keywords (terminologies) associated with the general themes of our corpora. In the context of the thematic mapping, we integrate terminologies from the study case domains and obtained by text-mining approaches. Finally, we propose a new transformation of the STR based on these terminologies to answer the following questions: What themes connect spatial entities? Can these new relations improve the mapping of the spatial dimension? To support our contributions, we evaluate the different aspects of the process using two corpora, including one highly heterogeneous corpus. The results obtained on these real data show that our proposals that integrate thematic information to the spatial dimension improve the task of linking heterogeneous data.

2 Thematic Relationships

2.1 Preliminaries

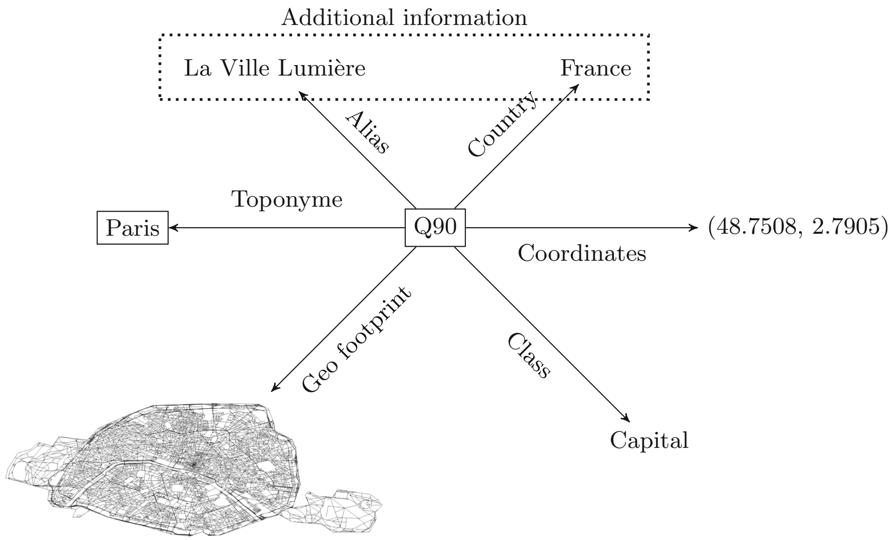
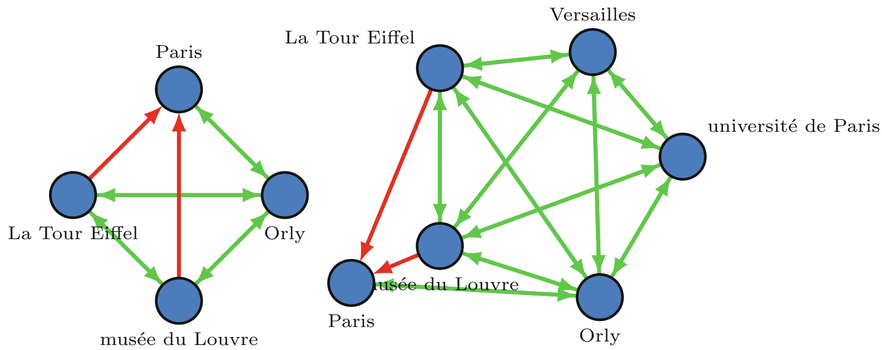


Fig. 1. Information related to the Paris spatial entity

To compare documents through the spatial dimension, a common representation is required. To this end, in previous work [6], we defined **S**patial **T**extual **R**epresentation (**STR**), a graph structure generated from spatial features identified in a document. A STR is composed of two components: **spatial entities** (i.e., vertices) and **spatial relations** (i.e., edges). A spatial entity is a named-entity located in a defined space [4], this associated with a set of 4 properties: a *toponym* (place name), a *geographic footprint* (coordinates), a *class* (e.g town, country) and *additional information* (e.g. alternates place name). An example of the available data is given in the Fig. 1. A spatial relation connects two spatial entities, *e.g., adjacency, inclusion, distance.* To generate a STR, a geographical knowledge database, also known as gazetteer, is required. Here, we propose GEODICT[1] [7], a gazetteer generated from three sources: *Wikidata*, *OpenStreetMap* and *Geonames*. In [6], we conducted experiments with graph matching algorithms [5] to match spatial features integrated into the STR. We designed two STR transformations: the Abstraction and the Extension. The abstraction increases the matching probability by replacing entities with their parent in the spatial hierarchy (Paris → France). To increase the matching probability, the extension adds new entities close to the existing ones. Figure 2 illustrates an STR and its associated extension (adjacency relations in green and inclusion relations in red).

The matching between corpus generated STR is performed using graph matching [10] algorithms. Different results based on the transformation of the STR and the graph matching algorithms are evaluated using a set of 4 criteria (See Sect. 2.4).



(a) A normal STR (b) Extended Version

Fig. 2. Example of a STR and its extension (Color figure online)

2.2 Thematic Entities

In order to improve the spatial matching process between two documents, we propose to integrate new information into the STR. The hypothesis assumes that linking thematic concepts to spatial entities will increase the number of relevant features in the representation for spatial matching. The main issue dwells with the automatic extraction of thematic entities related to the documents.

[1] Geodict is available at this address: http://dx.doi.org/10.18167/DVN1/MWQQOQ.

A thematic entity is associated with a phrase[2] as well as a set of variations used according to language or context (Table 1).

Table 1. The thematic entity: *land lot*

Language	Label	Alias
English	Land lot	Lot, plot tract of land, parcel of land, real property, real estate
French	Parcelle	

2.3 Extraction and Integration of Thematic Relations in the STR Structure

The thematic relation extraction process is context-sensitive. First, thematic entities need to be extracted. To identify thematic entities, we exploit terminology resources stored in specific dictionaries, thesauri and ontologies. Thus, we selected different terminologies for each specific corpus (See Sect. 3.1). Once the thematic entity extraction is performed, the next step is to connect identified thematic entities to the spatial entities.

In our work, a **thematic relation** exists between a spatial entity and a thematic entity if they belong to the same *context-window* in the document. In the following, we present two context-window

1. *Pattern based*

Example 1. [Determinant] [city or region or department] + [Determinant] + [**Spatial Entity**] + [Verb] + [Determinant] + [**Thematic Entity**]
⇒
"The city of **Montpellier** *retrieves* **organic waste** *to produce fertilizer [...]"*

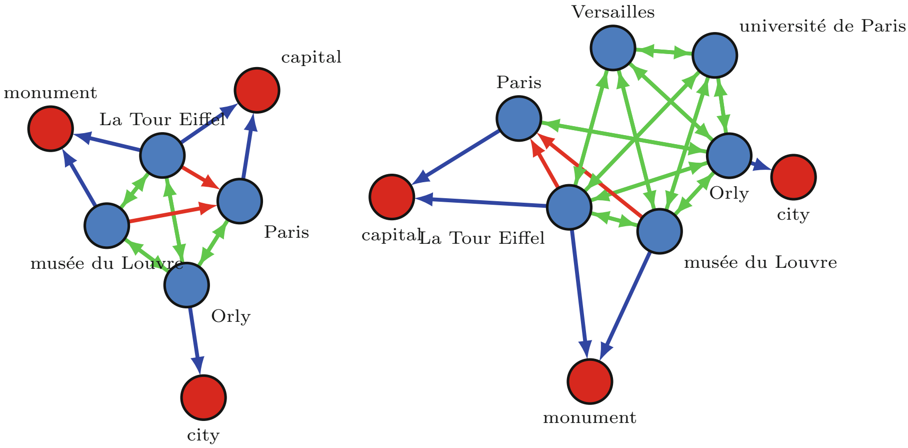
2. *Sliding window based.* This windows can be defined on a number of tokens or a specific structure (sentence, paragraph).

Example 2. The city of $\overbrace{\boxed{\textbf{Cerberus}\text{ helps }\textbf{farmers}}}^{\text{sliding window of size 5}}$ in their transition to organic agriculture.

Figure 3 illustrates the enrichment of the STRs presented Fig. 2 by adding the corresponding thematic nodes and links (thematic nodes in red and thematic links in blue).

If the corpus is highly heterogeneous, pattern-based extraction could be a real issue. For this reason, we chose to use the *sliding* window to identify relationships between the spatial entities and thematic entities.

[2] Group of morphemes or words that follow each other with a specific meaning.



(a) A STR with thematic information

(b) an extended STR with thematic information

Fig. 3. Example of STRs enriched with thematic nodes (in red) and thematic links (in blue) (Color figure online)

2.4 Matching Process: New Criteria

In [6], the matching quality was evaluated through four criteria:

- **Shared Spatial Entities (SSE).** The criterion is validated if the two STRs share at least one common spatial entity.
- **Close Spatial Entities (CSE)**. The criterion is validated if one or more spatial entities share proximities with a different entity in the other STR.
- **Significant Spatial Coverage (SSC).** Dense and significant groups of spatial entities located close to one another may be found in STR.
- **Strict Spatial Coverage (SCSC).** Distinct groups of spatial entities can be found in STRs.

In this paper, we add two new evaluation criteria based on the feedbacks of experts on the previous results given in [6]:

- **Average Distance (AM).** It is based on geographical distance average between the space entities belonging to the STRs. The distance is normalized between 0 and 1 then reversed (1 - average distance).
- **Percentage of shared spatial entities (PSSP).** To measure the similarity between the two sets of spatial entities of two corresponding STRs, we use the Srensen-Dice index (or Dice coefficient) defined as follows:

$$PSSP(STR_i, STR_k) = \frac{2 * |ES_{STR_i} \cap ES_{STR_k}|}{|ES_{STR_i}| + |ES_{STR_k}|} \tag{1}$$

with ES_{STR_i} the set of spatial entities of STR_i.

The evaluation process is then performed through 6 criteria. The best results are the ones maximizing the overall criteria, see Table 2.

3 Experiments

3.1 Datasets and Thematic Ressources

The proposed matching process aims to be generic, thus work on both heterogeneous and homogeneous data. Therefore, we select two corpora, one highly heterogeneous (AgroMada) and one homogeneous (PadiWeb). The first corpus, **PadiWeb**, corresponds to a set of press articles related to animal epidemiology. The second corpus, **AgroMada**, corresponds to a set of data produced during a project related to agroecology.

AgroMada. In the last decades, CIRAD has been involved in the development of sustainable agricultural practices in Madagascar. During this period, the CIRAD has produced a significant amount of data including theses, reports, technical manuals, field data, and presentations. Compared to Padi-Web, this corpus is highly heterogeneous. The original corpus is composed of 13 742 documents in different file formats. Based on this corpus, we selected documents corresponding to a specific thematic focus (i.e. agroecology) and a specific location (i.e. Madagascar). The final corpus, AgroMada, is composed of 5552 documents in English and French.

For this corpus, we have selected **4** vocabularies to evaluate the thematic entity extraction:

- INRA, Thesaurus formed from different dictionaries produced by INRA[3],
- Dev.Du., the vocabulary of sustainable development proposed by the French Ministry of Culture[4],
- Biotex BVLAC, vocabulary extracted using the Biotex software [8],
- Biotex + LDA BVLAC The Biotex software combined with LDA (Latent dirichlet allocation) [2].

Padi-Web. Padi-Web [1] is an epidemiology surveillance system implemented by CIRAD (Agricultural Research for Development) in collaboration with INRA (French National Institute for Agricultural Research). Padi-Web produces a classification and extracts information from unofficial sources (Google News) dealing with the emergence of epidemics to remedy delays in the publication of official decrees. To evaluate the volume and accuracy of the extracted information, a gold standard corpus, composed of 500 documents, was built.

Concerning this corpus, we selected **3** vocabularies and terminologie to evaluate the thematic entity extraction:

- BiotexPadi500 → Vocabulary extracted using the Biotex software on the PadiWeb corpus,
- BiotextPadi35k → Vocabulary extracted using Biotex software on a 35000 corpus document extracted using PadiWeb,

[3] https://dicoagroecologie.fr/en/.

