

Empiricism for descriptive social network models

Camille Roth^{a,*}

^a*European Center for Living Technology, Dorsoduro 3825, 30123 Venice, Italy;
and CREA (Center for Research in Applied Epistemology), CNRS/Ecole
Polytechnique, 1 rue Descartes, 75005 Paris, France.*

Abstract

Social complex system modeling is at the core of a recent interdisciplinary effort, in which statistical physics has been playing a notable role; yet, the unfrequent validation of behavioral hypotheses possibly leads to normative rather than descriptive models, plausibly less appealing to social scientists. An epistemological insight and hindsight on this whole program is proposed, emphasizing a strong empirical methodology which extends to low-level agent-based dynamics. Special attention is given to the recent interest in knowledge diffusion models.

Key words: social network models, epistemology & empirical methodology, social complex systems, knowledge diffusion.

PACS: 01.70.+w, 89.65.-s, 89.75.-k

Understanding how knowledge is collectively elaborated and diffused entails a novel perspective on social epistemology and social cognition, in which the modeling of underlying networks strongly relies on a fine apprehension of knowledge-based interactions at the level of agents — with respect both to the agent behavior and to the context of this behavior, i.e. the social network itself. As recent advances in computing capabilities and electronic data availability for several social systems (scientists, webloggers, online customers etc.) made possible a wide range of empirical validation experiments, these issues have known a renewed interest within an interdisciplinary effort aiming at modeling “social complex systems”, in which statistical physics has been playing a notable role, along with disciplines such as mathematical sociology and computer science, relying altogether extensively on graph theory.

Specifically, this research program is addressing a wide range of phenomena [for a review see 7], including the morphogenesis and structure of interaction

* *E-mail:* camille.roth@polytechnique.edu — *Fax:* +1 (413) 460-7085.

networks and communities studied at an aggregated level [12, 16, 22, 25, 27], to more local processes, such as individuals adopting opinions in so-called recommendation and influence networks [6, 14, 19, 21], or actors trying to establish common decisions in epistemic communities [2, 5, 17, 28]. Overall however, the behavioral hypotheses driving these models, even sociologically, cognitively or anthropologically credible, are often mathematical abstractions whose empirical measurement and justification seem to be rather occasional. My aim here is to provide an epistemological insight and hindsight on this whole program, by insisting on a strong empirical methodology concerning both higher-level phenomena and lower-level agent-based interactions. I shall first briefly recall and detail some goals and methods of (social) complex system modeling, then focus on the renewed interest in knowledge diffusion models.

1 Social complex system modeling

Epistemological approach. Recent approaches in social complex system modeling have been massively calling for social network- and agent-based models to appraise various kinds of stylized facts; including, to cite a few, characteristics of the connectivity of agents among their peers [22, 27], diffusion velocity of some knowledge within a social group [6], structure of communities [12, 25, 32], etc. These issues relate more broadly to the question of reconstruction: after pointing out and observing some relevant stylized facts for a given system, the aim is traditionally to propose a model which either rebuilds, explains and/or possibly predicts these facts. If the model is solid enough, one could even hope to reveal new and potentially counter-intuitive stylized facts that were not initially observed.

In social science, scientists are using more and more frequently methods of social network analysis (SNA) to infer and reproduce “high-level” phenomena which would traditionally have undergone a strictly high-level description: for instance qualifying the cohesion of a community, finding the roots of a crisis, explaining how roles are distributed, etc. By doing so, they are clearly exhibiting a formal relationship between traditional sociological descriptions and the more abstract structure of an underlying model, based on a graph binding nodes symbolizing agents and interactions — they reconstruct the “social structure” [11], benchmarked against already-established descriptions — the benefit being often that low-level information is easier to collect, more practical to observe and/or entails more robust descriptions [3].

Reconstruction and simulation. In general, this kind of reconstruction is a reverse problem consisting in successfully involving a lower level of agents and agent-based interactions in order to rebuild some descriptions concern-

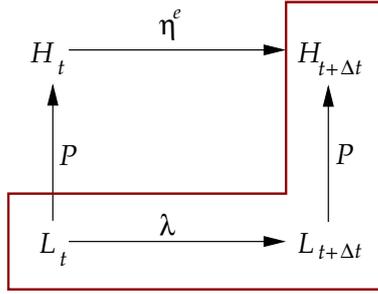


Fig. 1. Reconstructing high-level descriptions H and dynamics η^e from low-level states L and dynamics λ , through a mapping P (see [24, 30] for comprehensive discussions on this kind of diagrams).

ing either this lower level itself, or a higher level of macroscopic descriptions (communities, global structures). In formal terms, given some kind of high-level phenomena to be modeled “ H ”, and an empirical dynamics on them, η^e , agent-based models usually rely on distinct, interacting objects at a level “ L ” animated by some dynamics λ .¹ For a given state on objects L , a corresponding interpretation in terms of “classical” observables H is provided; let us call this transformation P , such that $P(L) = H$. Subsequently, a modeler would propose a dynamics λ on L such that the result matches the original dynamics of H : $P \circ \lambda(L) = \eta^e(H)$. This should eventually provide a commutative diagram, which is familiar in dynamical systems study [24, 30]: $P \circ \lambda = \eta^e \circ P$.

