Binding Social and Semantic Networks

Camille Roth*

Proceedings of ECCS 2nd European Conference on Complex Systems (09/2006, Oxford UK)

Abstract

Social and semantic networks have often been studied separately. We provide a theoretical framework to bind these two networks, suggesting that the analysis of knowledge community structure and underlying agent-based dynamics requires to take into account the reciprocal influence of both networks. We show how to characterize meaningful cultural communities using Galois lattices and briefly explain how models could render the coevolution of socio-semantic networks, building upon a generalized understanding of preferential attachment. This should enable the comprehension of stylized facts proper to knowledge networks, in particular how social and semantic structures jointly affect each other.

Keywords: social complex systems, network dynamics, Galois lattice or concept lattice, applied epistemology, community representation, knowledge diffusion, social cognition.

Introduction

Much attention has recently been given to real-world networks, considering them as complex systems in order to explain their formation and dynamics [1, 36]. Models have been developed that provide compelling insight and understanding of the topological properties of these networks, including in particular node degrees and the broadly shared *scale-free* property [2, 53] as well as, to cite a few, mean distance (shortest path length), largest connex component size (giant component), assortative mixing, existence of cycles, number of second neighbors, and one-mode community structure [9, 10, 19, 34, 38, 54].

In general social networks have rarely been treated in a different way than other real networks, while they exhibit particular features regarding for instance correlations in degrees of adjacent vertices [35] or clustering structure [55] (namely, the propensity of two agents to be connected together if they have common acquaintances). Often, agents are considered to behave in a way not more complex than molecules; even when taking into account the the behavioral complexity of agents, social network models do not seem to focus on the relationships between semantic and social features. Yet in order, for instance, to model the way beliefs propagate among social networks of agents, that is, explain how the social network structure affects concept propagation and in return how concept propagation affects

*Department of Social and Cognitive Science, University of Modena & Reggio Emilia, Via Allegri 9, I-42100 Reggio Emilia, Italy; *and* CREA (Center for Research in Applied Epistemology), CNRS/Ecole Polytechnique, 1, rue Descartes, F-75005 Paris, France. E-mail: *camille.roth@polytechnique.edu*
The author wishes to thank Paul Bourgine for very fruitful discussions. This work has been partially funded by the CNRS and the University of Modena.

the social network, one must be able to give an account of how knowledge networks form and evolve. Here, we need to consider arguments stemming from social psychology: attraction for same-profile people ("homophily") is key in the formation of social acquaintances [31], hence full attention should be given to the influence and evolution of semantic features.

We introduce here a network dual to the social network, the network of semantic items, denoted as concepts. We will focus on the study of communities of scientists, which has been made a realistic objective through the massive availability of electronic and free data. Apart from properties relative to the social network (such as node degrees) we can assume that one of the dominant criteria for choosing a scientific partner mostly depends on the cultural similarity of two agents. Some economic models of knowledge creation already take agent profiles into consideration, as elements of a vector space, to explain the structure of the economic network — in [12], agents match two by two to produce new knowledge according to their profile. Our goal is to bring a formal framework for studying the intertwining of social and semantic networks, both theoretically and empirically, as well as to point out stylized facts that would explain their reciprocal influence and the emergence of "cultural cliques" of agents. Since understanding knowledge network structure is a key step in this process, we first show how communities should be appraised when binding social and semantic networks. We consequently present some implications on how network dynamics should be considered and modeled. At the same time, we sketch out and describe a few potential empirical applications. In a broader view, this framework constitutes a step towards actually implementing the paradigm of cultural epidemiology [13, 50], therefore enabling us to proceed further with the study of knowledge diffusion.

1 Networks

In this section we present the social and semantic networks and links between and within them. We exemplify these structures by a socio-semantic complex system of scientists collaborating to produce articles.

Definition 1 (Social network)**.** *The social network $\mathcal{S}$ is represented by the network of coauthorship, where nodes are authors and links are collaborations. Thus $\mathcal{S} = (\mathbf{S}, \lambda_{\mathbf{S}})$, where $\mathbf{S}$ denotes the set of authors and $\lambda_{\mathbf{S}}$ denotes the set of undirected links.*

