

# The Evolution and Dynamics of Research Networks

Vladimir Batagelj<sup>1</sup>, Bettina Hoser<sup>2</sup>, Claudia Müller<sup>3</sup>, Steffen Staab<sup>4</sup> and Gerd Stumme<sup>5</sup>

<sup>1</sup> University of Ljubljana, Department of Mathematics  
1000 Ljubljana, Jadranska 19, Slovenia  
`vladimir.batagelj@fmf.uni-lj.si`

<sup>2</sup> University of Karlsruhe, Institute for Information Systems and Management  
(Information Services and Electronic Markets)  
76128 Karlsruhe, Kaiserstrasse 12, Germany  
`Bettina.Hoser@em.uni-karlsruhe.de`

<sup>3</sup> University of Stuttgart, Institute for IT-Services  
70550 Stuttgart, Allmandring 30 E, Germany  
`cmueller@iits.uni-stuttgart.de`

<sup>4</sup> University of Koblenz-Landau, Institute for Computer Science (Information Systems and Semantic Web)  
56070 Koblenz, Universitätsstrasse 1, Germany  
`staab@uni-koblenz.de`

<sup>5</sup> University of Kassel, Electrical Engineering and Computer Science (Knowledge & Data Engineering)  
34121 Kassel, Wilhelmshöher Allee 73, Germany  
`stumme@cs.uni-kassel.de`

**Abstract.** Existing collaboration and innovation in scientific communities can be enhanced by understanding the underlying patterns and hidden relations. Social network analysis is an appropriate method to reveal such patterns. Nevertheless, research in this area is mainly focused on social networks. One promising approach is to use homophily networks as well. Furthermore, extending the static to a dynamic network model enables to understand existing interdependencies in these networks. A mathematical description of possible analyses is given. Finally, resulting research questions are illustrated and the necessity of an interdisciplinary research approach is pointed out.

**Keywords.** Homophily networks, social networks, evolution, scientific community

## 1 Analysis of Research Networks

Nowadays, communities of researchers have been studied from many different angles for many years (e.g., [1][2], [3], [4]). This subject is very attractive to researchers, as it is relatively simple to obtain data for the analysis [5]. Furthermore

the collaboration of researchers can be improved based on the understanding of the underlying patterns and the relations in scientific communities.

However, most of the analyses focus on a single type of network (e.g., only on co-authorship or only on citation networks). In order to get a complete picture of a research community, all information contained in these different networks should be combined. Our long-term vision is to build a system that allows a researcher – who is not a specialist in Social Network Analysis (SNA) – to get insights in any research domain that he is not acquainted with yet. For instance, a researcher, who has a background in semantic web communities, wants to gain knowledge in data mining field. He would be satisfied to find out all the information about the new community i.e. main actors, the main lines of research, the most central papers, etc., without extensive literature work. Nowadays, there are many algorithms for different kinds of network analyses (e.g., [6], [7]) or knowledge discovery (e.g., [8]). All of them have a number of drawbacks like they are not interacting in a coherent way, they are usually restricted to one specific type of dataset, and their results are not easy to interpret by a non-SNA-specialist. A creation of a system that overcomes these drawbacks poses interesting implementation and research issues.

From an application perspective, we would require such a system to be able to fulfill at least the following tasks:

- Identify the main trends in a specific research community (like the semantic web community) in the last  $n$  years,
- Reveal the main trends of a specific topic (like “topic detection”) in the last  $n$  years (This differs from the previous task such that it may involve different communities who do not know each other),
- Observe how colleagues move through communities (“What is Tim Berners-Lee up to next?”),
- Analyse one community over time (decay, growth, stability),
- Discover the emergence, convergence/divergence and decay of communities. (mergers & acquisitions) and
- Recommend, for a given interest profile, the membership in a community, or the creation of a new community.

The following contribution summarizes the results of a lively two days lasting discussion about how scientific communities can be analysed on a timely basis and how existing changes in the topic and community structure can be identified. This contribution is organized as follows: First of all, we started to apply the idea on folksonomies and social platforms. As a result a general model is defined. Then an introduction in network evolution is given. The proposed methods allow to meet same tasks of the proposed application. We conclude with a short outlook. Open questions in the field of interdependent evolution of networks are discussed.