[4] http://www.culture.gouv.fr/Thematiques/Langue-francaise-et-langues-de-France/Politiques-de-la-langue/Enrichissement-de-la-langue-francaise/FranceTerme/Vocabulaire-du-developpement-durable-2015.

– Disease Infect. → Terminologie providing a list of names of infectious animal diseases, as well as their variations.

3.2 Results

In [6], we propose a matching evaluation protocol based on different criteria. In this protocol, we evaluate different combinations of STR types (normal, extended, generalised) and different graph matching algorithms. In this paper, new experimentations are presented including the proposed extension of the STR based on thematic entities. We divide the different algorithms in two categories. The first category regroups the **structure-based** algorithms that only compare vertices, edges and their attributes. The second category regroups the **pattern-based** algorithms that compare high-level graph concept such as sequence, shortest-path, topology. As in [6], we selected a panel of graph matching algorithms. Concerning *structure-based* algorithms, the Vertex/Edge Overlap [9] (VEO), the Maximum Common Subgraph [3] (MCS) or an approximation of the Graph Edit Distance were selected. As for *pattern-based algorithms*, graph kernels, graph embedding and an Bag of Cliques (BOC) have been selected. To evaluate the potential of such combinations (STR and graph matching), we use different baseline approaches, such as *BOWSE* that computes the cosine similarity between each document vector, composed of the frequency of toponyms in the document.

The Table 2 shows the combinations of graph matching algorithm, type of STR that obtains the best performance. Different filters based on the STR size (i.e. number of spatial entities) are used to observe if different behaviours appears.

Table 2. Best combinations (STR type, graph-matching algorithm, Vocabulary) obtained on the different datasets, and with or without filters

Corpus	# Spatial Entities (E)	Graph matching algorithm	Type of STR	Terminology
AgroMada	$\|E\| > 0$	MCS	STR	–
	$\|E\| > 46$	BagOfCliques	STR	–
	$\|E\| > 72$	BOWSE	STR Ext. 1	Biotex + LDA BVLAC
PadiWeb	$\|E\| > 0$	MCS	STR	Disease Infect
	$\|E\| > 4$	MCS	STR	Disease Infect
	$\|E\| > 7$	MCS	STR	Biotex + Padi 35k

As we concluded in [6], structure-based method like *MCS* obtains higher performance in text mapping based on the spatia dimension. Integrating relations between spatial entities and thematic entities in the matching process allows to increase matching performances based on the defined criteria, especially in the PadiWeb corpus. As for vocabulary, vocabularies extracted from BIOTEX are useful on larger STR (i.e. number of spatial entities). Conversely domain-vocabulary (i.e. built by expert) such as DISEASE INFECT. obtains better results on STR highly and lowly spatialized (STR size).

4 Conclusion

In this paper, we propose a new matching approach for textual document based on the spatial and thematic dimensions. We explore in more details how thematic information could improve the performances. Results are interesting and show disparate impacts according to the context of the data used. Specific corpora seem to have more powerful effects when taking the two dimensions into account. The vocabularies used for thematic entity extraction play a crucial role that needs to be deepened in future works.

References

1. Arsevska, E., et al.: Monitoring disease outbreak events on the web using text-mining approach and domain expert knowledge. In: European Language Resources Association (ELRA), Paris, France, May 2016
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
3. Bunke, H., Allermann, G.: Inexact graph recognition matching for structural pattern. Pattern Recognit. Lett. **1**(May), 245–253 (1983). https://doi.org/10.1016/0167-8655(83)90033-8
4. Casati, R., Varzi, A.C.: Spatial entities. In: Stock, O. (ed.) Spatial and Temporal Reasoning, pp. 73–96. Springer, Dordrecht (1997). https://doi.org/10.1007/978-0-585-28322-7_3
5. Fischer, A., Riesen, K., Bunke, H.: Improved quadratic time approximation of graph edit distance by combining Hausdorff matching and greedy assignment. Pattern Recognit. Lett. **87**, 55–62 (2017). https://doi.org/10.1016/j.patrec.2016.06.014
6. Fize, J., Roche, M., Teisseire, M.: Matching heterogeneous textual data using spatial features. In: 2018 IEEE International Conference on Data Mining Workshops, ICDM Workshops, Singapore, Singapore, 17–20 November 2018, pp. 1389–1396 (2018). https://doi.org/10.1109/ICDMW.2018.00197
7. Fize, J., Shrivastava, G.: GeoDict: an integrated gazetteer. Association for Computational Linguistics (2017)
8. Lossio-Ventura, J.A., Jonquet, C., Roche, M., Teisseire, M.: Biomedical term extraction: overview and a new methodology. Inf. Retr. J. **19**(1–2), 59–99 (2016). https://doi.org/10.1007/s10791-015-9262-2
9. Papadimitriou, P., Dasdan, A., Garcia-Molina, H.: Web graph similarity for anomaly detection. J. Internet Serv. Appl. **1**(1), 19–30 (2010). https://doi.org/10.1007/s13174-010-0003-x
10. Riesen, K., Jiang, X., Bunke, H.: Exact and inexact graph matching: methodology and applications. In: Aggarwal, C.C., Wang, H. (eds.) Managing and Mining Graph Data, vol. 40, pp. 217–247. Springer, Boston (2010). https://doi.org/10.1007/978-1-4419-6045-0_7

Semantic Customers' Segmentation

Jocelyn Poncelet[1,2(✉)], Pierre-Antoine Jean[1(✉)], François Trousset[1(✉)], and Jacky Montmain[1(✉)]

[1] LGI2P - IMT Mines Als - Université de Montpellier, Alès, France
{jocelyn.poncelet,pierre-antoine.jean,franois.trousset,
jacky.montmain}@mines-ales.fr
[2] TRF Retail - 116 Alle Norbert Wiener, Nîmes, France
jocelyn.poncelet@trfretail.com

Abstract. Many approaches have been proposed to allow customers' segmentation in retail sector. However, very few contributions exploit the existing semantics links that may exist between objects and resulting groups. The aim of this paper is to overcome this drawback by using semantic similarity measures (SSM) in customers' segmentation to provide clusters based on product' topology instead of numerical indicators usually used (*i.e.* monetary indicators). More precisely, we intend to show the main advantage of SSM with a product taxonomy in the retail field. Usually, traditional approaches consider as similar three customers buying respectively apple, orange and beer. However, human intuition tends to group customers who buy orange and apple because both are fruits. Our approach is defined to identify this kind of grouping through SSM and abstract concepts belonging to product taxonomy. Experiments are conducted on real data from a French Retailer store and show the relevance of the proposed approach.

Keywords: Customers segmentation · Semantic clustering · Semantic similarity measures · Retail

1 Introduction

As a recurrent issue, outcomes of statistical analyses lack of semantics and their interpretability for decision-making purposes remains challenging. Data summarizing tools are required to provide more explicit aggregated or global indicators [1]. However, they do not improve the knowledge base (Customer Relationship Management - CRM) about customers of a given supermarket brand. As a result, they provide too much macroscopic insights which remains useless to determine information to advantage retailers in attracting and retaining customers. For a long time, retailers know the importance of data driven decision-making to capture customers behaviors and then use results to propose a customer driven approach in the retail industry. However, they focus on identifying good customers by mainly considering numerical values such as revenue, margin and the

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 318–325, 2019.
https://doi.org/10.1007/978-3-030-34770-3_26

frequency of customers [2,3,6,23]. Furthermore they do not consider the semantics that can be associated to products. Indeed, a customer that regularly buys fish and vegetable has obviously a closer behavior to a customer that usually buys meat and salad than to a customer that only buys sanitary products or seasonal items.

This type of inference clearly refers to approximate reasoning based on product similarity that symbolic artificial intelligence can automatically carry out. In this paper, we intend to introduce such reasoning on products purchased to provide semantic group of customers, i.e. groups derived from the hierarchical abstraction structure of products into classes (*e.g.*, *Greek yogurt* is a kind of *yogurt* which is in return a type of *dairy product*, etc.). Our objective is to identify customers' clusters depending on their conceptual purchasing behavior. Products could be organized within a taxonomic partial order defining an abstraction hierarchy. Products sold are the most specific classes of the partial order. We make the following hypothesis: the more specific products customers share, the closer they are. Assessment of the similarity of two customers is based on approximation over this product taxonomy (Sect. 3.1). Such reasoning clearly induces new semantic clustering techniques of customers based on the hierarchical taxonomy. Some approaches also consider a taxonomy [2,23], but do not consider nor semantic measures nor information content to define compared sets of items. Thereby, this study proposes a semantic clustering approaches where customers are considered as digraphs of product classes and then clustered. The rest of the paper is organized as follows. In Sect. 2 we address related works concerning customer segmentation. Section 3 introduces preliminary definitions to carry out our semantical clustering proposal.

2 Related Works on Customer Segmentation

In the retail world, the clustering of customers, usually called 'Customer Segmentation', consists in dividing heterogeneous customers groups based on common attributes. It requires to handle a large variety of customers [9,10]. Usually, the following kinds of data are considered: (a) Demographic data (e.g. gender, age, marital status); (b) Psychographic data (e.g. social class, lifestyle and personal characteristics); (c) Geographic data (e.g. area of residence or work); (d) Attitudinal Data (e.g. perceived data gathered from surveys); etc. Many different segmentation approaches are also applied. To name a few, RFM Recency, Frequency, and Monetary or CLV Customer Lifetime Value criteria are mostly used [3–5], for clustering [6], classification (e.g. neural-networks, decision trees) [4], models based on associations (association rules, Markov chain) [4], sequence discovery, forecasting [5]. Other researches focus on the mix of the products or product categories that customers have bought in their whole purchase history [11]. Even if our goal is also customer segmentation, the data used are the products bought by the clients and our approach is driven by the product classes.

3 Preliminary Definitions and Proposal

3.1 Hierarchical Abstraction and Taxonomies

Our main objective is to provide interpretable clusters of customers using similarity measure guided by the product organization. The similarity measure relies on the taxonomical structure of concepts. A taxonomical structure defines a partial order of the key concepts of a domain by generalizing and specializing relationships between concepts. Taxonomies give access to consensual abstraction of concepts with hierarchical relationships, *e.g.*, *Vegetables* defines a class or concept that includes *beans*, *leeks*, *carots* and so on, that are more specific concepts. Taxonomies are central components of a large variety of applications that rely on computer-processable domain expert knowledge, e.g. medical information and clinical decision support systems [8,12]. They are largely used in Artificial Intelligence systems, Information Retrieval, Computational Linguistics... [13]. In our case, customers clustering will be based on product taxonomy that defines a partially ordered set (poset) of products. An example of partial product taxonomy is shown in Fig. 1. In retail world, product taxonomy can be achieved by different means. Retailers or other experts can build this commodity structure. Most approaches usually introduce the Stock Keeping Unit (SKU) per item [14] or product categories (e.g. *Meat*, *Vegetables*, *Drinks*, etc.). Some researchers adopt the cross-category level indicated by domain experts and/or marketers [15].

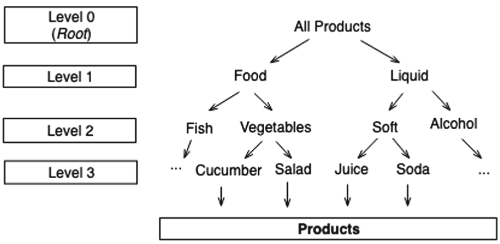


Fig. 1. Example of partial product taxonomy

3.2 Similarity and Informativeness Based on Taxonomy

Thanks to a product taxonomy, semantic similarity measures can be applied to define *similarity/dissimilarity* between customers. Those similarities compare sets of concepts associated to customers with *groupwise* measures [7]. These measures are themselves based on *pairwise* measures allowing the calculation of similarity between two concepts of the taxonomy [8]. Some of *pairwise* measures required Information Content (IC) associated to concepts, that is the amount of information associated to a concept (more a concept is specific, higher the information content of the concept is). There are several *groupwise* measures, in the same way there are several *pairwise* measures and several definitions of IC.