As time evolves, new articles are published, new nodes are possibly added to $\mathbf{S}$ and new links are classically created between each pair of co-authors [33]. We actually consider the temporal series of networks $\mathcal{S}(t)$ with $t \in \mathbb{N}$ (articles are published with a date, thus an integer) to observe the dynamics of the network (in the remainder of the paper we omit the reference to t, $\mathcal{S}$ depending implicitly on time). Depending on model goals and desired precision, we may want to take into account the fact that, for instance, two nodes have co-authored more than one paper (thus introducing *link strength*), or that their collaborations are more or less recent (thus introducing *link age* [22, 41]). Relationships should consequently be different according to whether agents have collaborated only once and a long time ago, or they have recently co-authored many articles. An easy and practical way for dealing with these notions is to use a weighted network. In a *non-weighted network*, we say that two nodes are linked as soon as there exists one coauthored article. Links can only be active *or* inactive. In a *weighted network*, links are provided with a weight $w \in \mathbb{R}^+$, possibly evolving in time. We can therefore easily represent multiple collaborations by increasing the weight of a link, or render the age of a relationship by decreasing this weight (for instance by applying an aging function). This method enables us to model a non-weighted network, by assigning weights of 1 or 0 respectively to active or inactive links, while leaving room for the *ex post* creation of a non-weighted network from a weighted network, by setting a threshold such that a link is active when its weight exceeds the threshold — otherwise inactive.

The semantic network is very similar to the social network, in a dual manner:

Definition 2 (Semantic network)**.** *The semantic (or conceptual) network $\mathcal{C}$ is the network of joint appearances of concepts within articles, where nodes are concepts and links are co-occurrences. $\mathcal{C} = (\mathbf{C}, \lambda_{\mathbf{C}})$, similarly to $\mathcal{S}$.*

When a new article appears, new concepts are possibly added to the network, and new links are added between co-appearing concepts. Here again, as in the case of the social network, one can use a weighted network to render various strength on co-occurrences. However, the whole point is to define precisely what a *concept* should be. Is it a paradigm like *"universal gravitation"*, a scientific field like *"molecular biology"*, or a simple word like *"interferon"*? In particular, what is a concept such that we can observe its appearance in articles? This notion needs be not too precise nor too wide. For instance, authors provide their articles with keywords: apparently, considering these keywords as concepts seems to constitute a relevant level of categorization while being a convenient idea. However, such keywords are likely to be unreliable for authors often arbitrarily omit important keywords, specify less relevant ones, etc.

In the field of scientometrics for instance, as a discipline dealing with large databases of scientific articles, words appearing in articles are more and more used as indicators of the topics raised by authors [30, 37, 47] — that is, the semantics is extracted from article contents rather than metadata. At first we should thus say that *each term is a concept*. This definition does not prevent us from observing higher-level concepts such as scientific fields or paradigms, since we can easily refer to these concepts *a posteriori* by considering sets of strongly connected terms. For example, we could interpret the set of frequently co-occuring words {*"cell"*, *"DNA"*, *"gene"*, *"genetic"*, *"genetics"*, *"molecular"*} as *molecular biology*. Moreover, we could proceed only with words present in what we consider to be the most relevant article data: the title and the abstract — setting aside article content, for it is rarely available and also because it could involve too many very precise though unrelevant words. Of course, we should also define a list of words to be ignored, or *"stop words"*, including grammatical and unsignificant words (*"is"*, *"with"*, *"study"*, etc.) as well as non-discriminating words (e.g., *"biology"* within a community of biologists).

Binding the two networks As the social network is the network of joint appearances of authors, so is the semantic network with concepts, establishing an obvious duality between the two networks which is key to bind them and explain their reciprocal influence. In the same manner as we did with the previous networks, we link scientists to the words they use, i.e. we add a link whenever an author and a word co-appear within an article. Hence considering the two networks $\mathcal{S}$ and $\mathcal{C}$, we deal with three kinds of quite similar links: (i) between pairs of scientists, (ii) between pairs of concepts, and (iii) between concepts and scientists; thus setting up three kinds of binary relations:

(i) a set of symmetrical relations $\mathcal{R}^{\mathbf{S}}_{\alpha} \subset \mathbf{S} \times \mathbf{S}$ from the social network to the social network, and such that given $\alpha \in \mathbb{R}$ and two scientists s and s', we have $s\ \mathcal{R}^{\mathbf{S}}_{\alpha}\ s'$ *iff* the link between s and s' has a weight w strictly greater than the threshold α.

(ii) a set of symmetrical relations $\mathcal{R}^{\mathbf{C}}_{\alpha} \subset \mathbf{C} \times \mathbf{C}$ from the semantic network to the semantic network, and such that given $\alpha \in \mathbb{R}$ and two concepts c and c', $c\ \mathcal{R}^{\mathbf{C}}_{\alpha}\ c'$ *iff* the link between c and c' has a weight $w > \alpha$.