Considering the SNA example again, suppose that H describes the community structure within a social group while L denotes the social network made of agents, links, and possibly individual properties. Some appropriate community finding algorithm can provide P by matching L with H [9, 10, 12, 25, 32]. The modeler would then try to design λ by describing a network morphogenesis mechanism (for instance link additions based on agent preferences) such that, through P , the model eventually reproduces η^e . In turn, the focus and expectations of the corresponding models vary greatly: some aim at reproducing either actual system states, or only a few statistical parameters (e.g. the exact distribution of a given variable, the same type of law, just some sort of power-law tail), functions which simply exhibit the same behavior, or even the mere existence of some class of attractors (equilibrium, structural robustness).

2 Empirical benchmarking

The success of the reconstruction endeavor depends on the capacity of “ $P \circ \lambda$ ” to rebuild η^e , which must be appraised with respect to an empirical benchmark

¹ Apart from agent-based models — in general, outside the “complex system” enterprise — L often corresponds directly to H , in which case $P = \text{Id}$.

— even a rough one. This argument remains valid whether the underlying model is simulation-based or purely analytical. Typically, one already has η^e — under the form of a series of empirical measurements, or at least as a well-established theory; that is, a more or less stylized η^e . Two attitudes are available:

- (i) Either rebuild η^e by proposing a *sufficiently* valid λ , in which case the model could be used to explain η^e from a different perspective (agent-based, obviously), i.e. to suggest that some stylized fact is “nothing but” due to the systemic integration of a particular kind of agent behavior. Then, the usefulness of the reconstruction stems from its ability to predict the future behavior of the whole system, or to suggest that some effect could be deleted by acting on λ in a certain way (going towards normative models).
- (ii) Or find a new, unexpected behavior on H : the η derived from $P \circ \lambda$ has some properties which were not known beforehand in η^e , yet happen to be empirically correct.

At this point, by choosing the ontology of the lower-level L the modeler should ensure that he also has an empirically valid mapping P .²

Realistic models. What are the responsibilities of each of those two attitudes? In the first case “(i)”, while it is already a great achievement to substantiate and re-discover the stylized facts from another viewpoint, it is also pivotal to check if λ is realistic rather than alleged. Otherwise, the agent-based approach might turn to be considered slightly superfluous: it could indeed be argued that η could be modeled directly without going through a possibly unchecked λ .

In practice however, this empirical endeavor on λ seems to be unsystematic. Many morphogenesis models have attempted to reproduce, for instance, the abnormally high clustering coefficient of some collaborative social networks (such as scientists [4, 22], movie actors [8] and corporate board members [23]) by using elaborated mechanisms based on dyadic interactions and subsequent addition of dyadic links. In contrast, real-world situations appear to feature n-adic interactions (article writing, simultaneous co-appearance in boards or movies) which correspond, in a classical graph, to clique additions. Some re-

² For instance, as suggested above, community-finding methods are designed to match a community structure H with a social-network-based description L . In this area of research, the “karate-club” example [12] is a classical empirical benchmark of P : given an already-known community structure of a given group of individuals, relevant for social scientists, for a given a priori definition of what a community should be, is a candidate algorithm able to find the same structure? Once this is checked, would that algorithm in turn reveal new, *a priori* unsuspected communities?

cent models using very basic n -adic interaction mechanisms appear to rebuild straightforwardly many statistical parameters such as a high clustering coefficient and realistic degree distributions [16, 28, 29, analytical proof for the general case in [15]] — which also suggest that hypergraphs are better-suited than graphs to model social network morphogenesis in such case, thus inducing a new design of L . More broadly, what would be the reach of dyadic-interaction-based models when their apparent high-level success is based on dubious low-level behavioral assumptions?

A questionable λ could thus impair the empirical value of the model, whereas a modest and simplistic yet faithful λ could be preferred. In other terms, λ *is certainly an approximation*, as with any model, but what λ *explicitly* describes should be valid, even if some part of reality is omitted. I would here suggest that, even before solving or running a model, λ should be designed in accordance with a strong program of empirical validation — this would allow descriptive models instead of normative models.³

In the second case, “(ii)”, this skeptic attitude is even more crucial: one would like to avoid that the system dynamics η induced by $P \circ \lambda$ correctly matches an empirical η^e but does so only for this very η^e , whereas new, unexpected properties of η are incorrect. It would therefore be worth checking the validity of both λ *and* the new η . Corresponding simulations and analytical solutions might otherwise turn to have limited benefits, or, repeat what one already knows (that is, η^e) *with no further generalizing power*.

