

Knowledge Communities and Socio-Cognitive Taxonomies

Camille Roth

1 Introduction

A significant portion of the state of the art in social network analysis (SNA) has been typically devoted to the characterization of groups of *similar* actors [68]—oftentimes denoted as *communities*. While similarities and communities can be defined in very diverse ways [6, 23], the literature can roughly be divided into two main classes of approaches, depending on whether groups stem from cohesive interactions or from cohesive affiliations. This dichotomy in turn refers to two distinct types of graphs and relations. On one hand, a large number of methods rely on *interaction networks*, which are purely social networks insofar as nodes are strictly actors and links indicate actor–actor relationships: actors know each other, they work with each other, they talk to each other, etc. Formally, this corresponds to monopartite graphs or the so-called one-mode networks. On the other hand, SNA scholars have made use of *affiliation networks* [68, Chap. 8] where nodes may be of two types: either actors or social attributes of some sort—be it an event, an organization, a team, an issue, an opinion, an interest, etc. Here, relationships denote the affiliation, in the broad sense, of an actor to an attribute; formalisms are based on bipartite graphs, or two-mode networks.

Approaches based on interaction networks traditionally seem to constitute the bulk of the literature on social group characterization from relational data in SNA [28, 68]. They focus on the shape and structure of relationships *between* actors and appear to pay little attention to the cognitive and property-based aspects

C. Roth (✉)

Sciences Po, Médialab, 84 rue de Grenelle, 75007 Paris, France

Centre Marc Bloch Berlin e.V., Friedrichstrasse 191, 10117 Berlin, Germany

e-mail: camille.roth@sciencespo.fr

© Springer International Publishing AG 2017

R. Missaoui et al. (eds.), *Formal Concept Analysis of Social Networks*,
Lecture Notes in Social Networks, DOI 10.1007/978-3-319-64167-6_1

of communities as such (even though they may indirectly uncover affiliation communities, i.e., social groups defined by similar attributes, for instance, as a result of homophily, or the fact that similar people interact more with one another [48]). Furthermore, many approaches based on affiliation networks nonetheless fall back on an interaction network by transforming two-mode into one-mode data, i.e., by building an actor network whose links denote shared affiliations among actors. Typical cases include co-appearance in a same movie, co-authorship of a same scientific article, or co-membership in a same team. Some further approaches aim at truly uncovering groups in the two-mode network [31], yet in general an actor-centric view eventually appears to prevail. In other words, the composition of the detected communities and, subsequently, the validity of the results are principally discussed in terms of actors, whereas attributes essentially remain in the background as an instrumental helper: somehow, semantic similarity is used as an indirect tool to uncover implicit interactional patterns.

This is where, we contend, lies one of the most crucial assets of dual approaches such as formal concept analysis (FCA [33]) with respect to SNA: the possibility of describing hybrid communities of actors and attributes in a simultaneous manner, without giving priority on one mode over the other, while tackling several of the key challenges raised by community detection in traditional SNA.

By focusing on knowledge communities and, more precisely, by emphasizing the possibility of formalizing the notion of *epistemic community* (EC), at the interface between SNA and FCA, this chapter aims at showing how FCA may particularly contribute to SNA in uncovering and describing social groups based on cognitive affiliation patterns. To this end, we first recall how structural approaches have formalized the notion of interactional community, discussing in particular the main quantitative issues and qualitative connections with sociological analysis. We then explain how FCA enables, by contrast, the description of social groups which are characterized by attribute similarity and for which it is more straightforward to use affiliation networks. We show how FCA still captures many important community features of interest to SNA. We illustrate this stance with a series of empirical examples.

2 Communities in Interaction Networks

2.1 *Explicit vs. Procedural Methods*

Algebraic Definitions of Social Groups Formal apprehension of the sociological notion of “community” principally stems from SNA [22], all the more as social interactions have progressively occurred in an increasingly networked fashion [69]. Historically, the introduction of graph theory in sociometry [27] paved the way to the first mathematical analyses of communities: the so-called sociogram of a given group of actors, which describes relationships such as acquaintances, friendships, collaborations, or exchanges, could be represented as a graph. Then, the

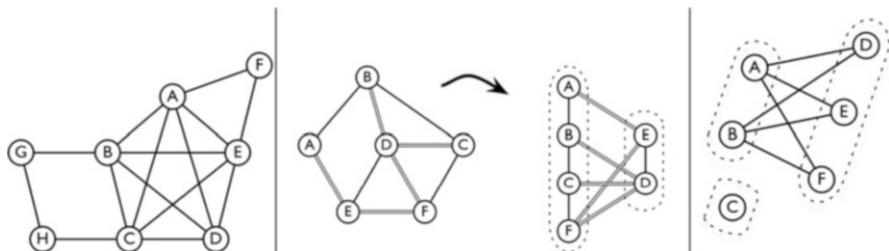


Fig. 1 Explicit, algebraic definitions of groups. *Left:* $\{A, B, C, D, E\}$ and $\{A, E, F\}$ are cliques, $\{B, C, G, H\}$ is a 2-clique. *Middle,* Cartwright and Harary’s decomposition [13] (positive or negative links are, respectively, represented by *single* or *double* lines). *Right:* structural equivalence classes of this network include $\{A, B\}$, $\{C\}$, and $\{D, E, F\}$

abstract study of its algebraic and topological structure could reveal the existence of “real” communities, by matching a given qualitative community definition with quantitative graphic properties.

The notion of *clique* [47] has played a foundational and prototypical role in this endeavor. Cliques shall sound familiar to FCA scholars, whose formal concepts are maximal bicliques of a bipartite graph isomorphic to the object-attribute matrix. Interactional cliques are even simpler patterns: they configure subsets of individuals who are *all* connected with one another, i.e., complete subgraphs of the interaction network (see an illustration in Fig. 1). Cliques can thus be seen as the most basic and strongest cohesive community unit. In practice, however, cliques larger than a dozen of actors are relatively rare. Furthermore, as may be the case with formal concepts, their computation, representation, and even interpretation often prove difficult. SNA thus quickly introduced less rigid notions of communities, starting with *n*-cliques [46] which allow for a looser connectivity among individuals belonging to the same group (they have to be at most at distance *n* from each other).