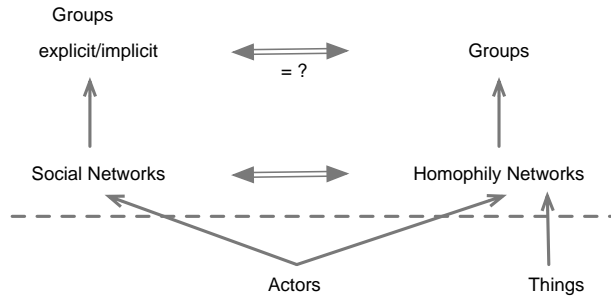
## 2 A Static Model of Homophily and Social Networks

We started our considerations with a definition of a general model. This model is based on folksonomies and social platforms, where the following two kinds of networks are recognized:

1. Homophily is the tendency of individuals to contact similar people at a higher rate than dissimilar people. Cultural, behavioral, genetic, or material information that flows through networks will tend to be localized. The distance in terms of social characteristics is translated into network distance [9].
2. A social network is a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of interdependency representing social acts such as communication (e.g. letters, e-mail, face-to-face), co-authorship of texts, trading or friendship.

In spite of the differences, empirical investigations in the past have shown that the two kind of networks exhibit correlations. How one network may give predictive insight into the other and how one network helps to explain and predict relationships in the other remains an open question up to now.

Both kinds of network are built from an overlapping set of entities. Social networks (typically) constitute one mode networks of actors and specific relationships between them. Homophily networks constitute higher mode  $n$ -partite networks of actors and things. Also, the homophily networks lead to relationships between actors – but only inferred relationships of closeness by interest, which do not necessarily imply that the one actor is aware of the other.

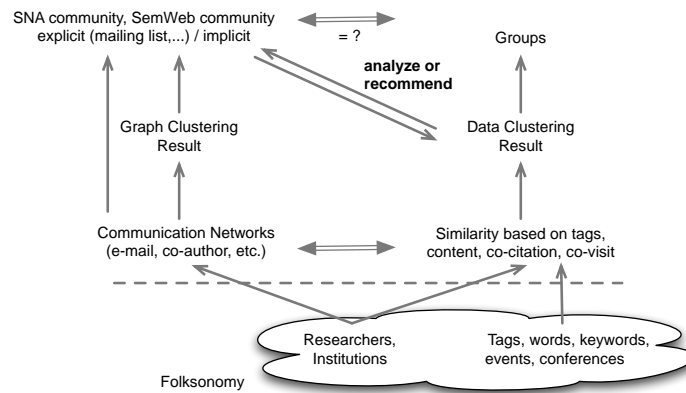


**Fig. 1.** Deriving Groups from Communities

In homophily networks, unsupervised data mining techniques, e.g. K-means, lead to clusters of actors into common interest groups based on the description of actors based on many things, e.g. tags, keywords, or articles they cite. In social networks, graph clustering techniques, e.g. spectral methods or cliques, lead to clusters of actors based on how the actors communicate. Both clustering techniques return groups. No matter how meaningful these groups are, they

need to be carefully evaluated. But clearly, the different groups lead to different insights into the underlying community under consideration.

An overview of interaction between the two kind of networks is sketched in Figure 1. Thereby, the figure also mentions some methods in order to illustrate its points, but the set of methods is only exemplary and not meant to be exclusive. The figure illustrates that one may define e-mail or co-authorship networks based on actors – here researchers – (first level) which are organized in institutions communication networks (second level). Researchers use systems like Bibsonomy to organize their publications and references. Using this information, the similarities of tags, content or co-citations can be obtained (homophily networks – level 3). The results of analyses are applied to identify implicit and explicit communities.



**Fig. 2.** Example Methods for Clustering into Groups

### 3 Evolution of Homophily and Social Networks

In the proposed approach, there are three levels: actor descriptions, homophily relations and social networks. Using the interplay of these three levels we are trying to answer several questions about the evolution of networks.

