On the Quality of a Social Simulation Model: A Lifecycle Framework*

Claudio Cioffi-Revilla

Center for Social Complexity and Dept. Computational Social Science
Krasnow Institute for Advanced Study, George Mason University
Fairfax, Virginia 22030 (Washington DC) USA
ccioffi@gmu.edu

Abstract. Computational social science grows from several research traditions with roots in The Enlightenment and earlier origins in Aristotle's comparative analysis of social systems. Extant standards of scientific quality and excellence have been inherited through the history and philosophy of science in terms of basic principles, such as formalization, testing, replication, and dissemination. More specifically, the properties of Truth, Beauty, and Justice proposed by C.A. Lave and J.G. March for mathematical social science are equally valid criteria for assessing quality in social simulation models. Helpful as such classic standards of quality may be, social computing adds new scientific features (complex systems, object-oriented simulations, network models, emergent dynamics) that require development as additional standards for judging quality. Social simulation models in particular (e.g., agent-based modeling) contribute further specific requirements for assessing quality. This paper proposes and discusses a set of dimensions for discerning quality in social simulations, especially agent-based models, beyond the traditional standards of verification and validation.

Keywords: Quality standards, evaluation criteria, social simulations, agent-based models, comparative analysis, computational methodology, Simon's paradigm.

1 Introduction: Motivation and Background

The field of social simulation in general, and agent-based modeling in particular, has begun to generate methodological proposals for assessing and promoting quality across diverse and related areas (Gilbert and Troitzsch, 2005; Taber and

* Prepared for the 9th Conference of the European Social Simulation Association, Warsaw School of Economics, Warsaw, Poland, September 16–20,2013. This paper was inspired by the First Workshop on Quality Commons, Maison de la Recherche, Paris, 28–29 January, 2010. Thanks to Petra Ahrweiler, Edmund Chattoe-Brown, Bruce Edmonds, Corinna Elsenbroich, Nigel Gilbert, David Hales, Dirk Helbing, Andreij Nowak, and Paul Ormerod for stimulating discussions at the Center for Mathematics and Analysis in the Social Sciences, University of Paris Sorbonne.

B. Kamiński and G.Koloch (eds.), *Advances in Social Simulation*,
Advances in Intelligent Systems and Computing 229,
DOI: 10.1007/978-3-642-39829-2_2,

Timpone, 1996).[1] For instance, proposals exist in the area of communicating social simulation models (Cioffi and Rouleau, 2010; Grimm et al., 2005), comparing models (Rouchier et al., 2008; Cioffi, 2011; Cioffi and Gotts, 2003), and assessing complex projects that involve large interdisciplinary teams (Cioffi, 2010). Consensus on quality standards in social simulation has not yet emerged, but a promising discussion by practitioners is already under way.

The properties of "Truth," "Beauty," and "Justice" have been proposed by Charles A. Lave and James G. March (1993) and are widely used for discerning quality in *social science mathematical models.* The three terms "Truth," "Beauty," and "Justice" (or "TBJ," for short) are labels for quality dimensions referring to fundamentally good—i.e., normatively desirable—features of social science modeling. Accordingly, the TBJ terms must be interpreted as labels, and not literally (Lave and March, 1993: Chapter 3).

Truth refers to the empirical explanatory content of a model—its contribution to improving causal understanding of social phenomena—in the sense of developing positive theory. For example, truth is normally judged by internal and external validation procedures, corresponding to axiomatic coherence and empirical veracity, respectively (Kaplan, 1964; Sargeant, 2004). Truthfulness is the main classical criterion for evaluating empirical science (Hempel, 1965; Cover and Curd, 1998; Meeker, 2002), whether the model is statistical, mathematical, or computational. "Truth" must be a constituent feature in a social science model, such that without it a model has no overall quality contribution.

Beauty refers to the esthetic quality of a model, to its elegance in terms of properties such as parsimony, formal elegance, syntactical structure, and similar stylistic features. Beauty is about art and form. For example, the mathematical beauty of some equations falls within this criterion, including features such as the style of a well-annotated system of equations where notation is clear, well-defined, and elegant. Unlike truth, beauty is not necessarily a constituent attribute, but is certainly a desirable scientific quality.

