Patterns and processes in socio-semantic networks

Camille Roth*

UK Social Network Conference — July 13–14, 2007 — Extended abstract proposition

Communities of agents manipulating, producing and exchanging knowledge are forming as a whole a socio-semantic complex system achieving widespread social cognition — with concepts being introduced and manipulated in a rather decentralized fashion. Scientists, software developers, political webloggers are examples, among others, of such distributed knowledge construction systems. Questions traditionally arise as to how these communities are forming, evolving, interacting, and, more broadly, what kind of processes are at work within them — particularly as regards knowledge diffusion. Networks, in particular, have emerged as an unavoidable formal framework to capture essential stylized facts of the structure of these knowledge communities, addressing issues pertaining to social epistemology, knowledge economics and cultural anthropology (Kitcher, 1995; Cohendet *et al.*, 2003), along with new toolboxes borrowing extensively to graph theory and systems dynamics (Newman, 2003).

Often, though, knowledge networks are treated like any other real network, with agents behaving in a way sometimes not much more complex than molecules. Even when the behavioral complexity of agents is taken into account, social network models seem to neglect epistemic features. Our goal is to emphasize the intertwining of social and semantic networks, both theoretically and empirically: we first suggest that binding these networks yields new kinds of patterns, showing notably how community structure may subsequently be appraised. We then present some implications on how structural dynamics and processes could be considered and modeled. At the same time, we will sketch out and describe an empirical application on a particular system of scientists working on a well-bounded domain.

Towards epistemic patterns. Network structure is classically described through "patterns", that is, statistical parameters computed on an underlying graph representing inter-agent relationships. Patterns are designed to relevantly match sociological descriptions such as, for instance, leadership position, transitivity and community cohesiveness. In this respect, existing patterns include, to cite a few, clustering coefficients and "cliquishness" (Watts and Strogatz, 1998; Robins and Alexander, 2004), node degree distribution and the broadly shared "*scale-free*" property (Barabási and Albert, 1999), largest connected component size and one-mode community structure (Pattison *et al.*, 2000; Girvan and Newman, 2002; Powell *et al.*, 2005). In turn, numerous models have subsequently been developed to provide an understanding of pattern formation by identifying key processes leading to some given shape (Skyrms and Pemantle, 2000; Albert and Barabási, 2002; Durlauf, 2001; Newman *et al.*, 2001, *inter alia*).

Yet, patterns and thus models are frequently adapted to single social networks only;[1] hence they do not seem to be able to account for the specificity of knowledge networks. To access new patterns, possibly prone to be adequate to knowledge networks, we introduce the notion of *epistemic network* as a compound of two co-evolving networks, a social network and its associated semantic network. Next, we show how epistemic networks are a natural framework for appraising knowledge community structure.

Epistemic communities. A knowledge community is a small, embedded sub-society of knowledge, with specific topics — partially independent, partially overlapping — which in turn appears to be structured in several im-

*Department of Sociology, University of Surrey, Guildford, GU2 7XH, United Kingdom. E-mail: *c.roth@surrey.ac.uk*

[1] Some economic models of knowledge creation already take agent profiles into consideration, to explain the structure of an economic network: Cowan *et al.* (2002) for instance feature agents who match two by two to produce new knowledge according to their profile, as elements of a vector space, whereas Boguna *et al.* (2004) introduce a general class of models where agents could connect with respect to other kinds of non-structural properties. In these cases however, such properties are principally used as agent-based features, rather than providing opportunity to design patterns which specifically take advantage of the duality.

plicit sub-communities, while expertise is heterogenously distributed over all agents. Boundaries appear between subgroups, both horizontally, with distinct areas of competence, and vertically, with various levels of specificity.

Such communities may be seen as "epistemic communities" (ECs), i.e. groups of agents who are working on and dealing with common topics, sharing a common goal of knowledge creation (Haas, 1992; Cowan *et al.*, 2000). While it remains questionable whether patterns relying on strict social networks could help detecting ECs, we suggest that a socio-semantic framework may be more adequate to this task. We therefore provide a formal definition of an EC as "the largest group of agents who share and work on the same concepts" — as such, this notion is not unrelated to structural equivalence (Lorrain and White, 1971). In this context, an EC is understood as a descriptive instance only, not as a coalition of people who have some interest to stay in the community: it is a set of agents who, simply, share the same knowledge concerns.