Groupwise measures allow comparison of sets of concepts related to objects (customers). There are two main categories: Direct and Indirect measure. Direct *groupwise* measures compare sets of concepts irrespective of their position in the taxonomy (*i.e.* Jaccard). Indirect *groupwise* measures aggregate similarity between concepts achieved by *pairwise* measures (*i.e.* BMA which stands for Best Match Average [16]). State-of-the-art approaches are divided in two kinds of *pairwise*: the first one based on the Information Content (IC) (*i.e.* Resnik *pairwise* measure [17]) and the other one based on the shortest path in the taxonomy (*i.e.* Wu & Palmer *pairwise* measure [18]). Finally, we discern two kinds of Information Content (IC): intrinsic and extrinsic. Intrinsic IC (*i.e.* Seco IC [19]) take only topological properties of the taxonomy into consideration while Intrinsic IC (*i.e.* Resnik IC [17]) used in addition frequency of concept in a observation bases (*i.e.* an ordinary corpus/data-set).

3.3 Methodology

Our goal is to identify sets of customers (*clusters*) with similar purchase behaviors. The main issue is to define appropriate similarity between customers. In practice, all customers' baskets are different whereas we intuitively make the difference between the two following customers: customer that mainly buys food product and customer who only purchases household laundry, cleaning products and textile.

Thanks to the use of the product taxonomy to compare customers, we obtain semantic clusters that are more interpretable by retailers. To this end, we used SSM with the Best Match Average [20] for the *groupwise* measure combined with the Resnik *pairwise* measure and Resnik IC [17]. This combination provides a matrix of *similarity/dissimilarity* between customers. Finally, this allows us to perform Hierarchical Clustering (HC) with Ward's method, which minimize distance inside each cluster (intra-clusters) while maximize distance between clusters (inter-clusters) [21]. The study tends to propose an approach of semantic clusters of customers close to the human intuition. In the following section, we present the application of semantic clustering on a real case. For more information about Semantic Similarity Measure, the reader is invited to refer to Harispe et al. studies ([8]).

4 Experiments and Results

Experiments have been conducted with real world data from one store, located in Paris (France) with 32 500 sales for 1 025 customers over a month. Table 1 describes some statistics about the dataset.

It contains data with different perspective: an overall vision, an average vision per customer and an average per sales receipt. For example, we can notice that the 1025 customers (fidelity cards) bought in average 25.36 different product categories (*Level 1 on the product taxonomy*). In other words, each customer probably purchase 25 different kinds of products. As explained above, to gather

Table 1. Some statistics about the dataset

	Number	Average per customer	Average per sale receipt
Number of customers	1025	-	-
Number of sale receipt	3692	5.75	-
Number of product bought	32552	48.54	8.63
Number of different categories	692	25.36	6.78
Number of different products	4979	31.70	7.47
Revenue	66590	102.45	18.22

customers, we used HAC method that cluster set of customers two by two, depending on their similarity. The key feature of the HAC is the identification of the ideal number of cluster. For the experiments, we used Ward's method besides a defined limit. Indeed, ideal clusters could be one customer per cluster but retailers try to identify a limited set of clusters. They try to figure it out "fictitious customers" that stand for as many customers as possible without loosing capital information. So, we varied the number of clusters from 2 to 25 and analyzed which number of clusters minimize the intra-class inertia and maximize the inter-class inertia. For the experiment, we obtained an optimal number of **4 clusters**, nominated respectively "C1" to "C4".

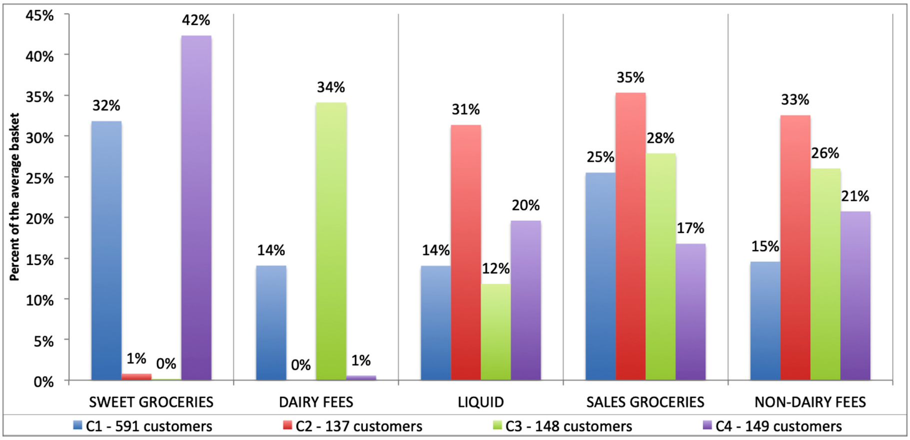


Fig. 2. Purchase frequency per product categories and clusters

The Fig. 2 represents the average frequency per product category level 1 (cf. Fig. 1) and per cluster. This information corresponds to the average basket of customers for each cluster. First of all, the analysis of the clusters' specificities underline that more than half of customers (591 customers) are gather in the cluster "C1". It seems that people from this "set of customers" came to purchase equally likely products from all different categories (*Level 1 on the product taxonomy*). The three others clusters "C2", "C3", "C4" have respectively 137, 148 and 149 customers which is approximately the same order of magnitude (≈15%). However, each of them has its own particularity. Customers from clusters "C2" mainly purchase products from *LIQUID*, *SALES GROCERIES* and *NON-DAIRY FEES*. This is the cluster that only purchase products from three categories (*Level 1 on the product taxonomy*). Clusters "C3" and "C4", for their part, are opposed by categories *SWEET GROCERIES* and *DAIRY FEES*. Indeed, customers from cluster "C3" will mainly purchase *DAIRY FEES* products while customers from cluster "C4" will mainly purchase *SWEET GROCERIES* products. We can notice that customers from those clusters ("C3" and "C4") will never purchase product from the discriminant category of the "opposite" cluster. Note that those customers, from clusters "C3" and "C4" shared their purchase between four categories Level 1 on the product taxonomy.

To make cluster more understandable by retailers, we give a label to each cluster. We used the most discriminant product categories to label clusters "C3" and "C4". Thereby, they have respectively the label "*DAIRY FEES* Customers" and "*SWEET GROCERIES* Customers". Specificities of Cluster "C2" came from the lack of purchase in previous category used. That's why we agreed to label it as "**N0**-*DAIRY FEES* & **N0**-*SWEET GROCERIES* Customers". Finally, cluster "C1" does not have any specific category. That's why we labeled it as "*ALL PRODUCTS* Customers".

5 Conclusion

The aim of the paper is to identify similar customers depending on their purchase behavior. Semantic clustering is based on the product taxonomy and should brings results more understandable for retailers. This approach allows retailers the identification of abstract purchase behavior (*i.e.*) thanks to the taxonomy. The final goal of customers' segmentation (or customer clustering) is to identify "good" customers based on retailers' preferences. With this semantic approaches results are based on their own business what underline the added value of the proposed approach.

Note that we used HAC to cluster customers, but other clustering methods and/or other configurations of Semantic Similarity Measures (IC, Pairwise, Groupwise) could be used. In this paper, our objective was to introduce semantic approach in retail. We believe that a comparison of clusters from different stores may allow improvement by defining generic "pattern" of customers. To go further, we can suppose that clusters' trajectory analyses could be done to identify changes in customers' behaviors. This will allow stores to be more preventive than reactive towards new trends (i.e. Vegan Customers). After all, if retailers validate resulting clusters, classification methods can be used to associate any new customers into thus clusters [22].

References

1. Berrah, L., Mauris, G., Montmain, J.: Monitoring the improvement of an overall industrial performance based on a Choquet integral aggregation. Int. J. Manag. Sci. OMEGA **36**, 340–351 (2008)
2. Griva, A., Bardaki, C., Pramatari, K., Papakiriakopoulos, D.: Retail business analytics: customer visit segmentation using market basket data. Expert Syst. Appl. **16**(1), 1–16 (2018)
3. Ching-Hsue, C., You-Shyang, C.: Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst. Appl. **36**(3), 4176–4184 (2009)
4. Chen, Y.L., Kuo, M.H., Wu, S.Y., Tang, K.: Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. Electron. Commer. Res. Appl. **8**(5), 241–251 (2009)
5. Khajvand, M., Zolfaghar, K., Ashoori, S., Alizadeh, S.: Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study. Procedia Comput. Sci. **3**, 57–63 (2011)
6. Lingras, P., Elagamy, A., Ammar, A., Elouedi, Z.: Iterative meta-clustering through granular hierarchy of supermarket customers and products. Inf. Sci. **257**, 14–31 (2014)
7. Harispe, S., Ranwez, S., Janaqi, S., Montmain, J.: Semantic similarity from natural language and ontology analysis. Synthesis Lectures on Human Language Technologies (2015)
8. Harispe, S., Snchez, D., Ranwez, S., Janaqi, S., Montmain, J.: A framework for unifying ontology-based semantic similarity. Study in the biomedical domain. J. Biomed. Inform. **48**, 38–53 (2014)
9. Hong, T., Kim, E.: Segmenting customers in online stores based on factors that affect the customers intention to purchase. Expert Syst. Appl. **39**(2), 2127–2131 (2012)
10. Aeron, H., Kumar, A., Moorthy, J.: Data mining framework for customer lifetime value-based segmentation. Expert Syst. Appl. **19**(1), 17–30 (2012)
11. Park, C.H., Park, Y.H., Schweidel, D.A.: A multi-category customer base analysis. Int. J. Res. Mark. **31**(3), 266–279 (2014)
12. Harispe, S., Ranwez, S., Janaqi, S., Montmain, J.: The Semantic Measures Library and Toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. Bioinformatics **30**, 740–742 (2014)
13. Harispe, S., Imoussaten, A., Trousset, F., Montmain, J.: On the consideration of a bring-to-mind model for computing the Information Content of concepts defined into ontologies. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8 (2015)
14. Kim, H.K., Kim, J.K., Chen, Q.Y.: A product network analysis for extending the market basket analysis. Expert Syst. Appl. **39**(8), 7403–7410 (2012)
15. lbadvi, A., Shahbazi, M.: A hybrid recommendation technique based on product category attributes. Expert Syst. Appl. **36**(9), 11480–11488 (2009)
16. Pesquita, C., Faria, D., Bastos, H., Falcao, A., Couto, F.: Evaluating go-based semantic similarity measures. In: Proceedings of 10th Annual Bio-Ontologies Meeting, vol. 37, p. 38 (2007)
17. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of IJCAI 1995, pp. 448–453 (1995)
18. Wu, Z., Palmer, M.: Verb semantics and lexical selection. In: 32nd Annual Meeting of the Association for Computational Linguistics, pp. 133–138 (1994)

19. Seco, N., Veale, T., Hayes, J.: An intrinsic information content metric for semantic similarity in WordNet. In: 16th European Conference on Artificial Intelligence, pp. 1–5 (2004)
20. Schlicker, A., Domingues, F.S., Rahnenfhrer, J., Lengauer, T.: A new measure for functional similarity of gene products based on Gene Ontology. BMC Bioinform. **7**, 302 (2006)
21. Murtagh, F.: Wards hierarchical agglomerative clustering method: which algorithms implement wards criterion? J. Classif. **31**, 274–295 (2014)
22. Silla Jr., C.N., Freitas, A.A.: A survey of hierarchical classification across different application domains. Data Min. Knowl. Discov. **22**, 31–72 (2011)
23. Srikant, R., Agrawal, R.: Mining generalized association rules. Future Gener. Comput. Syst. **13**, 161–180 (1997)

Analysis of Herd Behavior in Stock Prices Using Machine Learning

Gerard Rique(✉), Patrick Hosein, and Vaalmikki Arjoon

The University of the West Indies, St. Augustine, Trinidad and Tobago
{gerard.rique,patrick.hosein,vaalmikki.arjoon}@sta.uwi.edu

Abstract. In this paper, we consider the problem of herding behaviour in a Stock Exchange. Herding occurs when amateur investors follow the advice of financial gurus since they do not have the time, expertise or finances to do the research that is typically performed by these gurus. Although herding is well understood, many of the previous analyses have been through the use of statistical techniques. In this paper we have a second look using Machine Learning and demonstrate its effectiveness. We use a dataset obtained from the Singapore Stock exchange. Stocks were grouped into different portfolios based on the number of shares traded per day. Results from the algorithm show that herding is evident in each portfolio. We also find that herding is more pronounced among stocks that have higher volumes of shares traded.