Justice refers to the extent to which a model contributes to a better world—to improvement in the quality of life, the betterment of the human condition, or the mitigation of unfairness. Justice is a normative criterion, unlike the other two that are positive and esthetic. For example, a model may improve our understanding of human conflict, inequality, refugee flows, or miscommunication, thereby helping to mitigate or improve social relations and well-being through conflict resolution, poverty reduction, humanitarian assistance, or improved cross-cultural communication, respectively. Policy analysis can be supported by modeling.

These Lave-March criteria of truth, beauty, and justice are useful for evaluating the quality of social simulation models. For example, in the classic Schelling (1971) model of segregation all three criteria are well-recognized. This is a

[1] This paper focuses on social simulations, so the broader field of computational social science (e.g., social data algorithms or socioinformatics, complexity models, social networks, social GIS, and related areas of social computing) lies beyond the scope of this paper. Quality research in those other areas is subject to its own standards.

fundamental reason why Schelling's model is so highly appreciated. Other examples that satisfy the Lave-March TBJ criteria might also include the Sugarscape model (Epstein and Axtell, 1996), the Iruba model (Doran, 2005), and Pick-a-Number (Hoffmann, 2002, 2005).

However, a further challenge exists because social simulations have features that render truth, beauty, and justice insufficient as criteria for assessing quality. This is because social simulation models are instantiated or rendered in code (a computer program in some language), so one can easily imagine a social simulation that would be of high quality in terms of truth, beauty, and justice, but fail in overall quality because simulation models pose additional challenges beyond other social science models (i.e., beyond the features of statistical or mathematical models).

As illustrated in Figure 1 (UML class diagram), social simulations have properties that are shared with all models in science generally and social science in particular, based on inheritance as a specialized class, in addition to having other features of their own. For example, the specific programming language of an agent-based model (Java, C++, or other) would be a defining feature.

The inheritance relation between social science models and social simulations readily suggests the specific features that distinguish the latter from the former, as illustrated in Table 1.

Additional criteria for social simulations—i.e., criteria beyond classical standards for social social models—should allow us to judge quality in terms of "The Good, The Bad, and The Ugly."

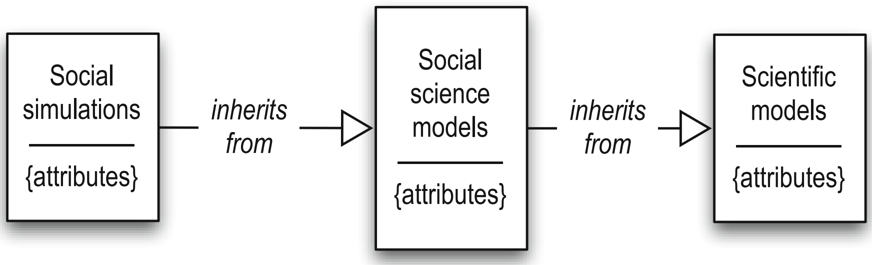

Fig. 1. UML class diagram illustrating the hierarchy of scientific models (left), social science models (center), and social simulations (right), each having increasingly specific standards for judging quality (left to right). *Source:* Prepared by the author.

Table 1. Quality Criteria for Evaluating Models in Domains of Science

Models in ...	Truth	Beauty	Justice	Additional
Science	Yes	Yes	No	No
Social science	"	"	Yes	"
Social simulation	"	"	"	Yes

Source: Prepared by the author.

Common practices such as verification and validation are well-known quality control procedures for assessing the quality of scientific models in general (Cover and Curd, 1998). However, verification and validation are insufficient criteria for assessing the quality of social science models, specifically for social simulations. An important implication is that current emphasis on model verification and validation is warranted (Cioffi, 2010; Sargent, 2004), but *verification and validation are insufficient by themselves for judging the quality of a social simulation model* (agent-based or other).

Therefore, a key methodological question concerning quality is: Which additional criteria—i.e., beyond Truth, Beauty, and Justice—could or should be used to assess the quality of a social simulation model? The next section addresses this question by proposing a set of dimensions for evaluating the quality of a given social simulation model.

2 Dimensions of Quality in Social Simulation Models

The quality of any complex artifact—whether a social simulation model or the International Space Station—is a multifaceted property, not a single dimension. Dimensions of quality can be used for evaluation as well as for a checklist of desirable attributes for building and developing a social simulation model. Arguably, there are two levels of quality assessment for computational social simulations, corresponding to the concepts of *a model* and *modeling*, respectively.