To implement this definition, we need a binary relation connecting agents on one side and "concepts" or knowledge units on the other side — as such a bipartite graph derived from the epistemic network. Assuming that such ECs are structured in fields and subfields of common concerns, we show that the Galois lattice structure is an appropriate method for representing knowledge network taxonomies, by automatically organizing a given group into hierarchic, embedded taxonomies of ECs. In addition, it accurately renders overlaps among epistemic communities, commonly called interdisciplinary fields. Applied on a community of embryologists interested in a particular model animal, EC taxonomies provide a surprisingly accurate epistemological description (Roth and Bourgine, 2006). On the whole, such patterns take advantage of the duality inherent to knowledge groups —in our case agents affiliated with topical categories (Breiger, 1974)— as opposed to single-network-based methods, using for instance social relationships or semantic proximity.

Interplay between structure and knowledge diffusion. Going further, as network structure generally affects propagation processes (Morris, 2000; Pastor-Satorras and Vespignani, 2001; Lloyd and May, 2001; Cowan *et al.*, 2002; Deroian, 2002), patterns proper to epistemic social networks may likewise be helpful to evaluate the impact of topological properties on information diffusion. Real-world knowledge networks are indeed plausibly behaving diversely from those created from classical morphogenesis models adapted to "universal" networks.

In particular, we will comment a recent study (Cointet and Roth, 2007) showing that, even for a simple knowledge diffusion protocol, several common network topology models do not make it possible to accurately reconstruct the behavior of a realistic network — in our case, the above-mentioned real network of embryologists. More precisely, we examine the diffusion dynamics of a single piece of information using a very basic knowledge transmission protocol on a simulation-based model: agents interact and get instant, perfect and irreversible knowledge of the information if their interlocutor has it. Using Erdös-Rényi random graphs (Erdős and Rényi, 1959, based on a uniform wiring probability p) and "scale-free" networks (Barabási and Albert, 1999, based on a power-law degree distribution $P(k) \sim k^{\alpha}$) as best approximations of the real network (that is, using empirically-measured parameter values), we notice that both models fail to reproduce diffusion phenomena as they happen in the corresponding empirical network.[2] Using however an improved network model based on joint events gathering groups of agents, which renders both the connectivity structure of a scale-free network and the local clustering of a cliquish network (Guillaume and Latapy, 2004), diffusion appears less dissimilar to real-network-based diffusion, thus indicating that *community-based* interaction behaviors may produce more realistic topologies.

Nevertheless, the performance of this latter model is better but remains fairly inaccurate. While patterns reproduced by these models may arguably be meaningful as such, they are possibly unsufficiently faithful to the original network so that they could serve as a basis for a knowledge diffusion model — even the simplest one. Obviously, even in the framework of epistemic networks (or co-evolving social and semantic networks), it is unlikely that there exists an "ultimate" morphogenesis model, i.e. one that ideally reconstructs any possible statistical parameter characterizing a network. Similarly, it should not be possible to exhibit ultimate topological properties. However, given a few, selected, relevant diffusion processes, it should be possible to identify a few relevant patterns which a morphogenesis model should reproduce, then benchmark the whole system behavior with real-world networks; thereby substantiating both morphogenesis and diffusion models (Roth, 2007).

In such a case, it might be necessary to diverge from universal statistical parameters such that the scale-free

[2]Worse, they both perform identically; suggesting that, for this kind of process, there is no improvement in considering a scale-free network over a uniform ER random network.

degree distribution, and explore more precise patterns adapted to knowledge networks, thus introducing new classes of (epistemic) networks. Considering altogether epistemic patterns and diffusion processes may constitute a crucial step in explaining how network structure affects concept propagation and, at the same time, how concept propagation in turn affects the network.

References

Albert, Reka and Barabási, Albert-Laszlo. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, *74*: 47–97.

Barabási, Albert-Laszlo and Albert, Réka. (1999). Emergence of Scaling in Random Networks. *Science*, *286*: 509–512.