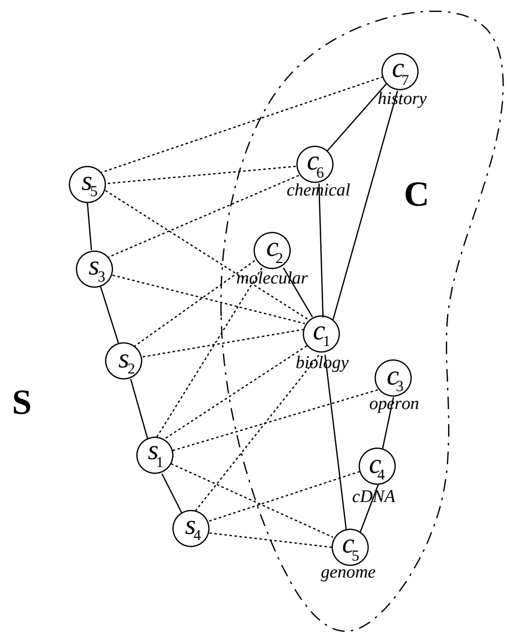

Figure 1: Sample network with $\mathbf{S} = \{s_1, s_2, s_3, s_4, s_5\}$, $\mathbf{C} = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\} = \{$biology, molecular, operon, cDNA, genome, chemical, history$\}$, and relations $\mathcal{R}^{\mathbf{S}}$ and $\mathcal{R}^{\mathbf{C}}$ (solid lines) and $\mathcal{R}$ (dashed lines).

(iii) a binary relation $\mathcal{R}_\alpha \subset \mathbf{S} \times \mathbf{C}$ from the social network to the semantic network, and such that given $\alpha \in \mathbb{R}$, an author s and concept c, $s\ \mathcal{R}_\alpha\ c$ *iff* the link between s and c has a weight $w > \alpha$.

With the special case $\alpha = 0$, noticing that $\alpha < \alpha' \Rightarrow \mathcal{R}^{(.)}_{\alpha'} \subset \mathcal{R}^{(.)}_{\alpha}$, then $\forall \alpha > 0, \mathcal{R}^{(.)}_{\alpha} \subset \mathcal{R}^{(.)}_{0}$: the relations $\mathcal{R}^{(.)}_{0}$ are maximal, i.e. two nodes are related whenever there exists a link binding them, whatever its weight. To ease the notation, we identify $\mathcal{R}^{\mathbf{S}}_0$ to $\mathcal{R}^{\mathbf{S}}$, $\mathcal{R}^{\mathbf{C}}_0$ to $\mathcal{R}^{\mathbf{C}}$, and $\mathcal{R}_0$ to $\mathcal{R}$.

2 Communities in epistemic networks

With these basic structures defined, we now need formal tools to formulate stylized facts about knowledge and people conveying it, notably by looking high-level patterns such as knowledge communities. As socio-semantic networks are primarily made of *dual-mode* data, we should indeed exhibit community structures that make use of this duality, instead of relying on one-mode projection which imply loss of crucial structural information. In particular, structurally equivalent [28] groups of agents using the same concepts constitute relevant dual-mode communities, as *epistemic communities*. To describe these facts, among other two-mode network data methods [14], Galois lattices appear as a suitable framework for agent-concept categorization — being also widely used as well in conceptual knowledge systems [57] and formal concept classification [20]. White & Freeman have already explored an application of this theory in mathematical sociology, grouping simultaneously agents and social events they attend [18].

The goal of this section is to present the Galois lattice theory and show how we can use it to describe efficiently knowledge community structure from relationships between $\mathcal{S}$ and $\mathcal{C}$. More broadly, we wish to suggest that community structure in knowledge-based social networks should be dealt with more deeply than by simply relying on single-mode characterizations [5].

2.1 Sets and relations

Let us first consider two finite sets A and B between which we have a binary relation $R \subseteq A \times B$. We introduce the operation "$\wedge$" such that for any element $x \in A$, $x^\wedge$ is the set of B elements R-related to x. Extending this definition to subsets $X \subseteq A$, we denote by $X^\wedge$ the set of B elements R-related to every element of X, namely:

$$x^\wedge = \{\, y \in B \mid xRy \,\} \qquad X^\wedge = \{\, y \in B \mid \forall x \in X, xRy \,\} \tag{1}$$

Similarly, "$\star$" is the dual operation so that $\forall y \in B, \forall Y \subseteq B$,

$$y^\star = \{\, x \in A \mid xRy \,\} \qquad Y^\star = \{\, x \in A \mid \forall y \in Y, xRy \,\} \tag{2}$$

By definition we set $(\emptyset)^\wedge = B$ and $(\emptyset)^\star = A$. These operations enjoy the following properties:

$$X \subseteq X' \Rightarrow X'^\wedge \subseteq X^\wedge \quad \text{(3a)} \qquad X \subseteq X^{\wedge\star} \quad \text{(4a)}$$
$$Y \subseteq Y' \Rightarrow Y'^\star \subseteq Y^\star \quad \text{(3b)} \qquad Y \subseteq Y^{\star\wedge} \quad \text{(4b)}$$

We also have:

$$(X \cup X')^\wedge = X^\wedge \cap X'^\wedge \qquad (Y \cup Y')^\star = Y^\star \cap Y'^\star \tag{5}$$

Accordingly $X^\wedge = (\bigcup_{x\in X}\{x\})^\wedge = \bigcap_{x \in X} x^\wedge$.