Methods which are more global and holistic were also proposed very early to partition a given network into various sub-communities, rather than just exhibiting local patterns such as cliques. Building upon the so-called balance theory introduced in psychology [37] and which allows for either positive or negative relationships between actors (friends/foes, i.e., valued links), Cartwright and Harary [13] were among the first ones to formalize communities at the network-level with their structure theorem. In a nutshell, when some composition laws on relationships hold (namely, foes of friends are foes), they showed that it is possible to split nodes into two groups such that intra-group (resp. inter-group) connections are positively (resp. negatively) valued. Here, communities follow from antagonistic rather than similar interactional configurations. Multiple refinements of this approach have later been introduced [17, 18, 20, 21], focusing, for instance, on the role of triads—again a very local pattern.

Beyond these foundational milestones, SNA has developed over the previous decades a very rich and diverse set of definitions of social groups, where patterns directly match explicit mathematical expressions. Many contributions within this

research program are variously evoked in [29, pp. 152–153], [68, Chaps. 7, 9, and 10], [30, pp. 743–744], or [22, pp. 206–207]. Let us mention, for instance, the notion of “equivalence class” of individuals, which describes groups of actors connected to other actors in an equivalent manner. This notably includes *structural equivalence* [45] where actors of the same class share exactly the same neighbors, *regular equivalence* [71] where actors of the same class are linked in a similar way to actors of another class, or *automorphic equivalence* [25], where actors of the same class occupy positions which are exactly interchangeable in the network (their labels may be exchanged without changing the relational structure). In a different direction, the *structural cohesiveness* of a set of actors [70] is defined as the number of individuals which have to be removed in order to get disconnected components, i.e., such that there exists at least one pair of individuals who are not indirectly connected through a chain of links. A group with a structural cohesiveness of k is called a *k-component*: here, communities are groups such that links between actors exhibit some redundancy.

Procedural Methods and Approximate Patterns Formalization does not necessarily imply quantification. In this respect, most of these algebraic approaches were fueled by mathematical sociologists who initially worked on case studies based on small datasets stemming from ethnographic observation, thereby featuring a limited number of actors. As a result, they are essentially adapted to small-sized networks and structures [50] since the number of patterns can grow quickly. How to deal, for instance, with the thousands of cliques which a small network of a hundred of nodes may contain; and what to deduce from their observation?

A more recent stream of research focused in all generality on the quantitative and large-scale study of the topology of social (and non-social) networks. This stance gained momentum during the 2000s, thanks to the joint availability of powerful computational resources and large relational datasets (even if this phenomenon could already be partly perceived as early as the 1970s [2, p. 116]). Under the term “community detection,” this literature addresses the issue of the discovery of cohesive structures in large graphs by applying data mining techniques developed to a large extent by computer scientists and statistical physicists [28]. Within this stream, groups or communities are consensually seen as aggregates of actors in the network: “*groups of vertices within which connections are dense, but between which connections are sparser*” [51]. This is aligned with a classical SNA definition: “*its members should have many relations with each other and few with non-members*” [2, p. 121]. Concretely, these approaches are based upon procedural methods and thus tend to blur the distinction between the formal definition of what these “dense” groups are and the algorithm which enables their detection. In contrast with explicit and closed mathematical definitions where “a *group/community* is a set of actors such that [...]” dense group patterns are almost entirely defined by the procedure—all the more when algorithms are stochastic and results vary from an execution to the other. This allows for scalability and, often, compactness of the partitions, to the expense of interpretability.

These algorithms may diversely feature the iterative construction of a series of embedded graph partitions, either by gathering structurally close individuals into

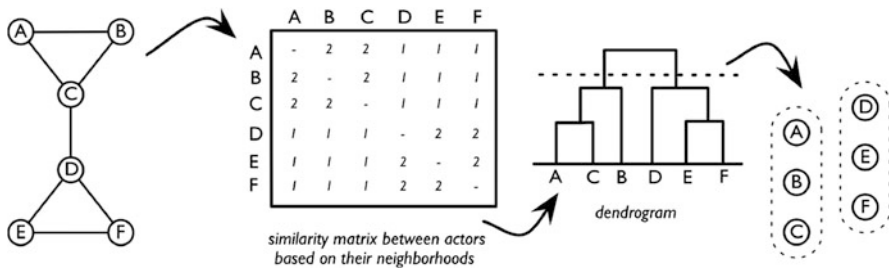


Fig. 2 Partition creation algorithm. *From left to right:* toy network; similarity matrix based on the number of node-independent paths between actors (i.e., paths involving strictly distinct node sets, except for the extremities); iterative construction of the dendrogram based on these similarities (progressively gathering nodes with highest similarities, randomly choosing in case of equality); groups of the partition found at the cut-off stage of the dendrogram, marked by a dashed line

increasingly larger groups [14, 72] or by dividing the whole graph into increasingly smaller groups [34]. This procedure is traditionally denoted as *hierarchical clustering* [68], it may be represented as a dendrogram; various criteria such as modularity [52] are then available to decide which partition to choose and where to cut the dendrogram (see Fig. 2 for a toy example). Other procedures can be based on network exploration [7], possibly inspired by percolation processes in order to find community boundaries [54], or holistic methods such as spectral decomposition based on some global properties of the graph adjacency matrix [12, 56].

2.2 Structural Properties of Groups

Structural methods may go beyond the mere partitioning of nodes: they may further be used to describe group structure in itself, i.e., the relationships between groups. Blockmodeling methods, for one, generalize partitioning by reducing the social graph into a meta-graph of groups called *blockmodel*, where nodes represent groups and links describe their relationships.

At the group level, more broadly, we may identify three classical qualitative phenomena which are an important and current research issue in SNA: (1) hierarchies between groups, (2) multiple membership of actors in groups, and (3) temporal dynamics of groups.

Group Hierarchies SNA makes it generally easy to describe social group orders and hierarchies, first and foremost by relying on set inclusion. A group can be “below” or “more specific than” another one if the former is included in the latter: a partial order may be defined where, say, $\{A, B\}$ and $\{B, C\}$ are included in $\{A, B, C\}$ while $\{A, B\}$ and $\{B, C\}$ cannot be compared with one another. Some methods naturally and implicitly define such an order: dendrograms configure increasingly finer partitions, while k -components are included in k' -components when $k' < k$. Traditionally, the resulting hierarchical structure is a *tree* comparable to Aristotelian

taxonomies (as in the traditional classification of scientific disciplines: e.g., “scientists” > “biologists” > “molecular biologists” > ...). Hierarchies may also be defined among items of a partition, especially when interactions are directed or valued: [18] uses link asymmetry to define levels between groups, such that “*admiration flows up levels*” as a consequence of differences in the underlying actor prestige or centrality [68, Chap. 5].