First, from a model's perspective, any set of quality dimensions for evaluating a social simulation must be based on its specific attributes or uniquely constituent features as a computational artifact in the sense of Simon (1996). Moreover, whether the overall quality of a given model should be an additive or a multiplicative function of individual qualitative features is less important than the idea that overall quality depends on a set of dimensions or desirable features beyond the Lave-March criteria, not on some single preeminent feature (e.g., simulation environment or programming language).

Second, from a modeling perspective, quality assessment should cover the broader modeling or model-building process as such, beyond the social simulation model that is produced in a narrow sense. This is because a computational model in final (i.e., committed) instantiated code is the result of a sequence of earlier modeling stages that precede the model itself, such as the critical stage of model design prior to implementation. Quality in design affects quality in the product of implementation, even when implementation *per se* is carried out in a proper manner (i.e., competently, with effectiveness and efficiency).

The following framework for quality assessment combines both perspectives by focusing on the classical methodological stages of social simulation model development:

1. Formulation
2. Implementation
3. Verification
4. Validation

5. Analysis
6. Dissemination

Such a Lifecycle Framework provides a viable checklist of quality dimensions to consider, based on the preceding methodological principles for social simulation. Note that verification and validation constitute only two contexts for assessing quality and, as shown below, some of the others involve quite a number of aspects regarding quality evaluation.

1. **Formulation.** Quality can be assessed starting from the formulation of a research problem that a given social simulation is supposed to solve. A first set of quality assessments regards research questions: Is the research question or class of research questions clearly formulated? Is the focal or referent empirical system well defined? Beyond clarity, is the research question original and significant? Originality should be supported by complete and reasoned surveys of prior extant literature to assess scientific progress. Every computational simulation model is designed to address a research question, so clarity, originality, and significance are critical. Motivation is a related aspect of problem formulation. Is the model properly motivated in terms of relevant extant literature? Or, is the simulation model the very first of its kind? If so, are there prior statistical or mathematical models in the same domain? Regrettably, incomplete, poorly argued, or totally missing literature reviews are rather common in social simulation and computational social science.
2. **Implementation.** Rendering an abstracted model in code involves numerous aspects with quality-related implications, starting with aspects of instantiation selection. Does the code instantiate relevant social theory? Is the underlying social theory instantiated using a proper program or programming language? Code quality brings up other aspects that may be collectively referred to as "The Grimson-Guttag Standards" (Guttag, 2013): Is the code well-written? Is the style safe/defensive? Is it properly commented? Can it be understood with clarity one year after it was written? In addition, what type of implementation strategy is used? I.e., is the model written in native code or using a toolkit? If toolkit (Nikolay and Madey, 2009), which one, why, and how good is the application? Is the choice of code (native or toolkit) well-justified, given the research questions? In terms of "nuts and bolts," quality questions include such things as what is the quality of the random number generator (RNG)? Think Mersenne Twister (Luke, 2011), MT19937, or other PRNG. Which types of data structures are used, given the semantics? Are driven threshold dynamics used? If so, how are the firing functions specified? In terms of algorithmic efficiency, What is the implementation difficulty of the problem(s) being addressed by the model? How efficient is the code in terms of implementing the main design ideas? In terms of computational efficiency, how efficient is the code in terms of using computational resources? This differs from algorithm efficiency. From the perspective of architectural design, is the code structured in a proper and elegant manner commensurate to the research question? In terms of object ontology, does the model instantiate the object-based ontology of the focal

system for the chosen level of abstraction? These quality-related questions precede verification and validation.