Closure operation More importantly, the following property holds true,[1]

$$((X^\wedge)^\star)^\wedge = X^\wedge \text{ and } ((Y^\star)^\wedge)^\star = Y^\star \tag{6}$$

and therefore we can define the operation "$\wedge\star$" as a *closure operation* [7], in that it is:

$$\text{extensive,} \qquad X \subseteq X^{\wedge\star} \quad \text{(7a)}$$
$$\text{idempotent} \qquad (X^{\wedge\star})^{\wedge\star} = X^{\wedge\star} \quad \text{(7b)}$$
$$\text{and increasing.} \qquad X \subseteq X' \Rightarrow X^{\wedge\star} \subseteq X'^{\wedge\star} \quad \text{(7c)}$$

We say that X is a *closed* subset if $X^{\wedge\star} = X$.

2.2 Galois lattices

We now consider the set of couples of subsets of A and B and build a new structure onto it: given two subsets $X \subseteq A$ and $Y \subseteq B$, a couple (X, Y) is said to be *closed iff* $Y = X^\wedge$ and $X = Y^\star$. Yet such a couple is actually a $(X, X^\wedge)$ where $X^{\wedge\star} = X$. Therefore, closed couples correspond obviously to couples of subsets of A and B closed under $\wedge\star$. This will allow us to define a new kind of lattice from A, B and R. We first recall the algebraic definition of a *lattice*:

[1]Indeed, (3a) applied to (4a) leads to $(X^{\wedge\star})^\wedge \subseteq X^\wedge$, while (4b) applied to $X^\wedge$ gives $(X^\wedge) \subseteq (X^\wedge)^{\star\wedge}$.

Definition 3 (Lattice)**.** *A set* $(L, \sqsubseteq, \sqcup, \sqcap)$ *is a* lattice *if every finite subset* $H \subseteq L$ *has a* least upper bound *in* L *noted* $\sqcup H$ *and a* greatest lower bound *in* L *noted* $\sqcap H$ *under partial-ordering relation* $\sqsubseteq$*.*

In this respect the set of subsets of a set X provided with the usual inclusion, union and intersection, $(\mathcal{P}(X), \subseteq, \cup, \cap)$, is a lattice — and so is a *Galois lattice* [4]:

Definition 4 (Galois lattice)**.** *Given a relation* R *between two finite sets* A *and* B*, the* Galois lattice $\mathcal{G}_{A,B,R}$ *is the set of every* closed *couple* $(X, Y) \subseteq A \times B$ *under relation* R*:* $\mathcal{G}_{A,B,R} = \{(X^{\wedge\star}, X^{\wedge}) | X \subseteq A\}$.

$\mathcal{G}_{A,B,R}$ provided with the natural partial order $\sqsubseteq$ such that $(X, X^{\wedge}) \sqsubseteq (X', X'^{\wedge}) \Leftrightarrow X \subset X'$ is indeed a lattice. As Wille points out [57], this structure constitutes a solid formalization of the philosophical appraisal of a concept characterized by its *extension* (the physical implementation or the group of things denoted by the concept) and its *intension* (the properties or the internal content of the concept). In a pair $g = (X, X^{\wedge})$ considered as a formal concept, X may be seen as the extension of g while $X^{\wedge}$ is its intension. For a given $X \subseteq A$, $X^{\wedge}$ will represent the set of properties shared by all objects of X, whereas for a given set of properties $Y \subseteq B$, $Y^{\star}$ will be the set of objects of A actually fulfilling them. Also, using the strict partial order $\sqsubset$, we can talk of *formal subconcept* by saying g *is a subconcept of* g' *iff* $g \sqsubset g'$. Hence g can be seen as a specification of g', since the number of its properties increases ($X^{\wedge} \supset X'^{\wedge}$, thus defining g more precisely) while less objects belongs to its extension ($X \subset X'$). Conversely, g' is a *"superconcept"* or a generalization of g; we have thus a tool of generalization and specification of formal concepts [56].

2.3 Applying lattices to $\mathcal{S}$ and $\mathcal{C}$

To apply these tools to our networks $\mathcal{S}$ and $\mathcal{C}$, we consider the two finite sets $\mathbf{S}$, $\mathbf{C}$, the relation $\mathcal{R}$ and $\mathcal{G}_{\mathbf{S},\mathbf{C},\mathcal{R}}$. First, for an author $s \in \mathbf{S}$, $s^{\wedge} = \{ c \mid s\mathcal{R}c \}$ represents the set of the concepts he talked about or the fields he dealt with. Proceeding identically with a concept $c \in \mathbf{C}$, $c^{\star} = \{ s \mid s\mathcal{R}c \}$ represents the set of scientists who used the concept c in at least one of their papers. Then, for a group of authors $S \subseteq \mathbf{S}$, $S^{\wedge}$ represents the words being used by every author $s \in S$, while for a set of words $C \subseteq \mathbf{C}$, $C^{\star}$ is the set of agents using every concept $c \in C$. Moreover, we can easily derive from (5) the words used by a community $S \cup S'$ by taking the intersection $S^{\wedge} \cap S'^{\wedge}$, or the authors corresponding to the merger of any two sets of concepts $C \cup C'$ by taking $C^{\star} \cap C'^{\star}$.