Keywords: Herd behavior · Machine learning · Regression

1 Introduction

The theory of market efficiency states that a stock market is efficient if the prices reflect all available information at any given time and investors therefore have a rational expectation about the evolution of future prices. Behavioral finance has been proposed as the solution to anomalies in stock markets that could not be explained by traditional financial models. Behavioral finance suggests that investors are not rational. They are influenced by their own biases and make errors that can lead to wrong decisions and loss of investments. Investor herding is one such behavioral anomaly.

Herding can be defined as a rationale of less sophisticated investors, who attempt to mimic financial gurus or follow the activities of successful investors, since using their own information or knowledge would incur a higher cost [4]. Herding patterns in stock markets can cause prices to stray from their inherent values. Herding information can therefore be beneficial to policy makers and regulatory bodies as it can aid in identifying reasons for pricing inefficiencies and their role in market stability. Herding therefore challenges the validity of the theory of market efficiency as investors are not necessarily rational in their decisions and do not always evaluate future share prices by rational analysis of firms, but by observing and following the actions of other investors [9].

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 326–331, 2019.
https://doi.org/10.1007/978-3-030-34770-3_27

Herding can be a result of either rational or irrational decision making. The former can result from employees or agents of financial institutions that make decisions that can best protect their reputation as their performance is evaluated based on their success in purchasing. The latter on the other hand is the tendency of investors who irrationally ignore their own beliefs and sources of information and follow the actions of others even though they may not agree with it. They may do this to eliminate uncertainty of their own decisions and fulfills the need to feel confident [9].

2 Previous Work and Contributions

There are different factors that can occur within a stock market that can cause investors to follow the actions of others. Incomplete information disclosure in emerging markets and macroeconomic information can cause investors to herd [3]. A study conducted on a stock market in China reveal that herding is more prominent when the market is falling and trading volume is high. When examining stock markets in India, it is revealed that herding in this stock market is irrelevant of level of trading volume [9]. Herding may therefore be a result of different factors in different countries. Studies conducted have reported evidence of herding behaviour on the Singapore stock market [4]. Other studies conducted however, report otherwise [8]. Herding may manifest itself in return data and it can be detected by examining the relationship between equity return dispersion and the overall market return [5].

Earlier studies conducted on herding behaviour use Cross Sectional Standard Deviation (CSSD) as a measure of dispersion of returns of securities in a stock market. This measure however, does contain a significant drawback as it is sensitive to outliers [6]. More recent studies have adopted the use of Cross Sectional Absolute Deviation (CSAD) as a measure of return dispersion as it is less susceptible to outliers [3]. The International Monetary Fund (IMF) have conducted studies where they have compared machine learning methods to traditional stochastic methods of forecasting macroeconomic variables, such as gross domestic product (GDP), in different countries [7]. From this study, the authors have reported that machine learning algorithms were able to outperform benchmark forecasts by the IMF World Economic Outlook Forecasts. The authors also report that further research can be done into the application of machine learning to the field of economics. A study has also been conducted comparing the performance of different machine learning algorithms on time series forecasting [1]. The authors report that training the models can have a significant impact on performance. The authors also report that Multi-Layer Perceptrons and Gaussian Processes perform best in time series forecasting.

A number of studies have been done to analyse herding patterns in various stock markets around the world using various statistical methods. This paper contributes to the literature by analysing herding patterns using machine learning techniques. This paper also attempts to contribute to the literature of applying machine learning to the field of economics. To the best of our knowledge, machine learning has never been applied to the analysis of investor herding.

3 Herd Behavior Analysis

3.1 Description of Stock Market Data

A very rich data set containing stock prices on over six hundred stocks from 2005 to 2015 in Singapore was used for this study. The data set contains the opening and closing price as well as the volumes traded for each stock on given days. Table 1 shows a sample of the data set.

Table 1. Data set sample

Symbol	Date	Open	High	Low	Close	Volume
5EF	21-May-15	0.029	0.03	0.029	0.03	300100
5EF	22-May-15	0.03	0.03	0.03	0.03	199600
5EF	25-May-15	0.031	0.032	0.031	0.031	2476900
5EF	26-May-15	0.031	0.031	0.03	0.03	1205100
5EF	27-May-15	0.03	0.03	0.029	0.03	1910300

The opening price of any stock refers to the price at which a stock is first sold on the opening of the exchange on a trading day. Similarly, the closing price refers to the price at which any stock is sold on the closing of exchange on a trading day. The volume column indicates the number of shares of a particular stock that has been traded on a given trading day. A significant amount of information is available on this stock market due to global analyst coverage which, in itself, can influence investor behavior.

3.2 Herd Behavior Detection

We do not limit this study to herding on the aggregate market. We attempt to find herding by grouping stocks into different portfolios based on volume of shares traded per day. We do this because investors may tend to herd in defined market subsets. We use the Cross Sectional Absolute Deviation (CSAD) as a measure of dispersion of stock prices on any given trading day. If market participants follow aggregate market behaviour and ignore their own information during periods of large price movements, the relationship between CSAD and market equity would be negative [3]. This negative relationship is shown in the following regression:

$$\mathrm{CSAD}_t = \alpha + \gamma_1 R_{mt} + \gamma_2 R_{mt}^2 \tag{1}$$

where CSAD_t refers to the cross sectional absolute deviation of the stock prices on trading day t and R_{mt} refers to the market return on trading day t.

A negative γ_2 parameter indicates the presence of herding. We build on this theory by applying a machine learning algorithm to develop a model that can best represent this relationship. A variation of the linear regression machine learning algorithm using gradient descent was used to develop a model to determine the presence of herding on the Singapore Stock Market (SGX). The cross sectional absolute deviation specified by [3] is calculated as:

$$\mathrm{CSAD}_t = \frac{1}{n}\sum_{i=1}^{n}|R_{it} - R_{mt}| \tag{2}$$

where n represents the number of stocks in a given portfolio, R_{it} represents the return of stock i on training day t and R_{mt} refers to the market return on trading day t. The returns for a stock on any given trading day is given by

$$R_t = 100 \times [\log(P_t) - \log(P_{-1})] \tag{3}$$

where R_t refers to the return of any stock at time t and P_t represents the daily closing stock price.

3.3 The Regression Algorithm

A non-linear regression algorithm was used to detect herding in our data. This algorithm allowed us to fit a second order polynomial to the dataset and determine the relationship between CSAD and market equity. Gradient descent was used to iteratively reduce the error of the regression model until we find the global optimum (smallest error) [2]. The sum of the squared errors was used as the performance metric.

4 Numerical Results

The algorithm described above was applied to different volume stock portfolios in the SGX dataset. The regression algorithm was able to detect significant evidence of herding on the Singapore dataset. Figure 1 shows the relationship between cross sectional absolute deviation and market equity for the stock portfolio with the highest volumes traded.

After training the model on only the positive market equity values, the model was able to detect a nonlinear increasing relationship between the CSAD and the market equity. This regression model is given by:

$$\mathrm{CSAD}_t = 0.6068 + 0.9693R_{mt} - 0.1343R_{mt}^2 \tag{4}$$

The algorithm was trained on three different stock portfolios. Coefficients after training for the three different portfolio are given in Table 2.

Results in Table 2 show a negative γ_2 coefficient for all three stock portfolios indicating the presence of herding. This table also shows that portfolios with higher volumes of trading have γ_2 coefficients that are more statistically significant than those portfolios with lower trading volumes.

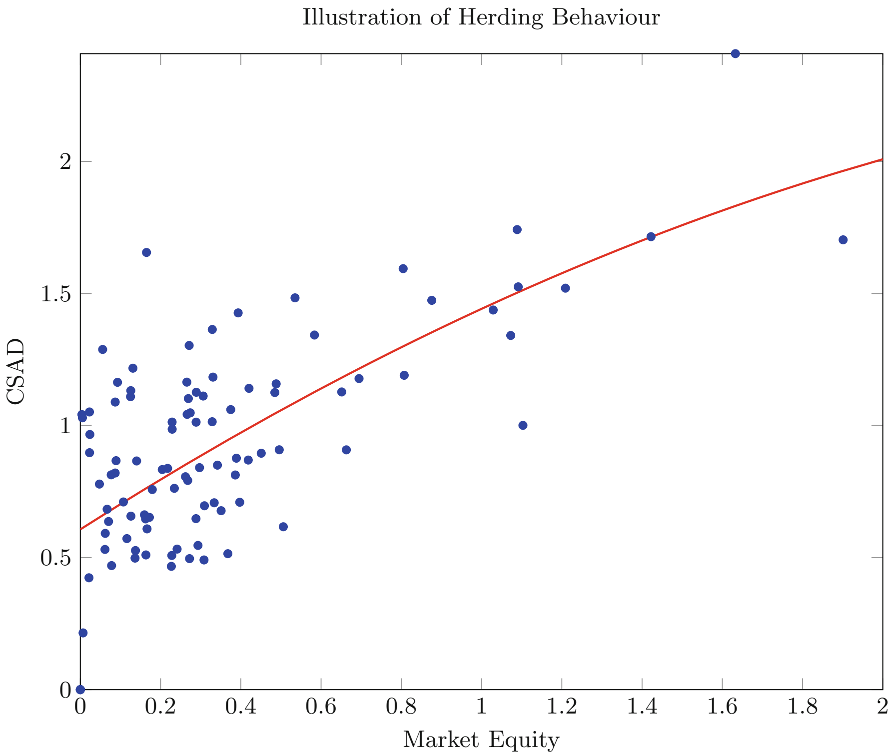


Fig. 1. Illustration of herding behaviour

Table 2. Results

Portfolio	α	γ_1	γ_2
1	0.3729	1.7897	−0.5477
2	0.6068	0.9693	−0.1342
3	0.0701	1.7329	−0.0845

5 Conclusion and Future Work

This paper investigates herding in the Singapore stock market using machine learning. The dataset containing stock prices for over 600 stocks was separated into three different stock portfolios based on the volume of shares traded per day. The algorithm was able to detect strong evidence of herding at the aggregate market level for all different stock portfolios. Results also show that the statistical significance of the R_{mt} coefficients are greater for stock portfolios with higher volumes traded. This may indicate that herding is more prevalent among stocks with greater volumes of trading. Therefore, we can conclude that investors tend to follow the behaviour of others in stocks with higher volumes of shares traded.

References

1. Ahmed, N.K., Atiya, A.F., Gayar, N.E., El-Shishiny, H.: An empirical comparison of machine learning models for time series forecasting. Econ. Rev. **29**(5–6), 594–621 (2010)
2. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Lechevallier, Y., Saporta, G. (eds.) COMPSTAT 2010, pp. 177–186. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-7908-2604-3_16
3. Chang, E.C., Cheng, J.W., Khorana, A.: An examination of herd behavior in equity markets: an international perspective. J. Bank. Financ. **24**(10), 1651–1679 (2000)
4. Chiang, T.C., Zheng, D.: An empirical analysis of herd behavior in global stock markets. J. Bank. Financ. **34**(8), 1911–1921 (2010)
5. Christie, W.G., Huang, R.D.: Following the pied piper: do individual returns herd around the market? Financ. Anal. J. **51**(4), 31–37 (1995)
6. Economou, F., Kostakis, A., Philippas, N.: Cross-country effects in herding behaviour: evidence from four South European markets. J. Int. Financ. Mark. Inst. Money **21**(3), 443–460 (2011)
7. Jung, J.K., Patnam, M., Ter-Martirosyan, A.: An Algorithmic Crystal Ball: Forecasts-based on Machine Learning. International Monetary Fund (2018)
8. Laih, Y.W., Liau, Y.S.: Herding behavior during the subprime mortgage crisis: evidence from six Asia-Pacific stock markets. Int. J. Econ. Financ. **5**(7), 71–84 (2013)
9. Lao, P., Singh, H.: Herding behaviour in the Chinese and Indian stock markets. J. Asian Econ. **22**(6), 495–506 (2011)

Health and Well-being

Personalized Music Experience for the Wellbeing of Elderly People

Stefania Bandini, Francesca Gasparini, and Marta Giltri(✉)

University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, Milano, MI, Italy
m.giltri@campus.unimib.it

Abstract. The ageing of our society, due to the technological advancements brought in our lives, its a sign of how better we can now live compared even just to a decade ago. Because of this, one of the important aspects that our society is now focused on is bringing proper care to all of the senior citizens that take advantage of these technological improvements.