Group Overlap Beyond partitions where individuals are meant to belong to a single group (as is the case with equivalence classes), a somewhat small part of the literature has addressed the question of multiple membership [3, 9, 26, 32, 54, for instance]. Here, actors may belong to one or more groups which can in turn partially overlap. While the relevance of taking into account such overlap is sometimes debated (e.g., [29, p. 153]), the relative weakness of scholarly interest in this issue may also be explained by concrete hurdles, such as how to properly justify thresholds triggering multiple membership, or how to deal with the potential combinatorial complexity.

Group Dynamics By definition, interactional analysis of social groups steers clear of intensional properties: in a dynamic perspective, this means that the old sociological question of the perpetuation of social groups¹ is appraised through the stability of interactional structures across time rather than the persistence of their attributes. Typically, inter-temporal correspondence may be assessed longitudinally (groups at t are associated with groups with similar members at t' [19, for instance]) or dynamically (the stability of relationships between t and t' defines the group, as in [49, 53], thereby assuming that social entities only exist by way of their temporal stability [1]). We shall show below how FCA brings a particular added value for this and the above issues, especially in the context of knowledge communities.

3 Reuniting Structure and Content

3.1 Affiliation Networks, Social Circles, and FCA

As mentioned in the introduction, interactional network analysis provides a robust set of methods to define social groups, yet by overlooking a priori their non-structural properties. In this way, since interactional SNA does not rely on intensional properties, it may fail to render the most semantic and cognitive aspects of communities—unless one assumes a strong redundancy between structural and

¹ “The most general case in which the persistence of the group presents itself as a problem occurs in the fact that, in spite of the departure and the change of members, the group remains identical. We say that it is the same state, the same association, the same army, which now exists that existed so and so many decades or centuries ago. This, although no single member of the original organization remains.” [64, p. 667]

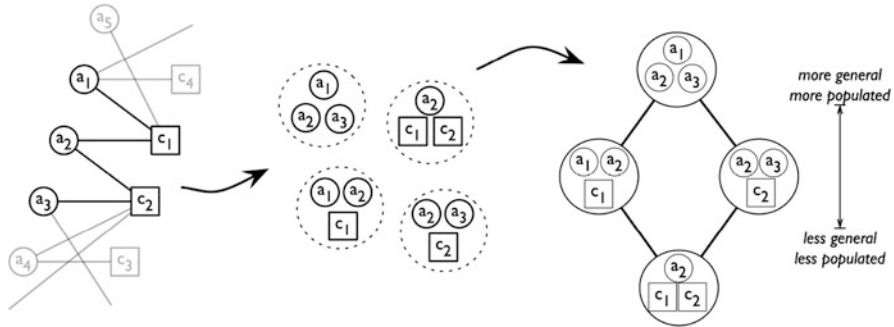


Fig. 3 Bipartite graphs and lattices. *Left:* toy bipartite graph: actors a_1, a_2, a_3, \dots and cognitive properties c_1, c_2, \dots . *Middle:* formal concepts corresponding to the black portion of the previous graph. *Right:* lattice-based hierarchical representation

non-structural properties. As such, a social group featuring semantic or cognitive affinity may only be found indirectly if the similarity is manifest in the interactional structure, for example, because of homophily: for instance, scientific collaboration networks exhibit some disciplinary cohesiveness [34, 57]. Semantic labels for Interactional groups are usually labelled a posteriori, if at all, and often by hand. Moreover, larger groups such as schools of thoughts, epistemic communities, interest circles and, more broadly, socio-cognitive groups may not correspond univocally to a single, well-defined interactional community.

The branch of SNA based on affiliation networks appears here as a robust relational framework able to combine structure and semantics. Technically, affiliation networks are bipartite graphs, where actors on one side are distinguished from affiliations on the other side (Fig. 3, left panel). A link may only connect an actor and an affiliation. This formalism is additionally dual, as are social circles [10], in the sense that affiliations are linked to actors just as actors are linked to affiliations.

Social circles are thus explicitly codified in the data: a single affiliation already constitutes an intensional group which denotes the shared participation in an event, membership in an organization, interest for a topic, adhesion to a belief. In this respect, looking for groups in affiliation networks may also be understood as the task of uncovering new (implicit) actor groups from the multiple intersections of social circles, which are thus seen as (explicit) intensional groups [5, 8, 32, 44]. From the viewpoint of SNA, this stance enables both a structural and a cognitive description of communities, which is the cornerstone of describing *socio-cognitive taxonomies*, i.e., joint taxonomies of actors and taxonomies of cognitive attributes—be it in the context of scientists working on research topics, bloggers posting about some issues, activists discussing political matters. Bipartite graphs are isomorphic to binary relations and to labeled hypergraphs (indeed, actor nodes affiliated with the same attribute in a bipartite graph univocally correspond to a labeled hyperedge)—the closeness with FCA is straightforward when considering actors as objects and affiliations as attributes.

While several studies aim specifically at detecting group patterns in bipartite graphs [43], they often tend to consider affiliations as an instrumental rather than fundamental feature. More precisely, many seem to discard the inherent duality either *ex ante*, by focusing on an actor–actor network derived from the original bipartite graph (through a projection of the two-mode network onto one of the modes), or *ex post*, by computing groups of actors with similar properties then discussing the validity of the detected groups principally in terms of actors.

Typically, FCA appears to be one of only a few current methods which aim at maintaining the duality of actors and affiliations along the whole process, from pattern detection to taxonomy interpretation. With respect to the above-mentioned SNA techniques (Sect. 2.1), it also relies on an explicit definition of what a group is, rather than relying on a procedural definition. We will discuss below how FCA also addresses the above-mentioned classical SNA challenges—dealing with group hierarchy, overlap, and dynamics. The resulting computational complexity is also an issue, which has been partly addressed by introducing the first practical application of stability [41] in the very case of socio-cognitive taxonomies and knowledge communities [42, 61].