An example is shown on figure 1. For instance, ${s_4}^{\wedge} = \{c_1, c_4, c_5\}$ and $\{c_1, c_6\}^{\star} = \{s_3, s_5\}$. If we consider the matrix R representing relation $\mathcal{R}$ as follows,

$$R = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

where $R_{i,j}$ is non-zero when $s_i \; \mathcal{R} \; c_j$, we can easily read ${s_i}^{\wedge}$ on rows and ${c_j}^{\star}$ on columns.

Closure and epistemic categories Seeing concepts as *properties* of authors who use them (skills in scientific fields as cognitive properties) and authors as *extensions* of concepts (implementation of concepts within authors), one can make a very fertile usage of the lattice $\mathcal{G}_{\mathbf{S},\mathbf{C},\mathcal{R}}$ by setting up an epistemic taxonomy with the help of formal concepts made of couples (S, C) with $S \subseteq \mathbf{S}, C \subseteq \mathbf{C}$.

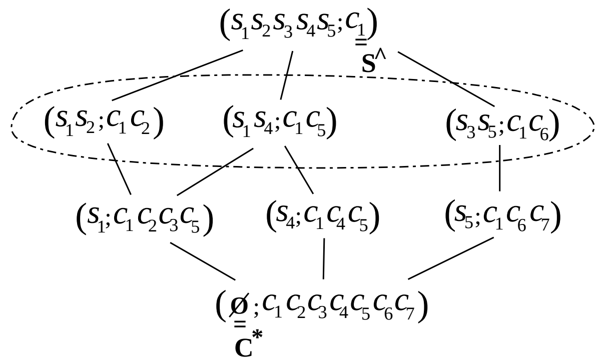

Figure 2: Representation of the whole Galois lattice of our example – the hierarchy is drawn according to the partial order $\sqsubset$, i.e. "bottom"$\sqsubset$"top". The cultural background $\mathbf{S}^{\wedge}$ is reduced to "biology". On the medium-level, we find formal concepts $(s_1, s_2$; "biology" , "molecular"), $(s_1, s_4$; "biology" , "genome"), $(s_3, s_5$; "biology" , "chemical").

We may indeed consider such formal concepts as *schools of thought* constituted by the community of agents S working and writing on the field C, a formal subconcept simply being a trend inside a school. By community we understand henceforth *epistemic community*, that is to say neither a department nor a group of research.

In addition, we recall that for such a closed couple from the Galois lattice, $C = S^{\wedge}$, $S = C^{\star}$ and finally $S = S^{\wedge\star}$. $S^{\wedge\star}$ actually represents the set of scientists using *at least* the same words as S, "$\wedge\star$" being a closure operation, $S^{\wedge\star}$ *closes* the set S by returning all the scientists related to every concept shared among S — once and for all from (7b).[2] Admittedly, for a single scientist s, $s^{\wedge\star}$ will certainly be equal to s, since there are strong chances that $\forall s' \in \mathbf{S}$, $\exists w \in s^{\wedge}$ and $\notin s'^{\wedge}$. Considering however a subset $S \subseteq \mathbf{S}$, as its cardinal increases there are more and more chances that the closure of S reaches an actual community of researchers. We conjecture that there is a relevant level of closure for which a set $S^{\wedge\star}$, and identically $C^{\star\wedge}$, is representative of a field or a trend. This idea is to be compared to Rosch's basic-level of categorization [42]. This medium level shall constitute a basic-level of epistemic categorization, whereas above it ("superordinate categories") the field would be too general, and too precise under it ("subordinate categories").[3]

Comparing dual-mode and single-mode communities Another point of interest is to see whether single network communities (based on the social or semantic network only) correspond to closed sets, i.e. whether a $\mathcal{S}$-community is also a $\wedge\star$-community, and whether a $\mathcal{C}$-community is also a $\star\wedge$-

[2]Note that given $S^{\wedge} = \{c_1, ..., c_n, c\}$ and $S'^{\wedge} = \{c_1, ..., c_n, c'\}$, we have $S' \notin S^{\wedge\star}$, S' not being in the closure of S, which might look quite strange as their domains of interest are similar. Yet, $(S \cup S')^{\wedge} = S^{\wedge} \cup S'^{\wedge} = \{c_1, ..., c_n\}$, thus $\{c_1, ..., c_n\}$ defines a (larger) epistemic community which includes both S and S'.
Another property may help understand better what the closure actually does: given $S^{\wedge} = \{c_1, ..., c_n\}$ and $S'^{\wedge} = \{c'_1, ..., c'_n\}$ such that $\forall (i,j) \in \{1, ..., n\}^2, c_i \neq c'_j$, we have $(S \cup S')^{\wedge\star} = \mathbf{S}$: the closure of the union of two sets of scientists working on totally different issues is the whole community $\mathbf{S}$ — "there is no way to distinguish S and S' from each other with respect to the rest of the community".