Music has become one of the reference points in elderly care, especially for the benefits it has brought in the treatment of neurodegenerative diseases. This is why in this paper the idea of a cyber physical system involving music playing and user monitoring is presented, in order to propose a designed system that could also be useful in the above mentioned elderly care.

Keywords: Cyber physical system · Elderly care · Music therapy

1 Introduction

We are currently living in a more and more ageing society. In the last few years our life expectancy has increased thanks to the multiple technological advancements and improvements we now have available, and our society has become more invested in designing and developing programs addressing the proper care of senior citizens. These programs may have different aims, just as providing entertainment, encourage mobility or increase socialization, but the more widespread are the ones concerning health and wellbeing. Because of the importance it has in their life [6], music is one of the things that has often been involved in the last mentioned programs for elderly care, especially those linked to the treatment of different types of dementia and other neurodegenerative diseases [4].

Because of this, we would like to present a project concerning the usage of music for different applications, which include the ones described above. In fact, the aim of the following paper is to present the idea of a cyber physical system (CPS), physical istantiation of the concept of agent in the enactive paradigm of artificial intelligence, whose task is to monitor the physiological reaction of a certain subject and to combine it with external inputs in order to manage a playlist of audio played to the subject him/herself. The management of this

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 335–340, 2019.
https://doi.org/10.1007/978-3-030-34770-3_28

playlist should then influence the emotional state of the user, helping him/her maintain or change the way he/she's feeling.

The paper is organized as follows: Sect. 2 will introduce the state of the art regarding different approaches in elderly care done with the support of music and a panoramic look at what are the performed experiments on music and physiological reactions; Sect. 3 will present the design of the proposed CPS, explaining its functioning and what kind of work needs to be performed in order to lay the foundations for its development; Sect. 4 will then present a closer look at the necessary preliminary steps. Finally, in the conclusion, the contents of the paper are sum up and a final nod to the challenges and the use of the system is made.

2 State of the Art

As discussed in [6], music has a powerful and influential role in the life of the elderly. The work presented by Hays and Minichiello focuses on understanding what kind of impact and role music has, and the participants to their informal interview reported how this particular medium affected several aspects of their life: identity and understanding of the self, connection to oneself and to others, wellbeing and therapy, health, emotions, fantasy and motivations, beauty, spirituality... From the words of these older people we can see how music can let us reflect upon ourselves and influence the way we interact with others, uplift our spirit and help during illness and treatment, elicit emotions and so on, in more ways than one depending on the person asked.

But the great impact of music is also proven when used as a practical therapeutic tool, not only for its positive effect on the psyche and on the spirit of the patients: in fact, music therapy is now a feasible non-pharmacological treatment to be used in providing temporary relief from the symptoms of neurodegenerative diseases. One of the most researched applications for music therapy regard the treatment of Alzheimer's Disease [5], where different studies assessed how patients subject to this kind of therapy showed improvement in their cognitive functions, memory and self consciousness, while their anxiety, depression and agitated behaviour seemed to diminish. Other than that, studies on the treatment of Huntington's Disease [2] and Parkinson's Disease [9] have been conducted, identifying the same beneficial impact brought by music to the patients' conditions and quality of life.

Turning now to the topic of audio listening and emotion recognition, there is no shortage of experiments that tried to tackle the particular problem of understanding what a subject is feeling while subjected to different stimuli. In fact, as showed in [7], there have been multiple experiments concerning emotion elicitation and recognition through the observance of physiological signals recorded during the sessions. The experiments have been various, with differences ranging from the stimuli used during the sessions to the sensors used for the signals recording, to the emotion classification model used and to the number of subject recorded.

Nevertheless, this approach of using physiological signals for emotion recognition has potential, and experiments using audios and music are already been performed [8], proving that this is a course worth exploring and studying.

3 The Cyber Physical System

As was said before, the aim of this paper is to present the design of a CPS whose task is to generate a playlist of audios able to maintain or influence the mood of the user by gaining informations from his/her physiological response and by interpreting external inputs which may give further data to take into account. This external input may include information about the time of the day, the activity the user is performing, the emotional state the user wants to reach and so on, giving context to which the playlist must adapt just as much as the physiological signals observed from the user.

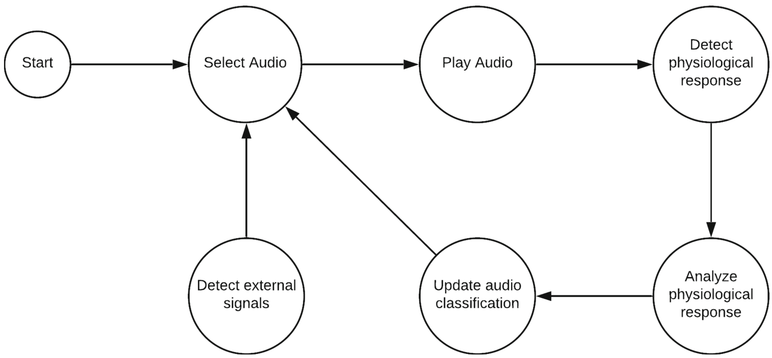


Fig. 1. Workflow cycle of the proposed CPS.

Figure 1 introduces what the CPS workflow looks like, and the process described in the diagram can be explained by the following sequence:

1. Select an audio for the playlist at random from a previously classified dataset to initiate the process
2. Play the audio to the user
3. Detect the physiological response of the subject and, subsequently, analyze it to gain information about his/her current emotional state
4. Update the audio emotional classification by using the current state of the user himself/herself if the user's response contradicts the emotion the audio was supposed to elicit
5. Detect eventual external inputs that could modify the system current course of action
6. Select a new audio for the playlist from the dataset and repeat the process.

This system should be able to work with every kind of audio, ranging from classical music to rock and to nature sounds, and should support enough interaction with the user to be able to obtain input commands to modify its strategy of playlist management.

The core of this CPS resides in the selection process, where the system has to incorporate reasoning computational models (rule based or ontology-based) in order to make the correct audio choice: taking into account the available audio tracks, the physiological signals coming from the user and the external inputs that need to influence the user's state, the CPS should be able to select the best audio required in the situation presented by processing all the information provided by the various entry channels described before. Therefore, the system should learn how to behave (select the audio) by looking at the environment (the physiological and external data) and acting on it (influencing the user), following a mode of action similar to the one implemented by enactive agents [3].

But, as the workflow above may suggest, this CPS should be able to do at least a couple of tasks more besides creating and managing the audio playlist. In fact, the system relies on two principal functions to ensure its proper operation and enhance the user experience:

1. Classification of new audios in order to consider them correctly when evaluating their inclusion in the playlist
2. Correlation of the physiological response of the user to an emotional label to correctly maintain or modify the system course of action

These two tasks fall under the *profiling* and *personalized experience* that this system should offer: because of different personal experiences, tastes and physical functions, everyone could approach the usage of this system in a different way, feeding it with unique data, and the CPS should be able to tailor its work to the final user's personal characteristics.

Because the experience customization is one of the principal characteristics of our system, we need to develop this aspect correctly by firstly executing some initial work aimed at gaining information and experience about this particular profiling task. Formally, we need to investigate (i) the emotional value that different audios can have, in order to understand if we can deduce what kind of emotion an audio can elicit starting from its objective features, and (ii) the relationship between the mood of the user and his physiological response, to better understand how the body of the subjects react when they are feeling a certain way.

4 Initial Work

Following the difficulties in the system working highlighted in the previous section, we identified two different preparatory tasks that have to be performed: the creation of a "ground truth" for the audio dataset and the interpretation of the physiological response of the subjects.

For the first task we could use an already labeled dataset like the one produced in [1] to train a classifier that could then be used to analyze and label new audios with the probability values regarding a given set of emotions, creating a classification based on objective audio features. The purpose of this training should be identifying in different audio tracks some objective features on which to base the emotional classification we need for our system. This task could be performed by using a cellular neural network (CNN) or a similar classification method. Figure 2 shows how this classification step should work.

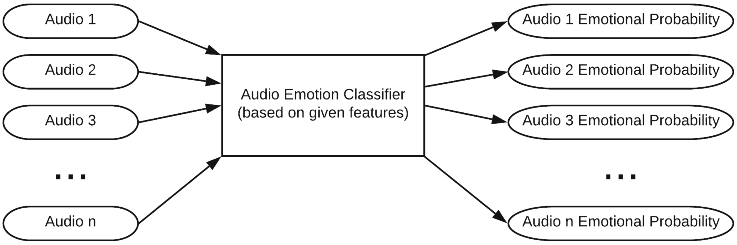


Fig. 2. Schema about audio classification.

For the second one, we can identify different states and link them to a subjective response from the person we are observing by correlating data from different sensors, acquired through an ad-hoc experiment. The purpose of this experiment should be becoming able to identify what kind of values we can expect from different physiological signals when observing a particular emotional state of a certain subject. This kind of work can be done by setting up an experiment structured as to follow part of the CPS functioning, in particular the steps from 1 to 3 presented in Sect. 3. Moreover, this experiment could be designed to already use as input certain songs or musics familiar to the elderly subjects in order to tune their response on audios which could be used in different therapy programs and, for example, in experiments like the ones highlighted in [4].

Figure 3 shows a diagram depicting the concept for the experiment.

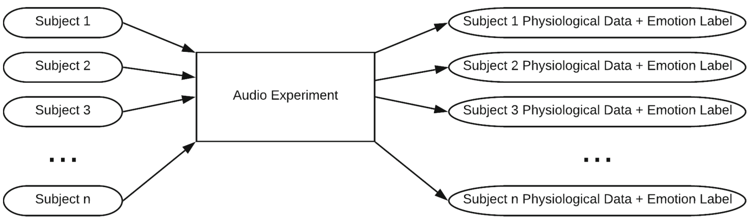


Fig. 3. Schema about subject response acquisition and classification.

5 Conclusion

In this paper, the design of a cyber physical system able to auto-adjust a music playlist considering external input and physiological response from its user is presented, and then two different preparatory tasks for its development are introduced.

Of course the tasks described above address just the foundation of the work that has to be done in order to render this CPS operational, since they are only concerned about emotion recognition and audio classification. But, as we said before, in order to properly create this system we need to fully understand how to tackle the problem of experience customization and profiling: being able to successfully understand how a user is feeling and how music could make him/her react is a fundamental step towards the creation of a functioning system. After this, the work regarding every other step in the workflow can be started and linked to the rest.

If effectively designed, this system could be used as a powerful tool in the treatment and medical area, especially after seeing how music can become a powerful support to elderly treatment and well-being, by providing further insight on the mental and emotional state of senior citizens.