3.2 Formal Concepts as Epistemic Communities

Before that, we first explain the plain application of FCA on affiliation networks. Formally, we consider the affiliation network as a pair of sets of actors \mathcal{A} and cognitive properties \mathcal{C} (described by, e.g, n -grams, lexical tags, topics, representations, etc.), i.e., agents and notions (or “concepts” in the generic sense of the word), and a binary relationship between them, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{C}$. The *intent* A' of a set of actors $A \subseteq \mathcal{A}$ is the intersection of all sets of cognitive properties associated with actors of A , i.e., $A' = \{c \in \mathcal{C} \mid \forall a \in A, a\mathcal{R}c\}$; dually, the *extent* C' of a set of cognitive properties C is the intersection of all actor sets associated with properties of C , i.e., $C' = \{a \in \mathcal{A} \mid \forall c \in C, a\mathcal{R}c\}$. Applying successively “'” yields a closure operator. For all subsets $A \subseteq \mathcal{A}$ and $C \subseteq \mathcal{C}$, (A'', A') and (C', C'') are called *formal concepts* and, equivalently, are maximal bicliques in the bipartite graph of the affiliation network.

In the context of knowledge communities, an efficient qualitative interpretation of formal concepts/biclique patterns consists in considering these socio-semantic groups as *epistemic communities* (EC). Introduced in [63] and later refined by [36] and used by many social scientists afterwards [15, 16], this notion essentially corresponds to actor groups who (1) share some interest for a certain set of topics or beliefs and (2) have a common goal of knowledge creation while obeying to some set of given rules agreed upon in the underlying community. In the very minimal sense, an EC may be formalized as a pair of agents and topics such that

all agents share all topics; that is, a biclique in the bipartite affiliation network ($\mathcal{A}, \mathcal{C}, \mathcal{R} \subseteq \mathcal{A} \times \mathcal{C}$). Each EC thus algebraically defined corresponds to a socio-cognitive group which is the closure of a set of actors or equivalently of cognitive properties—a socio-semantic pattern. See illustration in Fig. 3—middle.

Lattices and Socio-Cognitive Taxonomies This formalism addresses several of the issues exposed in Sect. 2.2 regarding interactional groups. In particular, it enables a hierarchical representation of groups through the natural inclusion-based partial order on formal concepts. Conceptually, this hierarchy induces a generalization/specialization relationship: it may be represented as a *lattice* [5]. The most general ECs (largest actor sets/extents, smallest attribute sets/intents) are found towards the top, while the most specific ECs are at the bottom (most specific extents, largest intents). See illustration in Fig. 3—right. This configures a socio-cognitive taxonomy relevant to social epistemology—for one, it is useful to represent distributed cognition activities [38] in a given knowledge production system, in particular the distribution of topics over actors.

Moreover, lattices configure non-Aristotelian taxonomies: ECs partially overlap. Of course, individuals may belong to more than one EC but, more importantly, ECs may also have more than one parent. Arguably, this property makes lattice-based taxonomies closer to cognitive categories, where ECs may simultaneously be subsets of *several* more general ECs.

Finally, it is possible to track the dynamics of these taxonomies by following the evolution of actor sets associated with a given attribute set, thus echoing the ambition of Simmel regarding the persistence of social groups (footnote 1). Note that this approach also inherits a drawback typical of community detection methods based on explicit definitions, especially in the case of cliques: computational complexity. Even for a small number of actors and properties, the number of ECs and the lattice size can be dramatically large [33], easily running in the thousands. This problem is typically critical for SNA scholars, who rarely use cliques, if any. In the next section, we discuss concrete strategies to tackle these issues efficiently.

4 Applications

From the viewpoint of FCA, knowledge communities typically feature either a significant number of actors, or of notions, or both: it is thus key to explain and emphasize how FCA can be of practical use despite combinatorial complexity, especially to compete or keep up with some of the above-mentioned SNA approaches, most notably those based on procedural methods. Data reduction is here a crucial issue, both in terms of input or output, i.e., at the level of the primary data or the computed results.

4.1 Datasets

We present three earlier empirical applications: two are related to scientific communities, one features political activists and motions. These case studies were diversely introduced in [42, 60, 62]: more specific details on each of them may be found in the respective references. In the meantime, FCA has been increasingly applied to groups of actors sharing some properties (e.g., [4, 55]).

In all cases, the empirical material consists of text documents describing, to some extent, who writes about or is interested in what. Actors of the corresponding affiliation networks are identified as document authors, while cognitive attributes are terms extracted from the plain text. A link between actor a and term c occurs whenever a authored a document mentioning c . In this respect, epistemic communities/formal concepts observed in these empirical case studies are strictly speaking socio-lexical patterns.

As the number of individual terms in the original data is always very large, especially with regard to FCA, we systematically apply some filtering relying on simple natural language processing (NLP) techniques. We lemmatize words, exclude stop-words, and eventually focus on the most frequent terms, additionally selecting the most meaningful ones with the help of a domain expert. The number of actors appears to be generally more tractable, yet when it is too large (as in the zebrafish case), we show how simple sampling strategies can be used. See Table 1 for basic statistics regarding the datasets.

The *zebrafish community* case study gathers embryologists who worked on an animal model called “zebrafish” over the years 1990–2003. This period corresponds to the early development of the field, whose population grew approximately tenfold [60]. Data was gathered from the publicly available bibliographical database MEDLINE by querying papers whose abstract includes “zebrafish”—assuming that in most cases authors who work on this animal would necessarily evoke the term in their abstract. The *ECCS* dataset focuses on scholars working on complex systems, focusing on the two first editions of the European Conference on Complex Systems, in 2005 and 2006. The conference organizers kindly provided us with submitted abstracts to both conferences, which we all used in the original study [42]. Finally, the *political motions* example is based on the six roadmaps submitted by six groups of members of the French socialist party towards the internal elections at their Congress in 2008 [62]. In these texts, signatories defend their vision of where

Table 1 Basic figures describing the size of the respective empirical datasets in terms of original documents, unique actors, and NLP-extracted terms

Dataset		Documents	Actors	Terms
ECCS	2005	194	413	92
	2006	187	401	109
Political motions		6	6	85
Zebrafish	1990–1995	533	1094	66
	1998–2003	4080	9689	67

the party should go in the coming years. We consider motions as actors of the corresponding affiliation network, i.e., six nodes; we also keep 85 pre-processed words appearing at least 32 times in the whole corpus.