Boguna, Marian, Pastor-Satorras, Romualdo, Diaz-Guilera, Albert and Arenas, Alex. (2004). Models of social networks based on social distance attachment. *Physical Review E*, *70*: 056122.

Breiger, Ronald L. (1974). The Duality of Persons and Groups. *Social Forces*, *53(2)*: 181–190.

Cohendet, Patrick, Kirman, Alan and Zimmermann, Jean-Benoît. (2003). Emergence, Formation et Dynamique des Réseaux – Modèles de la morphogenèse. *Revue d'Economie Industrielle*, *103(2-3)*: 15–42.

Cointet, Jean-Philippe and Roth, Camille. (2007). How Realistic Should Knowledge Diffusion Models Be? *Journal of Artificial Societies and Social Simulation*. (in press).

Cowan, Robin, David, Paul A. and Foray, Dominique. (2000). The Explicit Economics of Knowledge Codification and Tacitness. *Industrial & Corporate Change*, *9(2)*: 212–253.

Cowan, Robin, Jonard, Nicolas and Zimmermann, Jean-Benoît. (2002, July). *The Joint Dynamics of Networks and Knowledge*. Computing in Economics and Finance 2002 354. Society for Computational Economics.

Deroian, Frédéric. (2002). Formation of social networks and diffusion of innovations. *Research Policy*, *31*: 835–846.

Durlauf, Steven N. (2001). A Framework For The Study of Individual Behavior and Social Interactions. *Sociological Methodology*, *31(1)*: 47–87.

Erdős, Paul and Rényi, Alfred. (1959). On random graphs. *Publicationes Mathematicae*, *6*: 290–297.

Girvan, Michelle and Newman, Mark E. J. (2002). Community structure in social and biological networks. *PNAS*, *99*: 7821–7826.

Guillaume, Jean-Loup and Latapy, Matthieu. (2004). Bipartite structure of all complex networks. *Information Processing Letters*, *90(5)*: 215–221.

Haas, P. (1992). Introduction: epistemic communities and international policy coordination. *International Organization*, *46(1)*: 1–35.

Kitcher, Philip. (1995). Contrasting Conceptions of Social Epistemology. *Pages 111–134 of:* Schmitt, F. (ed), *Socializing Epistemology: The Social Dimensions of Knowledge*. Lanham, MD: Rowman and Littlefield.

Lloyd, Alun L. and May, Robert M. (2001). How Viruses Spread Among Computers and People. *Science*, *292(5520)*: 1316–1317.

Lorrain, F. and White, H. C. (1971). Structural Equivalence of Individuals in Social Networks. *Journal of Mathematical Sociology*, *1(49–80)*.

Morris, Stephen. (2000). Contagion. *Review of Economic Studies*, *67(1)*: 57–78.

Newman, Mark E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45(2)*: 167–256.

Newman, Mark E. J., Strogatz, S. and Watts, D. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, *64(026118)*.

Pastor-Satorras, Romualdo and Vespignani, Alessandro. (2001). Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, *86(14)*: 3200–3203.

Pattison, Philippa, Wasserman, Stanley, Robins, Garry and Kanfer, Alaina Michaelson. (2000). Statistical Evaluation of Algebraic Constraints for Social Networks. *Journal of Mathematical Psychology*, *44*: 536–568.

Powell, Walter W., White, Douglas R., Koput, Kenneth W. and Owen-Smith, Jason. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, *110(4)*: 1132–1205.

Robins, Garry and Alexander, Malcolm. (2004). Small Worlds Among Interlocking Directors: Network Structure and Distance in Bipartite Graphs. *Computational and Mathematical Organization Theory*, *10*: 69–94.

Roth, Camille. (2007). Empiricism for descriptive social network models. *Physica A*, *378(1)*: 53–58.

Roth, Camille and Bourgine, Paul. (2006). Lattice-based dynamic and overlapping taxonomies: The case of epistemic communities. *Scientometrics*, *69(2)*: 429–447.

Skyrms, Brian and Pemantle, Robin. (2000). A dynamic model of social network formation. *PNAS*, *97(16)*: 9340–9346.

Watts, Duncan J. and Strogatz, Steven H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*: 440–442.