Acknowledgement. This research is supported by the FONDAZIONE CARIPLO "LONGEVICITY-Social Inclusion for a Elderly through Walkability" (Rif. 2017-0938)

References

1. Aljanaki, A., Wiering, F., Veltkamp, R.C.: Studying emotion induced by music through a crowdsourcing game. Inf. Process. Manag. **52**(1), 115–128 (2016)
2. van Bruggen-Rufi, M., Roos, R.: The effect of music therapy for patients with Huntington's disease: a systematic literature review. J. Lit. Art Stud. **5**(1), 30–40 (2015)
3. De Loor, P., Manac'h, K., Tisseau, J.: Enaction-based artificial intelligence: toward co-evolution with humans in the loop. Minds Mach. **19**(3), 319–343 (2009)
4. Fang, R., Ye, S., Huangfu, J., Calimag, D.P.: Music therapy is a potential intervention for cognition of Alzheimer's disease: a mini-review. Transl. Neurodegener. **6**(1), 2 (2017)
5. Guess, H.: Alzheimer's disease and the impact of music therapy a systematic literature review (2017)
6. Hays, T., Minichiello, V.: The meaning of music in the lives of older people: a qualitative study. Psychol. Music **33**(4), 437–451 (2005)
7. Jerritta, S., Murugappan, M., Nagarajan, R., Wan, K.: Physiological signals based human emotion recognition: a review. In: 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, pp. 410–415. IEEE (2011)
8. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. **30**(12), 2067–2083 (2008)
9. Raglio, A.: Music therapy interventions in Parkinson's disease: the state-of-the-art. Front. Neurol. **6**, 185 (2015)

Cognitive and Physiological Response for Health Monitoring in an Ageing Population: A Multi-modal System

Aurora Saibene and Francesca Gasparini(✉)

Multi Media Signal Processing Laboratory, Department of Computer Sciences, Systems and Communications, University of Milano - Bicocca, Milan, Italy
a.saibene2@campus.unimib.it, francesca.gasparini@unimib.it
http://www.mmsp.unimib.it/

Abstract. In the last years, a rapid increase of life expectancy has been observed and thus ageing is becoming a hot topic in different research fields. Ensuring the wellbeing of elderly people has a positive impact on healthcare, economy and social equality and thus developing a system that takes into account and integrates physiological parameters, cognitive and physical training, elderly-user experience and perception of the system itself, may represent an innovative multi-modal mean through which ameliorate older people wellbeing.

Keywords: Ageing · Health monitoring · Multi-modal signal processing

1 Introduction

The population is ageing. Life expectancy is rapidly increasing as depicted by the World Population Prospects 2019 (https://population.un.org/wpp/) maps. Making a comparison between the projections in 2005–2010 (Fig. 1) and in 2020–2025 (Fig. 2), we can notice that the life expectancy at birth in the sole Europe will increase by 5 to 10 years.

Therefore, ageing is becoming a challenging topic, bearing an impact on healthcare, economy, social and environmental interactions.

Being able to ensure the wellbeing of elderly people requires an easier and more accessible healthcare monitoring, which should take into account the issues arising from the ageing process itself: difficulties in moving in a not-always elderly friendly environment, deterioration of physical and cognitive conditions, possible lack of social interactions maybe due to the aforementioned impairments.

Knowing all this, we suggest a system for ageing process monitoring decoupled from a clinical environment and thus serviceable in a personal space, focusing principally on the sensing devices and their data processing.

In the following we describe the system components and how they exploit low-cost physiological devices, signal processing techniques and networking to enable a multi-modal analysis for:

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 341–347, 2019.
https://doi.org/10.1007/978-3-030-34770-3_29

- constant health monitoring and data access by a physician (PU);
- elderly-user (EU) self evaluation thanks to health goals daily completion;
- training suggestion based on daily acquired data;
- environment parameters calibration, being the user provided with an in-home domotic system or some smart devices, like virtual assistants.

Although we are aware that a system like the one proposed requires high accessibility, dependability and security, taking into account the necessity of managing a great amount of personal data, we will not discuss in the present work how to solve the problems arising from the listed topics and we redirect you to the outstanding review by Dey et al. [4], which outlines similar critical issues in the Cyber Physical System domain while addressing some of them.

However, we highlight the importance of mental and physical training for a healthy ageing, considering the necessity of having elderly-user-friendly training applications.

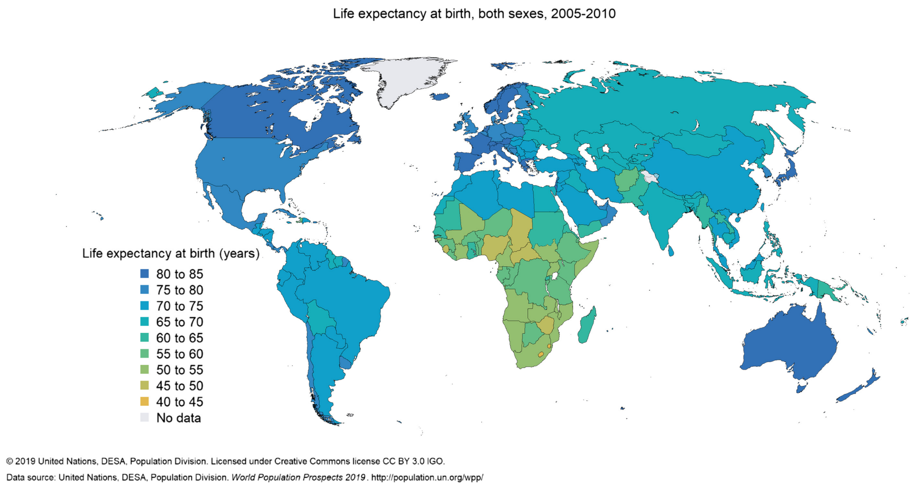


Fig. 1. Life expectancy at birth, both sexes, 2005–2010. Image obtained from https://population.un.org/wpp/. The color scale begins from a yellow color corresponding to age 40–45 and increases to a blue color representing an age range between 80 and 85. (Color figure online)

2 Background

To allow a better understanding of the related work, we give here an intuition of the proposed system.

Suppose that an elderly person wants to go to the local street market (in the following we call it LSM scenario). An immediate necessity is to know the weather forecast; also, it could be useful to obtain the time table of the available public transportation. Will these conditions influence the decision of going to the street market? Will they be cause of stress?

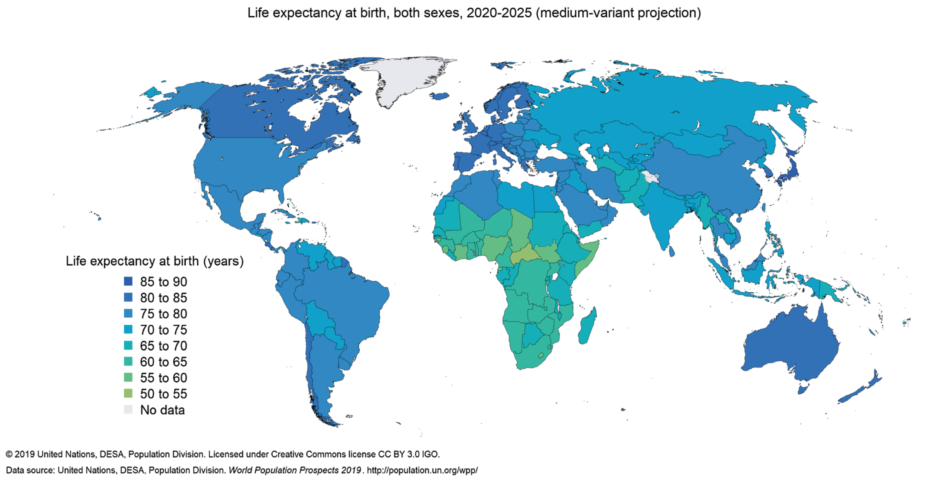


Fig. 2. Life expectancy at birth, both sexes, 2020–2025. Image obtained from https://population.un.org/wpp/. The color scale begins from a greenish color corresponding to age 50–55 and increases to a blue color representing an age range between 85 and 90. (Color figure online)

A way to predict and/or access these data is to have a system that combines different type of sensing devices and that takes into account the level of confidence that an elderly-user has in managing technology.

In particular, the possible devices that could help this type of analysis are:

- a video camera to monitor the EU behaviour while using his/her smart device to access the required information;
- some physiological sensors that could be intertwined to a shirt and/or part of a smart watch to allow a non-invasive recording of data like the galvanic response, known to be an easier stress indicator [6];
- a weather station to access the possible environmental influence for decision making.

The LSM scenario let us understand that the wellbeing of an elderly person is a summation of his/her physical and mental states, bounded to his/her everyday life and the environment he/she lives in.

Thus, the idea of having a system that allows health monitoring outside medical centers is justified.

3 Related Work

Nowadays, sensor technology has improved significantly and allows to measure physiological signals remotely as well as to register daily life activities (e.g. inertial sensors, video and audio recordings). Several sensors can be easily integrated into smartphones or wearable devices [8], making them more comfortable and usable even for elderly people or patients affected by differently impairing diseases. Moreover, according to de Veer et al. [3], even though the study was representative only of the Dutch population, a good number of older people use mobile devices and the 63.1% is open-minded towards the use of interactive applications for e-Health, in particular as healthcare provided through Internet.

Another interesting datum emerges from the aforementioned work: of the favorable sample, the 37.1% has stated that, even though the use of interactive applications would be an aid to maintain a certain independence and easily contact professionals, their experience in using them is not pleasant.

Thus, another important factor to take into consideration is the necessity of having elderly-user-friendly and centered applications to encourage older people interaction with the new media. In fact, numerous conditions may influence the choice of using or not an application. For the acceptance of personal health devices Sun et al. [7] identified factors that span from attitude towards technology to perceived usefulness.

In any case, e-Health, here generically intended as the remote healthcare based on information technology, became popular as a significant support to traditional healthcare assistance.

Several e-Healthcare Systems have been proposed in recent years with different purposes [5]. Firstly, most of them are devoted to constantly monitor physiological parameters of people affected by several types of chronic diseases such as obesity and diabetes, or degenerative ones, such as dementia and Alzheimer. Secondly, to reduce the cost of healthcare systems, a great effort has also been dedicated in developing systems able to make an early detection and diagnose peculiar diseases. In particular, several attempts have been proposed for monitoring elderly people daily life activities and vital signs, producing alarms or emergency help requests for relatives or physicians.

These systems are mainly focused on physiological parameters and their monitoring, but in the work of Bamidis et al. [1] the importance of preserving and enhancing brain health has been identified as a critical factor for elderly people wellbeing and associated to cognitive and physical training. This training could represent a mean through which delay the first cognitive declines, like memory loss, and enhance brain plasticity.

In fact, there is evidence that at least physical training with targeted exercises is an ally to maintain brain health during ageing, especially if the training is tailored to the person who has to perform it [2].

In the following, we describe the proposed system, which wants to exploit the discussed technologies, systems and knowledge, to have an integrated and tailored designed elderly-friendly health monitoring application.

4 System Proposal

We propose a system composed by three main modules, thought to allow a bidirectional interaction between the elderly-user and the system: subject-based sensing (SBS), environment-based sensing (EBS) and signal processing (SP) modules.

Before describing them in detail, we point out that a proper smart device is provided to the elderly-user to ensure the interaction with the system. The device allows:

- the collection of the answers given by the EU for a questionnaire defined by the physician, being a mean through which access the general wellbeing of the EU;
- the access to the training application for health goals completion (physical and cognitive exercises), that will change on a daily basis considering the feedback given by the modules to the system;
- the visualisation of the EU general wellbeing.

The PU has also access to the EU information through a web application and he/she will receive alerts in case of peculiar physiological recordings to consider a possible intervention.

Having these premises in mind, we now describe the modules that influence the whole system operation.

The SBS module is deputed to behavioural and physiological data recording. We define *behavioural sensing* a set of sensors that (1) allows the monitoring of the elderly-user interaction with the environment and the given smart devices and (2) acquires the answers to the questionnaire given from the physician to the EU. Thus, the SBS behavioural sensing considers data that could be uncontrolled or controlled by the EU.

In the first case the system will use in-home video cameras for non-verbal data acquisition and the EU access history to the dedicated application in a given smart device. The questionnaire filling is defined in the controlled condition case due to the necessity of having a precise subjective response.

For what concerns the SBS *physiological sensing*, the EU is provided with a set of physiological monitoring devices that works in an uncontrolled task-free or uncontrolled task-dependent condition. The smart shirt and watch represent the mean through which record galvanic, pulse, pressure and temperature signals in uncontrolled situations, while the low-cost wireless electroencephalographic device has to be worn in combination with the aforementioned tools only during the questionnaire filling (specific task) for cognitive workload assessment.