A last attention should be paid to model stability with respect to its hypotheses — even realistic. Put differently, would a model exhibit a continuous behavior with respect to continuous modifications in the hypotheses, which are themselves necessarily stylized to some extent? For instance, the celebrated Barabasi-Albert model of network formation [1] induces a power-law distribution of degrees when nodes join the system at a constant rate, respecting a linear preferential attachment behavior, that is, a preference to attach to other nodes proportionally to their degree. In the real world however, this behavior has been later measured to be slightly sub- or super-linear in many cases [2]; but even a slight discrepancy seems to bear crucial effects for the resulting degree distribution [20].

³ This stance should remain valid even for very stylized approaches and hypotheses: for instance, it is clearly wise for a model to show that the system behavior could drastically change for a certain critical level of a given quantity. In turn, it might be useful to assess the empirical meaning of this critical value: for a stylized parameter, how does a given value translate concretely? Can it even be reached?

3 The case of knowledge diffusion models

The question of appraising λ becomes critical when designing diffusion and influence models, since such phenomena intimately intertwine agent behavior and structural effects. More broadly, it is now widely accepted that the underlying network structure can significantly impact social system behavior, particularly knowledge diffusion [5, 31, 34]: “It is as unthinkable to study diffusion without some knowledge of the social structures in which potential adopters are located as it is to study blood circulation without adequate knowledge of the structure of veins and arteries” (Katz in [18], cited in [6]). In line with the previous arguments, even when assuming that the network structure is understood, again, an empirical stance should nonetheless be adopted to appraise agent-based transmission mechanisms.⁴ Yet, as Leskovic, Adamic and Huberman [21] put it, “[while former] models address the question of maximizing the spread of influence in a network, they are based on assumed rather than measured influence effects.”

Estimating behaviors in diffusion models. Agent-based knowledge diffusion models have been mainly introduced by Granovetter and his threshold model [13] at the end of the 1970s. In this model, individuals are subject to the influence of their social network neighbors; an agent is supposed to adopt a given cultural item if a certain proportion of his neighbors also have it (the “threshold”). Distinct kinds of agents could be identified, with distinct threshold values. Improvements focus on methods for weighting and counting the (non-linear) influence of neighbors — a feature not unfamiliar to formal neural network models. Cascade models, on the other hand, assume the existence of a given probability for each agent to believe each of his neighbors [19]; this mechanism is also close to a class of models stemming from biological epidemiology, based on SIS models (“susceptible-infected-susceptible”) [26, 34], for which several qualitative modifications are conceivable. Additionally some more recent models, linked to economics and cultural anthropology, involve the exchange of knowledge items or skills among former and possibly new neighbors [5, 29]; cultural “contamination” stems here from successive interactions of agents. Note that in contrast to above models, transmission of multiple

⁴ Indeed, knowledge diffusion can be a slow dynamics process, in that it can possibly occur at a timescale comparable to that of the evolution of the social network itself, it could be key to take into account social network morphogenesis *in coevolution with* knowledge transmission mechanisms, especially if the latter have an impact on the former (this should be particularly true for scientific ideas for instance, rather than for rumors, which seem to diffuse at a more rapid pace than friendship links). Clearly, in such settings asymptotic behaviors could be less informative than expected.

knowledge items is considered in a parallel yet non-independent manner.⁵

Are these mechanisms realistic for some real-world situation? For the threshold model for instance, there have already been qualitative and roughly quantitative estimations within some particular groups & situations [31], which provided an extremely valuable insight on key influence phenomena: existence of classes of agents with particular behaviors, estimations of threshold values for each class, *inter alia*. Yet, while all above-mentioned influence mechanisms sound intuitive and credible, even reasonably close to what social psychology or social epistemology could qualitatively suggest, they are plausibly contradictory one with another, and would thus not be relevant for all a general purpose. More to the point, a recent attempt at measuring the influence of neighbor recommendations for buying some products [21] reveals a decreasing adoption probability with respect to received influence. Such result appears to be inconsistent with existing influence models, unless one assumes a totally heterogeneous agent behavior — in which case it is unclear whether the data would allow to fit distinct adoption thresholds or probabilities for each pair of agents. New empirically consistent influence mechanisms are certainly required.