The basic level are the actor descriptions. Let  $V$  be the set of actors; then to each actor  $v \in V$  is assigned its description  $T_v$ . For example: the description  $T_v$  can consist of tags or keywords used by author  $v$ , but also some other of his properties.

In the proposed approach, we are trying to relate the development of social network to the homophily relation among actors induced by (dis)similarity of their descriptions. Assuming that the descriptions of actors are “correlated

with” (are inducing) the social network structure, we can try to operationalize homophily using an appropriate dissimilarity between the actors’ descriptions.

For example  $d(u, v) = \frac{|T_u \oplus T_v|}{|T_u \cup T_v|}$  where  $\oplus$  is the symmetric difference –  $A \oplus B = (A \cup B) \setminus (A \cap B)$ . The dissimilarity  $d$  determines a kind of homophily space or field  $(V, d)$ . From it we can derive “homophily” networks – for example:  $r$ -neighborhoods network or  $k$ -nearest neighbors network.

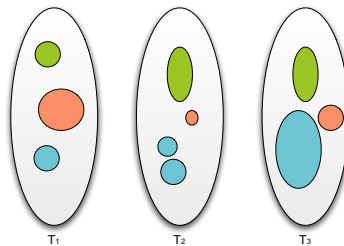
On the basis of the homophily relation under social actions of actors the social network  $(V, L)$  is constructed. The assumption is that closer actors in  $(V, d)$  have higher probability to establish a link in the social network.

An important question is, how good our assumption about the interconnections between the three levels is. Possible approaches to answer it are:

- Let  $N_r = (V, L_r)$  be the  $r$ -neighborhoods network –  $L_r = \{(u, v) : d(u, v) \leq r\}$  then we can observe how the “compatibility” index  $Com(r) = \frac{|L_r \cap L|}{|L|}$  is changing with  $r$ . There exists  $r^*$  such that  $Com(r) = 1$  for all  $r \geq r^*$ . Small value of  $r^*$  indicates strong influence of actors’ descriptions on formation of social network links.
- Another possibility is a kind of permutation test. We introduce the quantity  $S = \sum_{(u,v) \in L} d(u, v)$  measuring a stress in the social network; and, for permutation  $\pi : V \rightarrow V$ ,  $S(\pi) = \sum_{(u,v) \in L} d(\pi(u), \pi(v))$ . Now we can compare the value of  $S$  against the distribution of  $S(\pi)$  for random  $\pi$ s. If only few values of  $S(\pi)$  are smaller than  $S$ , this is an indication of strong influence of descriptions/homophily on the structure of social network.

The interconnection goes also in the opposite direction: actors linked in the social network will get more similar descriptions.

Till now we were looking at the static model. A sequence of such models indexed by time  $t$  describes the evolution of the system. An example is given in figure 3. Here we see how over time sub-communities of the larger community split, merge, grow and shrink respectively.