In the presented LSM scenario the SBS behavioural sensing is used in an uncontrolled condition: being interested in a spontaneous response to the environment and the smart devices usage and thus on video footage and dedicated application history. Also the SBS physiological sensing records the physiological data in an uncontrolled task-free condition, having to monitor an every day life situation.

Another important module is the environmental-based sensing one that works in an (1) uncontrolled condition, whenever the information about day, weather, location data are required for environmental influence assessment, and (2) in a controlled condition when the elderly-user has the necessity to decide the customisation of his/her smart devices, e.g. lights, roll-up shutters in case of a in-home domotic system or ask for a certain type of music to the virtual assistant.

In the LSM scenario the EBS module could be used in both its conditions: the uncontrolled information may present an alert about weather conditions, being for example a time range in which the sun is too strong or the risk of incurring in a downpour; the controlled condition could be exploited by the EU to roll-down the shutters while he/she stays out and automatically roll them up when he/she returns to his/her home.

The key components of the described modules are presented in Table 1 and in italics are highlighted the sensing and conditions used in the LSM scenario.

Lastly, we present the signal processing module that is deputed to managing the data acquired by the SBS and EBS modules and combine them in a multi-modal approach.

Having data that are time-varying, behavioural (discrete and categorical) and environmental (mainly scalar), proper algorithm are defined to analyse them both at a low-level (data pre-processing) and at a high-level (classification).

Table 1. System modules detailed description

Module	Sensing	Condition	Sensors and signals
SBS	*Behavioural*	*Uncontrolled*	Video monitoring for non-verbal data
		Controlled	Questionnaire
	Physiological	*Uncontrolled task-free*	Physiological (galvanic, electrocardiogram, body temperature and pressure) devices for health monitoring
		Uncontrolled task-dependent	Wireless electroencephalographic device for cognitive workload assessment and physiological devices for stress assessment during questionnaire filling
EBS	*Environmental*	*Uncontrolled*	Day, weather, location data for environmental influence assessment
		Controlled	Domotic system for parameters customisation

5 Discussion

New technologies and advances in the study of ageing effects are driving the interest of developing systems that could enhance the wellbeing of elderly people.

Moving from this principle, we propose a system that has the elderly-user as its main actor. A bidirectional interaction allows the EU to have tailored physical and cognitive exercises to promote an engaging experience more than an enforced training menu.

Even though the health monitoring is intended as both the training data analysis and the physiological parameter acquisition, the physician intervention is only required in case of great risk of danger for the EU health and for the definition of the questionnaire. Thus, the system is not intended for the diagnose and constant monitoring of diseased people, but as a tool to ameliorate elderly people wellbeing.

The new aspect of our proposal is to exploit different sensors, the already developed ideas about e-Healthcare systems, the studies about elderly people perception of technology to create a multi-modal and interactive analysis of cognitive and physical elderly people conditions while enhancing their user experience.

References

1. Bamidis, P., et al.: A review of physical and cognitive interventions in aging. Neurosci. Biobehav. Rev. **44**, 206–220 (2014)
2. Barha, C.K., Galea, L.A., Nagamatsu, L.S., Erickson, K.I., Liu-Ambrose, T.: Personalising exercise recommendations for brain health: considerations and future directions. Br. J. Sports Med. **51**(8), 636–639 (2017)
3. Cimperman, M., Brenčič, M.M., Trkman, P.: Analyzing older users' home telehealth services acceptance behavior—applying an extended UTAUT model. Int. J. Med. Inform. **90**, 22–31 (2016)
4. Dey, N., Ashour, A.S., Shi, F., Fong, S.J., Tavares, J.M.R.: Medical cyber-physical systems: a survey. J. Med. Syst. **42**(4), 74 (2018)
5. Ghamari, M., Janko, B., Sherratt, R., Harwin, W., Piechockic, R., Soltanpur, C.: A survey on wireless body area networks for ehealthcare systems in residential environments. Sensors **16**(6), 831 (2016)
6. Sharma, M., Kacker, S., Sharma, M.: A brief introduction and review on galvanic skin response. Int. J. Med. Res. Prof. **2**, 13–17 (2016)
7. Sun, N., Rau, P.L.P.: The acceptance of personal health devices among patients with chronic conditions. Int. J. Med. Inform. **84**(4), 288–297 (2015)
8. Yetisen, A.K., Martinez-Hurtado, J.L., Ünal, B., Khademhosseini, A., Butt, H.: Wearables in medicine. Adv. Mater. **30**(33), 1706910 (2018)

Nudges Driven Networks: Towards More Acceptable Recommendations for Inducing Targeted Social Communities

Italo Zoppis[1(✉)], Sara Manzoni[1], Giancarlo Mauri[1], Giada Pietrabissa[2], Andrea Trentini[3], Daniela Micucci[1], and Gianluca Castelnuovo[2]

[1] Department of Computer Science, Università degli Studi di Milano-Bicocca, Milan, Italy
italo.zoppis@unimib.it
[2] Department of Psychology, Catholic University of Milan, Milan, Italy
[3] Department of Computer Science, Univeristà degli Studi di Milano, Milan, Italy

Abstract. Mind Cognitive Impairment is one of the most common clinical manifestations affecting the elderly. In this paper, we report the work in progress (in the frame of our SENIOR project) to provide elderly with new Nudge theory driven advices for influencing their interest to a conscious and functional participation to "targeted" social communities where suggestions on the overall wellness can be shared, recognized as usefull by users and supported by health care providers.

Keywords: Nudge theory · Optimization problems · Community identification · Heuristic algorithms · SENIOR project

1 Introduction

Mind Cognitive Impairment (MCI) is rapidly becoming one of the most common clinical manifestations affecting the elderly. It is characterized by deterioration of memory and cognitive function that is beyond what is expected based on age and educational level. MCI does not interfere significantly with individuals' daily activities. It can act as a transitional level of evolving dementia with a range of conversion of 10%–15% per year [18]. Therefore, it is crucial to protect older people against MCI. The SENIOR (SystEm of Nudge theory based ICT applications for OldeR citizens) project will create an advanced coach ICT system, collecting and integrating significant physiological and behavioral data for interacting with elderly people and to provide them personalized suggestions about preventive measures, social participation and overall wellness according to the Nudge theory approach [15]. In this respect, the SENIOR virtual coach system will be a not invasive but disappearing technology and will assist senior citizens in a ergonomic manner to preserve healthy conditions and social well-being as long as possible, also in relationship with caregivers.

The personalized virtual coach is currently planned with two major components: a wearable device/app for the patients and a cloud back-end server for the

S. El Yacoubi et al. (Eds.): INSCI 2019, LNCS 11938, pp. 348–355, 2019.
https://doi.org/10.1007/978-3-030-34770-3_30

health professionals. The patient daily behavior is collected by the app, seamlessly, without user interaction; the app sends data continuously to the remote back-end. Finally, the retrieved data patterns are processed with machine learning techniques to produce the suggested actions to patients. The challenge of SENIOR project will be to move many elderly patients towards a conscious and functional use of new technologies by exploiting the advantages of being connected, not only for being clinically monitored, but for keeping on contacts with a targeted audience of friends and (clinically) profiled patients able to share homogeneous experiences, objectives and interests. This perspective follows an interesting research area where networks (i.e., conceptually, graphs) are taken as models for explainable recommendations [11,12,22]. Following these ideas, we will take advantage of specific user-user relationships to model the situation where patients could be interested to meet or contact some other users, for sharing information even with available health providers. In fact, most real domains are best described by structured data where instances of particular types are related to each other in complex ways (i.e., graphs). In these contexts, relationships (e.g. expressing similarities or differences between instances) can provide suitable source of information to inference mechanisms in data mining and machine learning problems. Such an approach has been proved useful for both classification and clustering in many situations, see e.g. [5,6] for molecular data analysis, [14] for network data, or [16] for a broader view of fundamentals in cluster analysis.

In our case, the SENIOR's social engine will operate to identify sub-groups of homogeneous users (e.g., friends or clinically profiled patients) and at the same time to support and stimulate them to a functional participation through a corresponding Nudge-based "activity". It is clear that, a proper handling of procedures and data is fundamental in order to convert available information into useful formulation that leads to particular (induced) communities. In fact, to this respect, the SENIOR engine is designed as a specific Nudge-based Recommender Systems (RS) able to leverage social public information to influence elderly for sharing similar experiences and interests, or promoting discussions about received treatments even to obtain more targeted suggestions from specialists and doctors.

More specifically, from a computational point of view the SENIOR's RS for its social engine functionality can be formally expressed as the problem of finding the largest cohesive sub-groups of users within an input network (e.g., [7,9]). Therefore, here we approach this issue from a theoretical perspective, first by formulating the corresponding computational problem, then by reporting both an heuristic algorithm for the considered problem and preliminary numerical results for a set of simulated data.

Please notice that the intent here is to introduce the fundamental concepts of the SENIOR's social "engine" (Sect. 2), in particular by detailing the optimization modeling as discussed in Sect. 2.2. Perspectives and future development of this work will be reported in Sect. 5.

2 Targeted and Optimized Communication

2.1 Nudge Driven Approach

The “Nudges” could be defined as the attempts at influencing people’s judgment, choice or behavior in a predictable way, that is made possible because of cognitive boundaris, bias, and habit in individual and social decision making, posing barriers for people to perform rationally in their own self-declared interest, and which works by making use of those boundaries, routines and habits as integral parts of such attempts.

The application of the Nudge theory have gained wide spread traction and have been disseminated into many different sectors of public policy including public health. Nevertheless, the systematic use of “Nudges” to induce behaviour change such as making helpful choices easier has yet to develop.

In this context, SENIOR will develop Nudge advises for the promotion of an active healty behaviour in older people with MCI. In particular, SENIOR will cope with an active social (clinical oriented) participation by designing a specific Nudge-based Recommender System (RS) able to leverage social (public) information to influence elderly for sharing similar experiences and interests, or promoting targeted discussions about treatments and even to obtain suggestions from specialists and doctors.

2.2 Optimized Communication

Although Nudges can effectively be applied to virtually any type of human relationships where the alteration of people’s thinking and decision-making may be beneficial for those people, and to wider society as a whole, much more work remains to be carried out for sharing “an optimized” information content in a social space.

In this regard, the SENIOR’s social engine is formulated as an optimization problem in such a way that the largest number of “similar” users can be identified, and at the same time supported by the corresponding Nudge advises to participate consciously and effectively in the identified cohesive social community. In other terms, the “best” social space (i.e., the largest community of elderly with MCI) will be suitably and continuously induced and updated by both Nudge-based activities and optimization procedures aiming to identify subgroups of homogeneous people who can share their interests, being characterized for example by similar clinical profiles or friendships.

2.3 Problem Formulation

Finding a network’s cohesive sub-graph is a well-known problem that has been applied in several contexts with different objectves and perspectives [1,20,21]. Computationally, graphs are abstract objects which represents relationships between elements, V (vertices), through a set of edges, E. A typical approach to compute dense sub-graphs is the identification of cliques. We recall that a graph is a clique if it is a complete graph, that is a graph where each pair of vertices is connected by an edge. Moreover, a graph is an independent set if no pair of vertices is connected by an edge. Therefore, given a graph $G = (V, E)$ and a set V' of vertices, with $V' \subseteq V$, we denote by $G[V']$ the graph induced by V', that is $G[V'] = (V', E')$, where

$$E' = \{\{v_i, v_j\} : v_i, v_j \in V' \wedge \{v_i, v_j\} \in E\}$$

Proceeding this modeling, our goal can be then formulated by identifying, within the observed population, the largest clique with particular properties. In more general terms, in order to adapt our formulation to a set of individuals that could be either cases (e.g., patients affected by some disease) or sampled controls, we furtherly require for our modeling to separate as much as possible healthy subjects from the case group. In our context, a clique might be ideally used to provide a set of homogeneous patients (i.e., clinical profiled elderly with similar interests or experiences). Similarly, an independent set of vertices, well separated from that clique, could ideally be used to instantiate other control subjects. Finally, before providing a formal description for our maximization objective we can give the following definitions.