4.2 Socio-Cognitive Taxonomies

Hierarchy and Overlap We first use the zebrafish case study to illustrate the hierarchy and overlap between groups which is made possible by FCA-based socio-cognitive taxonomies. The period 1990–1995 already features a thousand of actors and 66 attributes—something which yields about eight million ECs and, admittedly, can get neither drawn nor interpreted. A first reduction strategy may consist in operating at the level of the input data by sampling the actor set, assuming that a random portion of the population would still render a faithful taxonomy of the whole community (if needed, removed actors may later be assigned to the computed taxonomy). We use an affiliation subnetwork including a random share of 20% of the population and use it to compute a formal concept lattice made of about 200k ECs. This still represents a sizeable number of ECs, and further reduction may be needed. A second strategy may consist in filtering the output, for instance, by conserving formal concepts according to some relevance criterion. The so-called iceberg lattices [67] have been classically used, whereby a certain portion of the top of the lattice is conserved, assuming that this portion corresponds plausibly to the most interesting or the most meaningful part of the taxonomy. Extent size, i.e., population size of ECs, is a popular criterion; distance to the top may also be used. In Fig. 4, we show such a truncated lattice for the period 1990–1995, together with the last period 1998–2003, to exhibit the temporal evolution.

Let us first focus on the general structure for a given period, say 1998–2003, after the zebrafish community reached some maturity. This picture describes succinctly its main research axes, their representativity, overlaps, and hierarchical relationships. To put it shortly, we see three pillars: (1) comparative studies occupy an important position (*human/mouse/homologous genes*), (2) the study of the nervous system, around the *dorsal* and *ventral* plates, also gathers a certain proportion of scholars, and (3) systemic studies linked to signaling during embryonic development are well-represented (*signal/pathway/growth/receptor*).

Temporality Additionally, we may compare lattices for different periods within the same knowledge community. By focusing on identical attribute groups (intents) across time, FCA makes it possible to render the temporal evolution and relative stability of, at the macro-level, socio-cognitive taxonomies and, at the meso-level, social groups—a key issue in SNA as well. Here, however, the inter-temporal correspondance of groups will be based on attributes rather than interactions.

In practice, we represent evolution by coloring ECs corresponding to a given intent according to the growth of their population share (extent representativity).

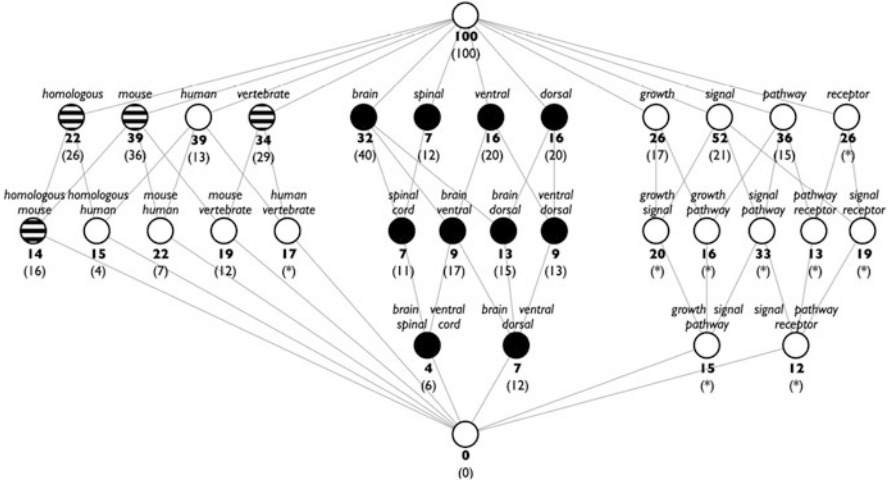


Fig. 4 Truncated lattice of the zebrafish community [59, 60]. *Lines* represent direct hierarchical (inclusion) relationships. Each EC is described by a word list (the intent of the formal concept) and its actor size as a percentage of the underlying population (extent of the formal concept) in *bold* for 1998–2003 and in *parentheses* for 1990–1995. *White disks* indicate growth from the first to the last period, *solid black disks* indicate decline, and *dashed disks* correspond to relative stagnation (arbitrarily defined as variation smaller than $\pm 15\%$). Percentage sums go over 100% as a result of multiple membership. By definition the *bottom node* gathers all attributes (not listed here)

We see in Fig. 4 that comparative studies have expanded within the zebrafish community, together with the analysis of systemic signals, which echoes a general trend in molecular biology at that time, whereas studies centered around the embryonic nervous system are progressively fading (also a general trend in the surrounding fields). While showing the diversity of the distribution of cognitive tasks within the community, this comparison also demonstrates that it enjoys a remarkable stability, given that the underlying population grew tenfold between the two periods.

Approximation We now turn to the political motion dataset to illustrate reduction strategies further. We use stability [41], a criterion which removes redundancy across the whole lattice and has been widely used in the FCA community since its inception [11, 61]. It indeed constitutes a robust approach to deal with potentially large lattices such as those emerging from empirical social data, while still paying attention to smaller yet plausibly meaningful and representative patterns which would be filtered out by top-down approaches based on, e.g., iceberg-like criteria.

In a nutshell, the (extensional) stability of a given formal concept (A'', A') is formally defined as $\sigma(A'', A') = |\{B \subseteq A'' | B' = A'\}| / 2^{|A''|}$, i.e., the proportion of subsets B of the actor set A'' of a given formal concept whose intent B' is identical to A' . Put differently, this criterion measures how much the existence of a given EC depends on its actors. The higher the σ , the more stable the EC, and the more likely it will be presented in the final results.