Finally, most models bear the underlying assumption that an individual is permanently under the influence of all his neighbors, friends, colleagues: social network connections are considered permanently active. While this feature is likely to be verified in neural networks and computer networks, in contrast, social networks are structures which simply represent past acquaintances, which could potentially induce future interactions: as Douglas White underlines, “[o]ne way of using the coding of networks is to regard them as the precipitates of past behavioral interactions” [33]. As such, the social network should be seen as a framework wherein social interactions may take place: for any given period, actual dyadic interactions within the system consist only of a subset of links present in the social network.

Concluding remarks. Empirical tests could wisely be suggested to come *before* any modeling attempt: modeled dynamics of *both* higher and lower levels should therefore adopt a strong discipline of empirical validation and realistic design, especially concerning λ . Otherwise, we could be likely to achieve appealing normative models, obviously useful in some settings, such as organizational optimization, but rather seldom sought by social scientists.

⁵ More comprehensive typologies of influence models are presented in [14, 31].

References

- [1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [2] A.-L. Barabási, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and T. Schubert. Evolution of the social network of scientific collaborations. *Physica A*, 311:590–614, 2002.
- [3] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *PNAS*, 99(3):7280–7287, 2002.
- [4] M. Catanzaro, G. Caldarelli, and L. Pietronero. Assortative model for social networks. *Physical Review E*, 70:037101, 2004.
- [5] R. Cowan and N. Jonard. Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28:1557–1575, 2004.
- [6] F. Deroian. Formation of social networks and diffusion of innovations. *Research Policy*, 31:835–846, 2002.
- [7] S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of Networks — From Biological Nets to the Internet and WWW*. Oxford: Oxford University Press, 2003.
- [8] P.-P. Z. et al. Model and empirical study on some collaboration networks. *Physica A*, 360(2):599–616, 2006.
- [9] M. G. Everett and S. P. Borgatti. Analyzing clique overlap. *Connections*, 21(1):49–61, 1998.
- [10] K. A. Frank. Identifying cohesive subgroups. *Social Networks*, 17(27-56), 1995.
- [11] L. C. Freeman. Social networks and the structure experiment. In L. C. Freeman, D. R. White, and A. K. Romney, editors, *Research Methods in Social Network Analysis*, pages 11–40. Fairfax, Va.: George Mason University Press, 1989.
- [12] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *PNAS*, 99:7821–7826, 2002.
- [13] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1987.
- [14] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of WWW2004*, NYC, NY, USA, May 17-22 2004.
- [15] J.-L. Guillaume and M. Latapy. Bipartite structure of all complex networks. *Information Processing Letters*, 90(5):215–221, 2004.
- [16] R. Guimera, B. Uzzi, J. Spiro, and L. A. N. Amaral. Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308:697–702, 2005.
- [17] P. Haas. Introduction: epistemic communities and international policy coordination. *International Organization*, 46(1):1–35, winter 1992.
- [18] E. Katz. The social itinerary of technical change: two studies on the diffusion of innovation. In S. Wilbur, editor, *Studies of Innovation and of Communication to the Public*. Institute for Communication Research, Stanford University, 1961.
- [19] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, New York, NY, USA, 2003. ACM Press.

- [20] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85:4629–4632, 2000.
- [21] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *ACM Conference on Electronic Commerce*, pages 228–237, 2006.
- [22] M. E. J. Newman. The structure of scientific collaboration networks. *PNAS*, 98(2):404–409, 2001.
- [23] M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(036122), 2003.
- [24] M. Nilsson-Jacobi. Hierarchical organization in smooth dynamical systems. *Artificial Life*, 11(4):493–512, 2005.
- [25] G. Palla, I. Dernyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 2005.
- [26] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.
- [27] W. W. Powell, D. R. White, K. W. Koput, and J. Owen-Smith. Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110(4):1132–1205, 2005.
- [28] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras. Self-organization of collaboration networks. *Physical Review E*, 70:036106, 2004.
- [29] C. Roth. Co-evolution in epistemic networks – reconstructing social complex systems. *Structure and Dynamics: eJournal of Anthropological and Related Sciences*, 1(3):art2, 2006.
- [30] A. Rueger. Robust supervenience and emergence. *Philosophy of Science*, 67(3):466–489, 2000.
- [31] T. W. Valente. *Network Models of the Diffusion of Innovations*. Hampton Press, 1995.
- [32] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.
- [33] D. R. White. Networks and hierarchies. In D. Lane, D. Pumain, S. van der Leeuw, and G. West, editors, *Complexity Perspectives on Innovation and Social Change*. Springer, To appear (2007).
- [34] F. Wu, B. A. Huberman, L. A. Adamic, and J. R. Tyler. Information flow in social groups. *Physica A*, 337:327–335, 2004.