3. **Verification.** Which passive and active tests were conducted to verify that the model is behaving in the way it is intended to behave? Social scientists also call this internal validity. Verification tests include but are not limited to the following: Code walkthrough, debugging, unit testing, profiling, and other common procedures used in software development (Sergeant, 2004). What where the results of such verification tests? Quality assessment should cover investigation of which verification procedures were in fact used, since results can range widely depending on the extent of verification methods employed. Unfortunately, most social simulations are reported without much (or any) information regarding verification procedures, as if "results speak for themselves."
4. **Validation.** Similarly, validation of a social simulation, what social scientist call external validation (or establishing external validity), consists of a suite of tests, not a single procedure. Such tests are important for assessing quality in a social simulation. Which tests (histograms, RMSE for assessing goodness of fit, time series, spatial analysis, network structures, and other forms of real vs. artificial pattern matching tests) were conducted to validate the model? What were the results? Validation tests are often the focus of reporting results, at the expense of all other phases in the lifecycle of a social simulation model.
5. **Analysis.** The preceding aspects provide a basis for establishing overall confidence in a given model. What is the level of confidence in the model's results, given the combined set of verification and validation tests? If networks are present and significant in the focal system, does the model exploit theory and research in social network analysis (Wasserman and Faust, 2005)? Does the model facilitate analysis of complexity in the system of nonlinear interactions and emergent properties? Which features of complexity (emergence, phase transitions, power-laws or other heavy-tailed distributions, criticality, long-range dynamics, near-decomposability, serial-parallel systems, or other structural features) are relevant to the particular model? If spatial features are significant, does the simulation employ appropriate spatial metrics and statistical tools for spatial data? What is the overall analytical plan in terms of simulation runs and how is it justified? How does computational analysis advance fundamental or applied understanding of social systems? In terms of overall effectiveness, does the model render what is necessary for answering the initial research question or class of research questions? This differs from efficiency. In terms of the simulation's computational facilities, does the model possess the necessary functionality for conducting extensive computational analysis to answer the research questions or go even beyond? How powerful is the model in terms of enabling critical or insightful experiments? For example, in terms of parameter exploration (evolutionary computation) and record-keeping. What is the quality of the physical infrastructure that renders the most effective simulation experience?

6. **Dissemination.** Finally, the quality of a social simulation should be assessed in terms of its "life-beyond-the-lab." For instance, in terms of pedagogical value: Does the model teach well? I.e., does it teach efficiently and effectively? In terms of communicative clarity and transparency: Are useful flowcharts and UML diagrams of various kinds (class, sequence, state, use case) provided for understanding the model? Are they drawn with graphic precision and proper style (Ambler, 2005)?In terms of replicability, what is the model's replication potential or feasibility? How is reproducibility facilitated? Aspects related to a model's graphics are also significant for assessing quality, not just as "eye candy." In terms of GUI functionality, is the user interface of high quality according to the main users? Is the GUI fundational for answering the research questions? More specifically, in terms of visualization analytics: Is visualization implemented according to high standards (Thomas and Cook, 2005)? This does not concern only visual quality (Tufte, 1990), but analytics for drawing valid inferences as well (Cleveland, 1993; Few, 2006; Rosenberg and Grafton, 2010). In terms of "long-term care:" What is the quality of the model in terms of curatorial sustainability? How well is the model supported in terms of being easily available or accessible from a long-term perspective? In which venue (Google Code, Sourceforge, OpenABM, Harvard-MIT Data Center/Dataverse, or documentation archives such as the Social Science Research Network SSRN) is the model code and supplementary documentation made available? Finally, some social simulations are intended as policy analysis tools. Is the model properly accredited for use as a policy analysis tool, given the organizational mission and operational needs of the policy unit? Does the model add value to the overall quality of policy analysis? Does it provide new actionable information (new insights, plausible explanations, projections, margins of error, estimates, Bayesian updates) that may be useful to decision-makers?

3 Discussion

Verification and validation are obviously essential dimensions of quality in social simulation models; but they are just two among other dimensions of interest in assessing quality across the full spectrum of model development stages. Thus, and contrary to common belief, verification and validation may be viewed as necessary but insufficient conditions of a high-quality social simulation. The many dimensions of quality in social simulations extend far beyond and much deeper than the Lave-March TBJ criteria. Some of these criteria may be viewed as utilitarian (e.g., based on resources), while others are non-utilitarian (based on style).

Quality has dimensions because it is a latent concept, rather than a single directly measurable property. Therefore, proxies (i.e., measurable dimensions or attributes) are needed. The quality dimensions proposed in the preceding section provide a viable framework while social simulation develops as a field, not as a permanent set of fixed criteria. Each stage in the Lifecycle Framework contains numerous dimensions for quality assessment because social simulations are complex artifacts, in the sense of Simon.