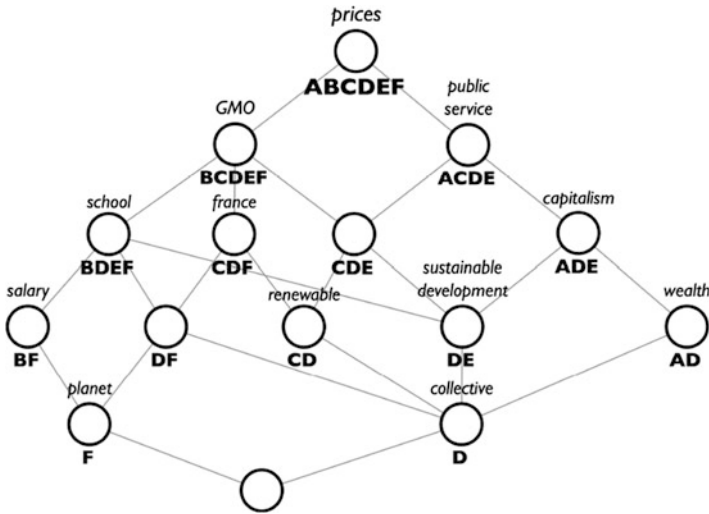


Fig. 5 Stabilized concepts for the political motion dataset (after [62]). Letters A–F refer to the six motions. Here, for the sake of readability, cognitive attributes of nodes are labeled in a parsimonious way: we only indicate terms which are *added* to formal concept intents from *top* to *bottom*; for instance, the node “(france, CDF)” corresponds to an EC gathering motions CDF and terms “france,” “GMO,” and “prices” (terms are thus inherited from *bottom* to *top*)

Figure 5 presents a reduced taxonomy based on stability. It remains readable while featuring specialized groupings quite deep down from the top. At the most general level, the structure exhibits the omnipresence of issues related to purchase power—all motions talk about “prices”—or GMOs, used by all but motion A. Yet, we also see progressively smaller groupings: for instance, school-related issues (used by motions B, D, E, and F), and then, even lower, joint use of “sustainable development” by D and E, or “salary” by B and F. Dually, we see that motion D is present in almost all ECs by addressing issues present in all other motions.

Combining Both We finally use the ECCS case to illustrate the application of both principles: temporality and approximation (see [42] for more details). Figure 6 shows the 15 most stable concepts for lattices computed over all authors in each year.

On the whole, the main pillars of this scientific field revolve around “networks,” “models,” and their “dynamics,” as well as, to a weaker extent, “structure” and “distribution” (which, in this context, most often refer to scale-free distributions). At the global level, structures for both time periods are relatively comparable. A finer examination reveals some differences: several specific ECs (subconcepts) disappeared in 2006 ({network, dynamics}, {dynamics, model, process}, {dynamics, process}, and {information}) while others appeared ({interaction}, {network, social}, {model, agent}, and {simulation, model}). Focusing on specific intents also provides extra information on the epistemological evolution in 2006: for instance,

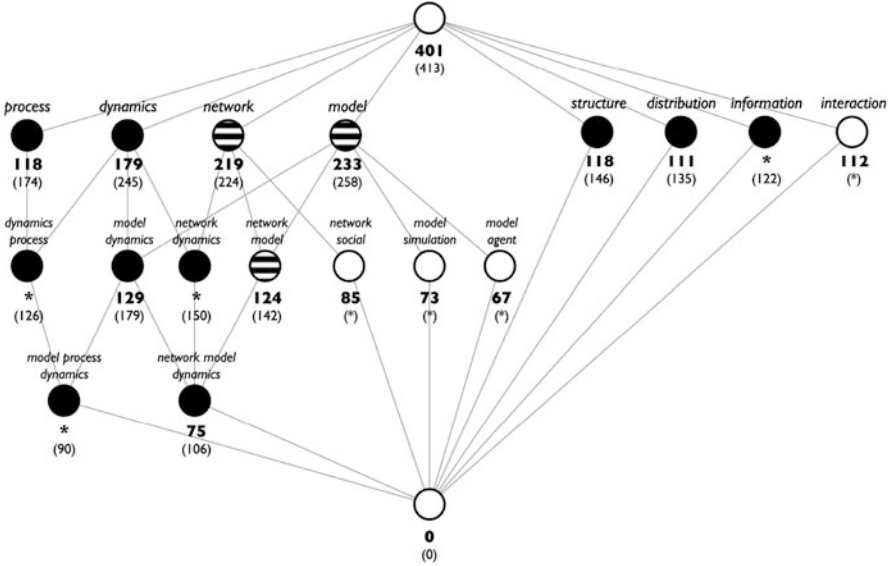


Fig. 6 Stabilized concepts for ECCS-2005 and ECCS-2006 (after [42]). Figures are absolute size of ECs

the EC on {network, dynamics} does not exist anymore on its own, while {network, dynamics, model} still does, suggesting that network dynamics is entirely subsumed by dynamic network models.

5 Concluding Remarks

Beyond the diversity of the SNA literature on the detection of groups, we could draw a fundamental dichotomy between interaction-based and affiliation-based group definitions. In the very case of scientific communities, social scientists argue for a similar dichotomy [40] between “*taxonomic collectives*,” which are relevant at a high level of observation, and interaction groups, in which actors are embedded and which are also relevant at the local level to understand actor behavior. The case studies presented here show how the notion of EC and, behind this, FCA applied to affiliation networks provide a description of the configuration of actor groups in knowledge communities in a manner at least similar to what is possible through classical interactional SNA, while taking actor attributes into account.

With these dichotomies in mind, we can sketch some of the issues where FCA could create a most relevant bridge over SNA for the study of knowledge networks. This includes, first and foremost, the study of the correlation between affiliation and interaction communities. In other words, describe to what extent socio-cognitive communities are also strongly cohesive in interactional terms, how taxonomic

collectives may be interaction groups and whether epistemic communities do cover various interaction communities. More generally, do one-mode communities correspond to two-mode communities or formal concepts [58]? Here, some empirical answers have been recently proposed in this direction [65] relying on the so-called alpha concept lattices [66].

Second, on a more practical and theoretical level, the development of approximation strategies is key to guarantee the acceptance of FCA by SNA scholars. This is all the more true in socio-cognitive contexts where result interpretability, both in terms of social groups and in terms of cognitive taxonomies, needs to be manually tractable and therefore involve a sensibly limited number of categories. Stability-based pruning is an option among many, especially in the case of noisy data stemming from social behavior [39]. The design of scalable selection criteria [35] adapted to a socio-cognitive context could be another promising direction of research.

Third, much remains to be done with respect to the dynamics, for instance, by digging further the intensional stability of communities across time. As could be seen here, socio-cognitive taxonomies plausibly evolve slowly, even in the case of a high turnover of actors from a period to the other. On the FCA side, this touches the issue of inter-lattice comparisons [73] and their temporal analysis [24, 74], even though this area remains relatively nascent in FCA. On the side of SNA, group evolution is mainly assessed through a single-network lens. Most likely, appraising simultaneously the joint evolution of social and cognitive patterns, possibly to the point where social groups are even defined by the dynamic stability of socio-cognitive patterns, would constitute a fruitful contribution to social analysis.