[3]Interestingly, $\mathbf{S}^{\wedge}$ also represents the concepts the whole community shares — the *"background"*, obviously too common to be discriminating. This set could actually constitute an appropriate companion to the list of stop words mentioned in §1. On the other hand, $\mathbf{C}^{\star}$ should not enjoy in general any such property and should be empty: the contrary would mean that there is at least one author linked to *all* concepts in use among the whole community.

community. In other words, are schools of thought and scientific fields also socially and semantically strongly linked or not? We might also want to compare different definitions of a one-mode community, using for instance criteria such as k*-connectivity*[4] [39] or various clustering algorithms [19].

2.4 Empirical application

A detailed empirical case study, beyond the toy example we presented here, would take too much room in this paper. Interested readers may nonetheless find a study based on this approach for a real-world community of embryologists working on the "zebrafish", between 1990 and 2003, both for a static case [45] and for a longitudinal study [46] — the taxonomic evolution of this scientific field given through Galois lattices has been successfully compared to taxonomies given by domain experts. In addition, the conjecture of the existence of a relevant basic-level of categorization, linked to a particular distribution of agent set sizes of epistemic communities, is also empirically confirmed.

3 Network dynamics

3.1 Intertwining both networks

A model should first try to account for the network evolution by building upon existing models while including the improvements offered by the present theoretical framework. For instance, the pioneering growing-network model proposed by Barabasi & Albert [3] and many subsequent models [6, 10, 11, 15, 16, 17, 21, 23, 25, 29, 40, 51], are traditionally directed by two key phenomenas: (i) a *constant rate of growth* (the number of nodes at any time t is αt), justified by the fact that real networks "grow by the continuous addition of new nodes" [1]; and (ii) *a preferential attachment* — external (new nodes join the system) as well as internal (links appearing between existing nodes). The preference between agents is usually based on the preference for already well-connected agent (as more recognized, famous, reliable or simply efficient), the number of common neighbors, a notion of fitness, the value of centrality or distance, etc.

Yet agents also interact according to preferences based on non-structural features and usually prefer to interact with similar agents. Using semantic features is required to characterize this phenomenon, which is traditionally denoted by the term "homophily". While it is already well-documented in social science [26, 52, 31], very few models make use of it [8, 49] and none seems to take into account the evolution of semantic characteristics themselves. If we assume yet that homophily is essential to the system dynamics, the preferential attachment must be modified in order to take into account similarity between agents or between concepts: nodes will indeed join preferentially more connected but also more similar nodes. Thus, the preferential attachment probability of a node to another node Π should depend *inter alia*, for a given scientist $s \in \mathbf{S}$, both (i) on the degree of other scientists $s' \in \mathcal{S}$ (using $\mathcal{R}^{\mathbf{S}}$) and (ii) on the *"distance"* between $s^\wedge$ and $s'^\wedge$ – or *dual distance* between s and s' (using $\mathcal{R}^{\mathbf{C}}$ and $\mathcal{R}$). We need not focus on a particular definition for this distance, as long as it decreases with the number of shared concepts and that it equals 0 for identical intensions and 1 for disjoint intensions — it could be based on the classical Jaccard coefficient for example, i.e. $(s, s') \in \mathbf{S}^2$, $d(s, s') \in [0; 1] = \frac{|(s^\wedge \setminus s'^\wedge) \cup (s'^\wedge \setminus s^\wedge)|}{|s^\wedge \cup s'^\wedge|}$; which fulfills the above requirements.

Empirically, a different behavior should be expected with respect to such a parameter d: if communities do exist, preferential attachment is indeed likely to depend on a parameter which favors

[4] The smallest number of nodes to withdraw from a connected (sub)graph to get a disconnected one.

the reinforcement of similar agents.[5] This has been shown to be the case in [44], with a stronger propension of interaction between agents sharing more concepts.

3.2 Implementing dual-mode network dynamics

More generally, specifying the list of properties is nevertheless a process driven by the real-world situation *and* by the stylized facts the modeler aims at rebuilding and considers relevant for the morphogenesis of both networks. While we only discussed the example of two significant properties (node degree and semantic distance), measuring preferential interaction behavior relatively to other parameters could be very relevant as well — including social distance, common acquaintances, etc. The goal is also to exhibit credible as well as non-overlapping, non-correlated properties.