Interestingly, Osgood's first dimension in cognitive EPA-space is Good-Bad (evaluation). This is why quality evaluation (good-bad-ugly) is essential (Osgood, May, and Miron, 1975). The proposed criteria should allow a classification of social simulations into categories of good, bad, or outright ugly.

As computational social scientists we need to better understand the micro-processes that compose the overall quality of social simulation:

- How is a problem chosen for investigation?
- How is the problem-space reduced by abstraction?
- How is the model designed?
- How well are the entities and relations understood?
- How is the simulation language chosen?
- How is the model implemented?
- How are verification and validation conducted?
- How are simulation runs actually conducted?
- How is the model being maintained?

Requiring additional quality criteria for social simulation models is not an argument against the unity of science. It is a plea for greater specificity and more rigor in the evaluation of quality in the field of social simulation.

From a methodological perspective, quality criteria could also help support the (virtual) experimental function of social simulations in terms providing ways for assessing the veracity of artificial worlds. Computational experiments could thus be framed within the context of a social simulation characterized by a set of quality features, taking all lifecycle stages into consideration, as in evaluating experimental results *conditional upon* the quality of the social simulation model. Such a function would enhance the value of social simulation as an experimental method and highlight its scientific usefulness.

Finally, the topic of quality in social simulations also motivates a broader discussion concerning similarities and differences between social simulation and other scientific approaches in science, such as statistical and mathematical models. While all scientific approaches share some of the same quality criteria, each has also unique quality criteria that are not applicable in other approaches. For social simulations there are aspects such as quality of code or visualization dashboards that are *sui generis* to the approach itself. The lifecycle approach to assessing quality in social simulations could also shed new light on parallel efforts in statistical and mathematical models of social systems, since there too we find a similar sequence of stages, from the formulation of research questions to analyzing and communicating model results, albeit with significant variations in technical details if not in the overall process.

4 Summary

Computational social science arises from a number of research traditions that have roots in The Enlightenment and even earlier origins in Aristotle's comparative analysis of social systems. Therefore, our existing standards of scientific quality and excellence have been inherited through the history and philosophy of science in terms of basic principles, such as formalization, testing, replication, and dissemination.

More specifically, the properties of Truth, Beauty, and Justice proposed for mathematical social science (Lave and March, 1993) in an earlier generation remain equally valid quality criteria for assessing social simulation models. But useful as such classic standards of quality may be, social computing adds new scientific features (e.g., emphasis on understanding complex adaptive systems, object-oriented ontologies, network structures that can evolve in time, nonlinear dynamics) that require development as new standards for quality evaluation. Social simulation models in particular (e.g., agent-based modeling) require further specific requirements for judging quality. This paper proposed a set criteria for discerning quality in social simulations, especially agent-based models, based on universal stages of simulation model development. These criteria are offered as an initial heuristic framework to consider and develop as a work-in-progress, not as a finalized set of fixed criteria.

Acknowledgements. Funding for this study was provided by the Center for Social Complexity of George Mason University and by ONR MURI grant no. N00014-08-1-0921. Thanks to members of the Mason-Yale Joint Project on Eastern Africa (MURI Team) for discussions and comments, and to Joseph Harrison for converting the earlier LaTeX article into LNCS style. The opinions, findings, and conclusions or recommendations expressed in this work are those of the author and do not necessarily reflect the views of the sponsors. This paper is dedicated to the memory of Auguste Comte (1798–1857), who lived at the Hotel de Saint-Germain-des-Pres, Paris, where this paper originated during the 2010 Quality Commons workshop.

References

1. Ambler, S.W.: The Elements of UML 2.0 Style. Cambridge University Press, Cambridge (2005)
2. Cioffi-Revilla, C.: Simplicity and Reality in Computational Modeling of Politics. Computational and Mathematical Organization Theory 15(1), 26–46 (2008)
3. Cioffi-Revilla, C.: On the Methodology of Complex Social Simulations. Journal of Artificial Societies and Social Simulations 13(1), 7 (2010)
4. Cioffi-Revilla, C.: Comparing Agent-Based Computational Simulation Models in Cross-Cultural Research. Cross-Cultural Research 45(2), 1–23 (2011)