Acknowledgements The present contribution partially relies on ideas introduced in a book chapter originally published in French and entitled “Communautés, analyse structurale et réseaux socio-sémantiques” [59].

References

1. Abbott, A.: Things of boundaries. *Soc. Res.* **62**(4), 857–882 (1995)
2. Alba, R.D.: A graph-theoretic definition of a sociometric clique. *J. Math. Sociol.* **3**, 113–126 (1973)
3. Arabie, P., Carroll, J.D.: Conceptions of overlap in social structure. In: Freeman, L.C., White, D.R., Romney, A.K. (eds.) *Research Methods in Social Network Analysis*, pp. 367–392. George Mason University Press, Fairfax, VA (1989)
4. Balamane, A., Missaoui, R., Kwuida, L., Vaillancourt, J.: Descriptive group detection in two-mode data networks using biclustering. In: *Proc. of 2016 IEEE/ACM Intl. Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE Computer Society, San Francisco (2016)
5. Barbut, M., Monjardet, B.: *Algèbre et Combinatoire*, vol. II. Hachette, Paris (1970)
6. Bell, C., Newby, H.: *Community Studies: An Introduction to the Sociology of the Local Community*. Allen & Unwin, London (1972)

7. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008)
8. Boeck, P.D., Rosenberg, S.: Hierarchical classes: model and data analysis. *Psychometrika* **53**(3), 361–381 (1988)
9. Bonacich, P.: Using boolean algebra to analyze overlapping memberships. *Sociol. Methodol.* **9**, 101–115 (1978)
10. Breiger, R.L.: The duality of persons and groups. *Soc. Forces* **53**(2), 181–190 (1974)
11. Buzmakov, A., Kuznetsov, S.O., Napoli, A.: Is concept stability a measure for pattern selection? *Proc. Comput. Sci.* **31**, 918–927 (2014)
12. Capocci, A., Servedio, V., Caldarelli, G., Colaiori, F.: Detecting communities in large networks. *Physica A* **352**, 660–676 (2005)
13. Cartwright, D., Harary, F.: Structural balance: a generalization of Heider’s theory. *Psychol. Rev.* **63**, 277–292 (1956)
14. Clauset, A.: Finding local communities in networks. *Phys. Rev. E* **72**, 026132 (2005)
15. Cohendet, P., Créplet, F., Dupouet, O.: Organisational innovation, communities of practice and epistemic communities: the case of Linux. In: *Economics with Heterogeneous Interacting Agents*, pp. 303–326. Springer, Berlin (2001)
16. Cowan, R., David, P.A., Foray, D.: The explicit economics of knowledge codification and tacitness. *Ind. Corp. Chang.* **9**(2), 212–253 (2000)
17. Davis, J.A.: Clustering and structural balance in graphs. *Hum. Relat.* **20**, 181–187 (1967)
18. Davis, J.A., Leinhardt, S.: The structure of positive interpersonal relations in small groups. In: Berger, J., Zelditch, M., Anderson, B. (eds.) *Sociological Theories in Progress*. Houghton Mifflin, Boston, MA (1970)
19. Doreian, P.: On the evolution of group and network structure. *Soc. Netw.* **2**, 235–252 (1979)
20. Doreian, P., Mrvar, A.: A partitioning approach to structural balance. *Soc. Netw.* **18**(2), 149–168 (1996)
21. Doreian, P., Mrvar, A.: Partitioning signed social networks. *Soc. Netw.* **31**, 1–11 (2009)
22. Edling, C.R.: Mathematics in sociology. *Annu. Rev. Sociol.* **28**, 197–220 (2002)
23. Elias, N.: Towards a theory of communities. In: Bell, C., Newby, H. (eds.) *The Sociology of Community: A Selection of Readings*. Routledge, London (1974)
24. Elzinga, P., Wolff, K., Poelmans, J.: Analyzing chat conversations of pedophiles with temporal relational semantic systems. In: *Proc. 1st IEEE European Conference on Intelligence and Security Informatics*, pp. 242–249. Odense, Denmark (2012)
25. Everett, M.G.: Role similarity and complexity in social networks. *Soc. Netw.* **7**, 353–359 (1985)
26. Everett, M.G., Borgatti, S.P.: Analyzing clique overlap. *Connections* **21**(1), 49–61 (1998)
27. Forsyth, E., Katz, L.: A matrix approach to the analysis of sociometric data: preliminary report. *Sociometry* **9**(4), 340–347 (1946)
28. Fortunato, S.: Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010)
29. Freeman, L.C.: The sociological concept of ‘group’: an empirical test of two models. *Am. J. Sociol.* **98**(1), 152–166 (1992)
30. Freeman, L.C.: Un modèle de la structure des interactions dans les groupes. *Rev. Fr. Sociol.* **36**, 743–757 (1995)
31. Freeman, L.C.: Finding social groups: a meta-analysis of the Southern women data. In: Breiger, R., Carley, K., Pattison, P. (eds.) *Dynamic Social Network Modeling and Analysis*, pp. 39–97. The National Academies Press, Washington, DC (2003)
32. Freeman, L.C., White, D.R.: Using Galois lattices to represent network data. *Sociol. Methodol.* **23**, 127–146 (1993)
33. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin (1999)
34. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. *PNAS* **99**, 7821–7826 (2002)