In any case, a first step would be to determine empirically the shape of Π so that we can infer fertile intuitions for designing an analytical value for Π which could be introduced back into a model. Note that, although not detailed here, the reasoning holds the same for the preferential attachment in $\mathbf{C}$. As McPherson & Smith-Lovin suggest [31], homophily has been for long understood as a very general setting. Consequently, several authors have already carried the empirical measurement of this phenomenon [24, 27, 39, 44, 48] — on the side of methods, it is possible to design a credible and realistic preferential attachment behavior in evolving social graphs using methods of [32] extended in [44], for any kind of property. Hence, extending these notions to a dual-network framework is rather straightforward.

In turn, enriching the low-level behavior from a strict social network dynamics to a co-evolving socio-semantic network is definitely within reach, as is carried e.g. in [43]: eventually, reconstruction of stylized facts *relevant for epistemic communities* cannot be done without considering the morphogenesis of the whole socio-semantic structure.

Conclusion

Most studies carried onto social networks or semantic networks have considered each of these networks independently. We proposed here a framework for binding them and pointing out their very duality as well as expressing stylized facts about them. The Galois lattice theory has proved useful in helping introduce key notions such as schools of thought through closure, basic-level of categorization of a scientific field, and in general for characterizing scientific communities. As such, we defined a high-level structure (epistemic communities) from low-level descriptions (relationships between agents and concepts) specifically using the duality of socio-semantic networks. Next we showed how to apply this framework to model the *coevolution* of social and cultural networks, suggesting that low-level dynamics should reflect empirically measured generalized preferential interaction behavior. Instead of considering social and semantic networks separately, we suggested that interactions should take into account the reciprocal influence of both networks — therefore introducing, for instance, the notion of dual distance.

More than providing a theoretical framework, we intend to enable the comprehension of stylized facts proper to knowledge networks that distinguish them from several other classes of real-world networks. Going further in this effort of pointing out some applications of such a dual framework in

[5]Newman [32] for instance considers the number of common acquaintances as an explanatory argument for clique formation; instead, we may assume that collaborations do essentially occur on account of homophily, while this assumption does not contradict structural arguments such as those based on common acquaintances: two agents are all the more likely to have the same profile that they share many acquaintances.

the observation, description and eventually modeling of epistemic network dynamics, this paper is a preliminary attempt at naturalizing and operationalizing cultural epidemiology — describing and explaining propagation of concepts through the social network *as well as* the deformation of the scientist network by the joint structure of the semantic network; in other words, study these networks as an integrated epistemic network.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[2] A.-L. Barabási, Z. Dezso, Z. Oltvai, E. Ravasz, and S.-H. Yook. Scale-free and hierarchical structures in complex networks. In *Sitges Proceedings on Complex Networks*, 2004.

[3] 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.

[4] M. Barbut and B. Monjardet. *Algèbre et Combinatoire*, volume II. Hachette, Paris, 1970.

[5] V. Batagelj, A. Ferligoj, and P. Doreian. Generalized blockmodeling of two-mode networks. *Social Networks*, 26(1):29–54, 2004.

[6] N. Berger, C. Borgs, J. Chayes, R. D'Souza, and R. Kleinberg. Competition-induced preferential attachment. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming*, pages 208–221, 2004.

[7] G. Birkhoff. *Lattice Theory*. Providence, RI: American Mathematical Society, 1948.

[8] M. Boguna and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68:036112, 2003.

[9] M. Boguna, R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas. Models of social networks based on social distance attachment. *Physical Review E*, 70:056122, 2004.

[10] G. Caldarelli, A. Capocci, P. D. L. Rios, and M. A. Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89(25):258702, 2002.

[11] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo. Network structures from selection principles. *Physical Review Letters*, 92(19):198701, 2004.

[12] R. Cowan, N. Jonard, and J.-B. Zimmermann. The joint dynamics of networks and knowledge. Computing in Economics and Finance 2002 354, Society for Computational Economics, July 2002.

[13] R. Dawkins. *The Selfish Gene*, chapter 11: Memes, The New Replicator. Oxford: Oxford University Press, 1976.

[14] P. Doreian, V. Bategelj, and A. Ferligoj. *Generalized Blockmodelling*. Cambridge: Cambridge University Press, 2005.

[15] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks with aging of sites. *Physical Review E*, 62:1842–1845, 2000.

[16] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633–4636, 2000.

[17] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *ICALP '02: Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, pages 110–122, London, UK, 2002. Springer-Verlag.

[18] L. C. Freeman and D. R. White. Using Galois lattices to represent network data. *Sociological Methodology*, 23:127–146, 1993.