Definition 1. *Given a graph $G = (V, E)$, G is a* Sparse Split Graph *if V can be partitioned in two sets V_1 and V_2 such that $G[V_1]$ is a clique, $G[V_2]$ is an independent set and, for each $v_i \in V_1$ and $v_j \in V_2$, $\{v_i, v_j\} \notin E$.*

Definition 2. *Given a graph $G = (V, E)$ and a graph $G' = (V, E')$, we define the* agreement *of G and G', denoted by $A(G, G')$, as follows:*

$$A(G, G') = |E \cap E'| + |((V \times V) \setminus E) \cap ((V \times V) \setminus E')|$$

Informally, $A(G, G')$ represents the number of edges that belong to both E and E' plus the number of pairs of vertices in V that are not connected by an edge both in G and in G'.

Now, we are ready to give the following definition that could be used to optimize our objective.

Problem 1. Maximum Sparse Split Graph Problem ($\mathcal{MAX} - \mathcal{SSG}$)
Input: a graph $G = (V, E)$.
Output: a graph $G' = (V, E')$, such that G' is a sparse split graph.
Objective function: $A(G, G')$ is maximized.

The goal of our approach is then to obtain a suitable partitions of the given input network in two sets, a clique (elderly) and an independent set (controls), such that no edge connecting vertices of the clique with vertices of the independent set are allowed.

3 A Genetic Algorithm for the Community Detection

$\mathcal{MAX} - \mathcal{SSG}$ is the complement of the Minimum Sparse Split Graph Problem [13], where given a graph, it is asked to partition the graph in two sets, a clique and an independent set, such that there is no edge connecting a vertex of the clique with a vertex of the independent set, and the number of modified relations (edges removed or inserted with respect to the input graph) is minimized. Since Minimum Sparse Split Graph is NP-hard [13], the same complexity holds also for the Maximum Sparse Split Graph problem, thus making optimization potentially impracticable for large input instances. We present here an approach based on genetic algorithms [17] to seek faster approximation solutions[1].

Given an input graph $G = (V, E)$, the proposed GA represents a solution, i.e., a partition $V = \{V_1, V_2\}$, in such a way that V_1 is a clique and V_2 is an independent set, as a binary chromosome c of size $n = |V|$, and $v_i \in V_1$ if the ith-vertex is part of a clique; 0 otherwise. Note that, with a slight abuse of notation, we will denote by $G[c]$ the sub-graph of G induced by the representation of chromosome c. Similarly, $V[c]$ and $E[c]$ will denote the set of vertices (V) and edges of $G[c] = G[V]$.

During the offspring generation, chromosomes are interpreted as hypotheses of feasible solutions, undergoing to mutation, crossover and selection. The mutation process applied here aims to modify parsimoniously the set that a vertex belongs to: if a vertex belongs to an independent set then it is moved to the clique; if it belongs to the clique, then it is moved to an independent set. More specifically:

- A random sample of vertices with low degree in the original graph takes value 0 over the chromosomes (i.e., these vertices are considered as part of the independent set).
- A random sample of vertices from cliques of size three of the original graph takes value 1 over the chromosomes (i.e., these vertices are considered as part of the clique as they can be computed efficiently - in time $O(n^3)$, and they represent possible subsets of larger cliques.).
- Mutation: Each individual from current population is modified with probability 0.1.
- Crossover is applied with probability 0.8.
- Elitist selection (or elitism) when constructing a new population from an existing one is applied to allow the best individuals in the current population to be part of the next generation. It is known that such an approach can be used to guarantee that the solution quality does not decrease from one generation to the next [2].
- Finally, the fitness value for chromosome c is defined as the negative (maximization problem) of the Hamming distance between the adjacency matrix of the current evaluated graph $G[c]$ and the adjacency matrix of the input graph.

[1] See, for example, [10] for further details.

Table 1. Models and parameters. ER: Erdos-Renyi model; BA: preferential attachment (*Barabasi*). GA parameters: population size, generation, elitism, crossover probability and mutation probability.

Type	Pop. Size	Generat.	Elitism	Cross Pr	Mut Pr
$ER(10, 1/2)$	100	17.7	10	0.8	0.1
$ER(50, 1/2)$	500	64.3	50	0.8	0.1
$ER(100, 1/2)$	1000	106.6	100	0.8	0.1
$ER(200, 1/2)$	2000	176.2	200	0.8	0.1
$BA(10)$	100	15.1	10	0.8	0.1
$BA(50)$	500	41.1	50	0.8	0.1
$BA(100)$	1000	53.9	100	0.8	0.1
$BA(200)$	2000	46.4	200	0.8	0.1

Table 2. Models and results. ER: Erdos-Renyi model; BA: Preferential attachment (*Barabasi*). Results: Fitness value, CPU time: user, system, elapsed (in seconds) and Relative mutation rate, percentage of mutations (edge addition and edge deletions) in a graph among all the potential mutations in the original graph.

Type	Final	CPU			Relative Mut. rate
	Fitness Value	User	Sys	Elaps.	
$ER(10, 1/2)$	−13.0	1.777	0.00	1.778	28.9
$ER(50, 1/2)$	−508.0	29.918	0.470	30.442	41.6
$ER(100, 1/2)$	−2196.1	153.196	0.360	153.727	44.7
$ER(200, 1/2)$	−9182.5	1365.656	62.717	5277.984	46.4
$BA(10)$	−8.0	1.512	0.000	1.517	17.8
$BA(50)$	−54.0	25.207	0.040	26.378	4.1
$BA(100)$	−130.1	73.594	0.112	272.099	2.5
$BA(200)$	−372.0	231.403	6.769	239.079	2.2

4 Simulation and Numerical Results

The genetic algorithm described in the previous paragraph was coded in R using the Genetic Algorithm package [19]. We run the process using synthetic data by sampling graphs from two typical random models (Erdos-Renyi and Barabasi). *Erdos-Renyi* model, $ER(n, p)$, [4] is used to generate graphs with n vertices with constant edge probability p. *Barabasi*, $BA(n)$, uses a simple random mechanism to generate graphs with n vertices through *preferential attachment* [3]. For each model, we sampled 10 graphs using different values of n in $\{10, 50, 100, 200\}$. We increased the size of the population accordingly (see Table 1). The results are given in Table 2. We stop the execution of GA when the fitness value does not increase after 15 generations. As shown in the same table, the final fitness value (as expected), decreases as the number of vertices increases. Moreover, we observe that for ER model this value is smaller than BA. Accordingly, the required CPU time for convergences is greater for graphs sampled from ER than for graphs from BA. Our results show that the genetic algorithm has good performances on simulated data. Even for graph with 200 of vertices. In this case the algorithm was able to converge in a reasonable time (approximately 1365 s for samples from ER and approximately 231 s for samples from BA). In the analysis we consider the *relative mutation rate*, that is the percentage of the number of modifications (edge additions and edge deletions) in a graph among all the potential mutations in the original graph. Interestingly, this value is significantly lower in graphs from BA than in graphs from ER. Moreover, in graphs from ER this value increases as the number of vertices increases, while in graphs from BA we observe the opposite behavior. This fact suggests that graphs from BA have a structure that is closer to that of a Sparse Split Graph with respect to the graphs from ER; furthermore, this property becomes stronger as the number of vertices increases.

5 Conclusions

SENIOR's social ambition is the creation of targeted Nudge driven communities of elderly with MCI where social information is leveraged to foster engagement and partnership with consumers, to change their behaviors and to fight against unhealthy life styles.

We reported how an optimization procedure based on a soft computing method can be applied to get homogeneous communities of empowered patients supported by specialists. In this respect, a new Nudge based inductive engine is fundamental, not only to vehicolate social information but in order for elderly citzens to easily recognize suggestions as correct and useful. In this paper, we considered this issue from a computational point of view: first by defining a variant of a well known computational (optimization) problem. Then by applying (due to the intrinsic computational hardness of the formulated question) a GA-based heuristic for obtaining approximated solutions (i.e., see also [7,8]). Our interest for future implementation of this research will be focused on the validation of the SENIOR's social engine using real and more complex structured data.

References

1. Bader, G.D., Hogue, C.W.V.: An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. **4**, 2 (2003)
2. Baluja, S., Caruana, R.: Removing the genetics from the standard genetic algorithm. In: Machine Learning, Proceedings of the 12th International Conference on Machine Learning, Tahoe City, California, USA, pp. 38–46 (1995)
3. Barabási, A.-L., Albert, R.: GA: a package for genetic algorithms in R. Science **286**, 509–512 (1999)
4. Bollobas, B.: Random Graphs. Cambridge University Press, Cambridge (2001)

5. Cava, C., Zoppis, I., Gariboldi, M., Castiglioni, I., Mauri, G., Antoniotti, M.: Copy-number alterations for tumor progression inference, pp. 104–109 (2013)
6. Cava, C., Zoppis, I., Gariboldi, M., Castiglioni, I., Mauri, G., Antoniotti, M.: Combined analysis of chromosomal instabilities and gene expression for colon cancer progression inference. J. Clin. Bioinform. **4**, 2 (2014)
7. Dondi, R., Mauri, G., Zoppis, I.: Clique editing to support case versus control discrimination. In: Czarnowski, I., Caballero, A.M., Howlett, R.J., Jain, L.C. (eds.) Intelligent Decision Technologies 2016. SIST, vol. 56, pp. 27–36. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39630-9_3
8. Dondi, R., Mauri, G., Zoppis, I.: Orthology correction for gene tree reconstruction: theoretical and experimental results. Procedia Comput. Sci. **108**, 1115–1124 (2017)
9. Dondi, R., Mauri, G., Zoppis, I.: On the tractability of finding disjoint clubs in a network. Theor. Comput. Sci. **777**, 243–251 (2019)
10. Eiben, A.E., Smith, J.E., et al.: Introduction to Evolutionary Computing, vol. 53. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-662-05094-1
11. He, X., Chen, T., Kan, M.-Y., Chen, X.: TriRank: review-aware explainable recommendation by modeling aspects. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1661–1670. ACM (2015)
12. Heckel, R., Vlachos, M., Parnell, T., Dünner, C.: Scalable and interpretable product recommendations via overlapping co-clustering. In: IEEE 33rd International Conference on Data Engineering (ICDE), pp. 1033–1044. IEEE (2017)
13. Hüffner, F., Komusiewicz, C., Nichterlein, A.: Editing graphs into few cliques: complexity, approximation, and kernelization schemes. In: Dehne, F., Sack, J.-R., Stege, U. (eds.) WADS 2015. LNCS, vol. 9214, pp. 410–421. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21840-3_34
14. Kolaczyk, E.D.: Statistical Analysis of Network Data: Methods and Models. Springer, New York (2009). https://doi.org/10.1007/978-0-387-88146-1
15. Leonard, T.C., Thaler,R.H., Sunstein, C.R.: Nudge: improving decisions about health, wealth, and happiness (2008)
16. Long, B., Zhang, Z., Yu, P.S.: Relational Data Clustering: Models, Algorithms, and Applications. Chapman & Hall/CRC, Boca Raton (2010)
17. Mitchell, M.: An Introduction to Genetic Algorithms. Complex Adaptive Systems. MIT press, Cambridge (1996)
18. Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. **56**(3), 303–308 (1999)
19. Scrucca, L.: GA: a package for genetic algorithms in R. J. Stat. Softw. **53**(4), 1–37 (2013)
20. Sharan, R., Shamir, R.: Center CLICK: a clustering algorithm with applications to gene expression analysis. In: Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology, La Jolla/San Diego, CA, USA, 19–23 August 2000, pp. 307–316 (2000)
21. Spirin, V., Mirny, L.A.: Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci. **100**, 12123–12128 (2003)
22. Wang, X., He, X., Feng, F., Nie, L., Chua, T.-S.: TEM: tree-enhanced embedding model for explainable recommendation. In: Proceedings of the 2018 WWW Conference on WWW, pp. 1543–1552. International WWW Conference Steering Committee (2018)

Author Index