35. Gnatyshak, D., Ignatov, D.I., Semenov, A., Poelmans, J.: Gaining insight in social networks with biclustering and triclustering. In: Aseeva, N., Babkin, E., Kozyrev, O. (eds.) *Perspectives in Business Informatics Research BIR 2012: 11th Intl. Conf., Nizhny Novgorod, Russia, Sept 24–26*, pp. 162–171. Springer, Berlin (2012)
36. Haas, P.: Introduction: epistemic communities and international policy coordination. *Int. Organ.* **46**(1), 1–35 (1992)
37. Heider, F.: Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946)
38. Hutchins, E.: Distributed cognition. In: Smelser, N.J., Baltes, P.B. (eds.) *International Encyclopedia of the Social and Behavioral Sciences*, pp. 2068–2072. Elsevier, Amsterdam (2001)
39. Klimushkin, M., Obiedkov, S., Roth, C.: Approaches to the selection of relevant concepts in the case of noisy data. In: Kwuida, L., Sertkaya, B. (eds.) *Proc. 8th Intl. Conf. Formal Concept Analysis. LNCS/LNAI*, vol. 5986, pp. 255–266. Springer, Berlin (2010)
40. Knorr-Cetina, K.: Scientific communities or transepistemic arenas of research? A critique of quasi-economic models of science. *Soc. Stud. Sci.* **12**(1), 101–130 (1982)
41. Kuznetsov, S.: Stability as an estimate of degree of substantiation of hypotheses derived on the basis of operational similarity. *Nauchn. Tekh. Inf.* **2**(12), 21–29 (1990)
42. Kuznetsov, S., Obiedkov, S., Roth, C.: Reducing the representation complexity of lattice-based taxonomies. In: Priss, U., Polovina, S., Hill, R. (eds.) *Conceptual Structures: Knowledge Architectures for Smart Applications: 15th Intl. Conf. on Conceptual Structures, ICCS 2007, Sheffield, UK. LNCS/LNAI*, vol. 4604, pp. 241–254. Springer, Berlin (2007)
43. Latapy, M., Magnien, C., Vecchio, N.D.: Basic notions for the analysis of large two-mode networks. *Soc. Netw.* **30**(1), 31–48 (2008)
44. Lehmann, S., Schwartz, M., Hansen, L.K.: Biclique communities. *Phys. Rev. E* **78**, 016108 (2008)
45. Lorrain, F., White, H.C.: Structural equivalence of individuals in social networks. *J. Math. Sociol.* **1**(49–80) (1971)
46. Luce, R.D.: Connectivity and generalized cliques in sociometric group structure. *Psychometrika* **15**, 169–190 (1950)
47. Luce, R.D., Perry, A.: A method of matrix analysis of group structure. *Psychometrika* **14**, 95–116 (1949)
48. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001)
49. Mitra, B., Tabourier, L., Roth, C.: Intrinsically dynamic network communities. *Comput. Netw.* **56**(3), 1041–1053 (2012)
50. Moody, J.: Peer influence groups: identifying dense clusters in large networks. *Soc. Netw.* **23**, 261–283 (2001)
51. Newman, M.E.J.: Detecting community structure in networks. *Eur. Phys. J. B* **38**, 321–330 (2004)
52. Newman, M.E.J.: Modularity and community structure in networks. *PNAS* **103**(23), 8577–8582 (2006)
53. Palla, G., Barabási, A.L., Vicsek, T.: Quantifying social group evolution. *Nature* **446**, 664–667 (2007)
54. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005)
55. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: a survey on applications. *Expert Syst. Appl.* **40**(16), 6538–6560 (2013)
56. Pothen, A., Simon, H.D., Liou, K.P.: Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.* **11**(3), 430–452 (1990)
57. Rodriguez, M.A., Pepe, A.: On the relationship between the structural and socioacademic communities of a coauthorship network. *J. Informet.* **2**, 195–201 (2008)

58. Roth, C.: Binding social and semantic networks. In: Proceedings of ECCS 2006, 2nd European Conference on Complex Systems, Oxford (2006)
59. Roth, C.: Communautés, analyse structurale et réseaux socio-sémantiques. In: Sainsaulieu, I., Salzbrunn, M., Amiotte-Suchet, L. (eds.) *Faire communauté société – Dynamique des appartenances collectives*, pp. 113–128. Presses Universitaires de Rennes, Rennes (2010)
60. Roth, C., Bourguine, P.: Lattice-based dynamic and overlapping taxonomies: the case of epistemic communities. *Scientometrics* **69**(2), 429–447 (2006)
61. Roth, C., Obiedkov, S., Kourie, D.G.: Towards concise representation for taxonomies of epistemic communities. In: Yahia, S.B., Nguifo, E.M. (eds.) *Proc. CLA 4th Intl. Conf. on Concept Lattices and Their Applications*. LNCS/LNAI, vol. 4923, pp. 240–255. Springer, Berlin (2006)
62. Roth, C., Cointet, J.P., Obiedkov, S., Romashkin, N.: Analyse textuelle des motions du Congrès de Reims du PS (2008). <http://tinyurl.com/39g6lch>
63. Ruggie, J.G.: International responses to technology: concepts and trends. *Int. Organ.* **29**(3), 557–583 (1975)
64. Simmel, G.: The persistence of social groups. *Am. J. Sociol.* **3**(5), 662 (1898)
65. Soldano, H., Santini, G.: Graph abstraction for closed pattern mining in attributed networks. In: ECAI, pp. 849–854 (2014)
66. Soldano, H., Ventos, V.: Abstract concept lattices. In: Valtchev, P., Jäschke, R. (eds.) *Proc. Intl. Conf. on Formal Concept Analysis (ICFCA)*. LNAI, vol. 6628, pp. 235–250. Springer, Heidelberg (2011)
67. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with TITANIC. *Data Knowl. Eng.* **42**, 189–222 (2002)
68. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge (1994)
69. Wellman, B., Carrington, P.J., Hall, A.: Networks as personal communities. In: Wellman, B., Berkowitz, S.D. (eds.) *Social Structures: A Network Analysis*, pp. 130–184. Cambridge University Press, Cambridge (1988)
70. White, D.R., Harary, F.: The cohesiveness of block in social networks: node connectivity and conditional density. *Sociol. Methodol.* **31**, 305–359 (2001)
71. White, D.R., Reitz, K.P.: Graph and semigroup homomorphisms on networks of relations. *Soc. Netw.* **5**, 193–234 (1983)
72. White, H.C., Boorman, S.A., Breiger, R.L.: Social-structure from multiple networks. I: blockmodels of roles and positions. *Am. J. Sociol.* **81**, 730–780 (1976)
73. Wille, R.: Concept lattices and conceptual knowledge systems. *Comput. Math. Appl.* **23**, 493 (1992)
74. Wolff, K.: Applications of temporal conceptual semantic systems. In: Wolff, K., Palchunov, D.E., Zagoruiko, N.G. (eds.) *Knowledge Processing and Data Analysis*. LNAI, vol. 6581, pp. 59–78. Springer, Heidelberg (2011)