[19] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *PNAS*, 99:7821–7826, 2002.

[20] R. Godin, G. Mineau, R. Missaoui, and H. Mili. Méthodes de classification conceptuelle basées sur les treillis de Galois et applications. *Revue d'Intelligence Artificielle*, 9(2):105–137, 1995.

[21] 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.

[22] B. Huberman and L. Adamic. Growth dynamics of the World-Wide Web. *Nature*, 399:130, 1999.

[23] E. M. Jin, M. Girvan, and M. E. J. Newman. The structure of growing social networks. *Physical Review E*, 64(4):046132, 2001.

[24] G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311:88–90, 2006.

[25] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85:4629–4632, 2000.

[26] P. F. Lazarsfeld and R. K. Merton. Friendship as a social process: a substantive and methodological analysis. In M. Berger, editor, *Freedom and Control in Modern Society*, pages 18–66. New York: Van Nostrand, 1954.

[27] E. Lazega and M. van Duijn. Position in formal structure, personal characteristics and choices of advisors in a law firm: a logistic regression model for dyadic network data. *Social Networks*, 19:375–397, 1997.

[28] F. Lorrain and H. C. White. Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1(49–80), 1971.

[29] S. S. Manna and P. Sen. Modulated scale-free network in euclidean space. *Physical Review E*, 66:066114, 2002.

[30] K. W. McCain, J. M. Verner, G. W. Hislop, W. Evanco, and V. Cole. The use of bibliometric and Knowledge Elicitation techniques to map a knowledge domain: Software Engineering in the 1990s. *Scientometrics*, 65(1):131–144, 2005.

[31] M. McPherson and L. Smith-Lovin. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–440, 2001.

[32] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review Letters E*, 64(025102), 2001.

[33] M. E. J. Newman. The structure of scientific collaboration networks. *PNAS*, 98(2):404–409, 2001.

[34] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89:208701, 2002.

[35] M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(036122), 2003.

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

[37] E. C. M. Noyons and A. F. J. van Raan. Monitoring scientific developments from a dynamic perspective: self-organized structuring to map neural network research. *Journal of the American Society for Information Science*, 49(1):68–81, 1998.

[38] P. Pattison, S. Wasserman, G. Robins, and A. M. Kanfer. Statistical evaluation of algebraic constraints for social networks. *Journal of Mathematical Psychology*, 44:536–568, 2000.

[39] 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.

[40] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras. Self-organization of collaboration networks. *Physical Review E*, 70:036106, 2004.

[41] S. Redner. How popular is your paper? An empirical study of the citation distribution. *European Phys. Journal B*, 4(131–134), 1998.

[42] E. Rosch and B. Lloyd. Cognition and categorization. *American Psychologist*, 44(12):1468–1481, 1978.

[43] C. Roth. *Co-evolution in Epistemic Networks*. PhD thesis, Ecole Polytechnique, Paris, France, 2005.

[44] C. Roth. Generalized preferential attachment: Towards realistic social network models. In *ISWC 4th Intl Semantic Web Conference, Workshop on Semantic Network Analysis*, Galway, Ireland, 2005.

[45] C. Roth and P. Bourgine. Epistemic communities: Description and hierarchic categorization. *Mathematical Population Studies*, 12(2):107–130, 2005.

[46] C. Roth and P. Bourgine. Lattice-based dynamic and overlapping taxonomies: the case of epistemic communities. *Scientometrics*, 69(2), 2006.

[47] G. Salton, A. Wong, and C. S. Yang. Vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

[48] T. A. Snijders. The statistical evaluation of social networks dynamics. *Sociological Methodology*, 31:361–395, 2001.

[49] B. Söderberg. A general formalism for inhomogeneous random graphs. *Physical Review E*, 68:026107, 2003.

[50] D. Sperber. *La contagion des idées*. Paris: Odile Jacob, 1996.

[51] H. Stefancic and V. Zlatic. Preferential attachment with information filtering–node degree probability distribution properties. *Physica A*, 350(2-4):657–670, 2005.

[52] J. C. Touhey. Situated identities, attitude similarity, and interpersonal attraction. *Sociometry*, 37:363–374, 1974.

[53] A. Vázquez. Disordered networks generated by recursive searches. *Europhysics Letters*, 54(4):430–435, 2001.

[54] D. J. Watts, P. S. Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

[55] D. J. Watts and S. H. Strogatz. Collective dynamics of “small-world” networks. *Nature*, 393:440–442, 1998.

[56] R. Wille. Concept lattices and conceptual knowledge systems. *Computers Mathematics and Applications*, 23:493, 1992.

[57] R. Wille. Conceptual graphs and formal concept analysis. In *Proceedings of the fourth International Conference on Conceptual Structures*, number #1257 in Lecture Notes on Computer Science, pages 290–303. Berlin: Springer, 1997.