5. Cioffi-Revilla, C., Gotts, N.M.: Comparative Analysis of Agent-Based Social Simulations: GeoSim and FEARLUS Models. Journal of Artificial Societies and Social Systems 6(4) (2003)
6. Cioffi-Revilla, C., Rouleau, M.: MASON RebeLand: An Agent-Based Model of Politics, Environment, and Insurgency. International Studies Review 12(1), 31–46 (2010)
7. Cleveland, W.S.: Visualizing Data. AT&T Bell Laboratories, Murray Hill (1993)
8. Cover, J.A., Curd, M. (eds.): Philosophy of Science: The Central Issues. W. W. Norton, New York (1998)
9. Doran, J.E.: Iruba: An Agent-Based Model of the Guerrilla War Process. In: Troitzsch, K.G. (ed.) Representing Social Reality: Pre-Proceedings of the Third Conference of the European Social Simulation Association (ESSA), pp. 198–205. Verlag Dietmar Foelbach, Koblenz (2005)
10. Epstein, J.M., Axtell, R.: Growing Artificial Societies: Social Science from the Bottom Up. The Brookings Institution, Washington, D.C. (1996)
11. Few, S.: Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly, Sebastopol (2006)
12. Gilbert, N., Troitzsch, K.: Simulation for the Social Scientist, 2nd edn. Open University Press, Buckingham (2005)
13. Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D.L.: Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science 310, 987–991 (2005)
14. Guttag, J.V.: Introduction to Computation and Programming Using Python. MIT Press, Cambridge (2013)
15. Hempel, C.G.: Aspects of Scientific Explanation. Free Press, New York (1965)
16. Hoffmann, M.J.: Entrepreneurs and the Emergence and Evolution of Social Norms. In: Urban, C. (ed.) Proceedings of Agent-Based Simulation 3 Conference, pp. 32–37. SCS-Europe, Ghent (2002)
17. Hoffmann, M.J.: Self-Organized Criticality and Norm Avalanches. In: Proceedings of the Symposium on Normative Multi-Agent Systems, pp. 117–125. AISB (2005)
18. Kaplan, A.: The Conduct of Inquiry. Chandler, San Francisco (1964)
19. Lave, C.A., March, J.G.: An Introduction to Models in the Social Sciences. University Press of America, Lanham (1993)
20. Luke, S.: Mersenne Twister (2011), `http://www.cs.gmu.edu/~sean/research/` (accessed May 30, 2011)
21. Meeker, B.F.: Some Philosophy of Science Issues in the Use of Complex Computer Simulation Theories. In: Szmatka, J., Lovaglia, M., Wysienska, K. (eds.) The Growth of Social Knowledge: Theory, Simulation, and Empirical Research in Group Processes, pp. 183–202. Praeger, Westport (2002)
22. Nikolai, C., Madey, G.: Tools of the Trade: A Survey of Various Agent Based Modeling Platforms. Journal of Artificial Societies and Social Simulations 12(2), 2 (2009)
23. Osgood, C.E., May, W.H., Miron, M.S.: Cross-Cultural Universals of Affective Meaning. University of Illinois Press, Urbana (1975)
24. Rouchier, J., Cioffi-Revilla, C., Gary Polhill, J., Takadama, K.: Progress in Model-to-Model Analysis. Journal of Artificial Societies and Social Simulation 11(2), 8 (2008)

25. Rosenberg, D., Grafton, A.: Cartographies of Time: A History of the Timeline. Princeton Architectural Press, New York (2010)
26. Sargent, R.G.: Verification and Validation of Simulation Models. In: Henderson, S.G., Biller, B., Hsieh, M.H., Shortle, J., Tew, J.D., Barton, R.R. (eds.) Proceedings of the 2007 Winter Simulation Conference. IEEE Press, Piscataway (2004)
27. Schelling, T.C.: Dynamic models of segregation. Journal of Mathematical Sociology 1(2), 143–186 (1971)
28. Simon, H.A.: The Sciences of the Artificial, 3rd edn. MIT Press, Cambridge (1996)
29. Taber, C.S., Timpone, R.J.: Computational Modeling. Sage Publications, Thousand Oaks (1996)
30. Thomas, J.J., Cook, K.A. (eds.): Illuminating the Path. IEEE Computer Society, Los Alamitos (2005)
31. Tufte, E.R.: Envisioning Information. Graphics Press, Cheshire (1990)
32. Wasserman, S., Faust, K.: Social Network Analysis. Cambridge University Press (1994)