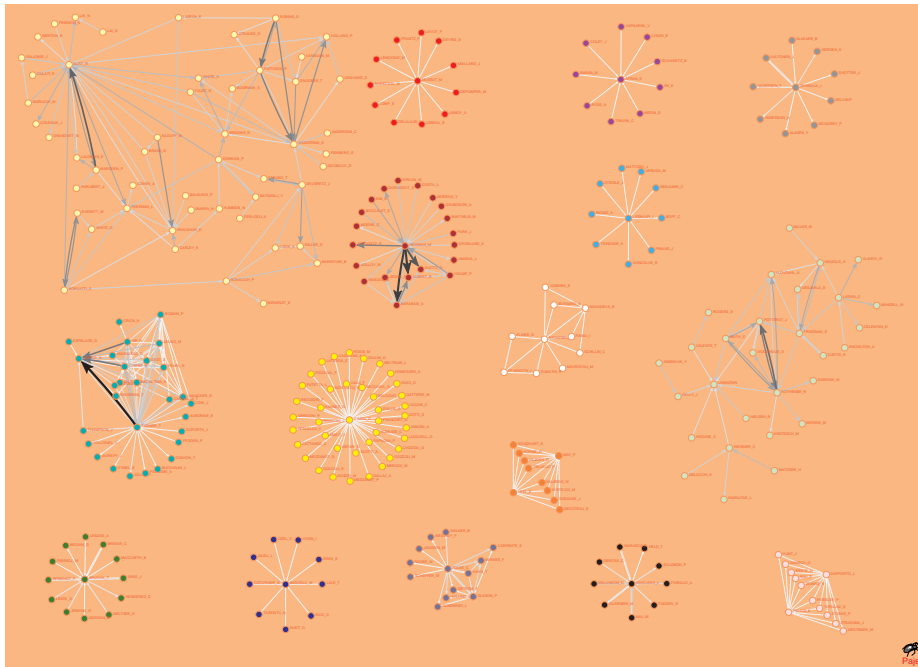


**Fig. 3.** Evolving communities

Assume that in each time point we have a clustering  $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$  where  $\emptyset \subset C_i \subseteq V$ . Note that,  $\mathcal{C}$  need not to be a partition, and some clusters  $C_i$  can even overlap.

The clustering  $\mathcal{C}$  can be explicit (given as data) or implicit (determined by some procedure, such as islands algorithm [10] or clustering with relational constraint [11] or  $k$ -cliques [12] or cores [13] etc.). In directed networks, an interesting approach could be the single-center clusters

We also assume that the evolution relation  $\rightsquigarrow$  among clusters from consecutive time points is known:  $C_i(t) \rightsquigarrow C_j(t+1)$ , meaning that cluster  $C_j(t+1)$  evolved from cluster  $C_i(t)$ . Again it can be given as data or computed. There are nontrivial problems in computationally determining the evolution relation, which are obtained by the leaders strategy in constrained clustering.



**Fig. 4.** Islands in authors' citations network of Social Networks field

What are the properties of the evolution relation? Different patterns can be noticed in the evolution of clusters [14]: birth of new cluster, merging/fusion of clusters, growth of cluster, splitting of cluster, contraction/decay of cluster, death of cluster, stability of cluster, internal change in cluster, etc.

The clusterings can be analyzed at the level of social network based on structural characteristics of clusters: single-kernel (strong component, clique, etc.), centrality/centralization, etc. For example:

- not single-kernel cluster can indicate a future splitting of the cluster and
- the change in the cluster’s kernel structure can indicate a change of cluster’s topics.

Additional analyses can be based on the other two levels:

- $S(C)$  can be used to measure the stress/divergence in a cluster  $C$ ,
- to each cluster  $C$  we can assign its vocabulary  $T(C) = \bigcup_{v \in C} T_v$ . Analyzing the changing of vocabulary through time  $(T(C_t))_{t \in T}$ , we can get insight into the life of the cluster: convergence to common topics, changing of interest, etc.

All of the described analyses can be done already using existing tools. But in several places these analyses are too crude – new methods, more tuned to the data and problems, need to be developed.

Here, we described only the simplest model which can be extended in some directions:

- in the same model we can consider different sets of vertices: authors, works, tags, journals/conferences, institutions, countries, etc. and
- in the same model we can consider different social networks: co-authorship, citations, communications, etc.

## 4 Outlook

The future research questions will focus on the problem of interdependent evolution of networks. This means that the analysis of the evolution of different networks is defined on the same set of actors. An example in the research community could be the co-evolution of the co-authorship-network, the co-citation network, the co-participation in research projects and for instance the communication network (this is most probably difficult to obtain). Questions of research might be:

- How does the evolution and the dynamics in one network influence the others?
- Is there a “leading” network, which mainly drives the co-evolution of the coupled system?
- How do social processes provide an explanation to the results found?
- Are there predictors for success or failure of a coupled system, given that success or failure have been defined in the context?

To even go a step further, one could take a look at the question: what will happen when the set of actors is unstable or even unpredictable? In the “real world” actors join and leave groups constantly. How can this behavior be modeled into the dynamics of co-evolving networks?

These questions call for an interdisciplinary research approach. Coupled systems and their behavior are a well established field within physics. The social

aspects of the question are being investigated in sociology and psychology. The data acquisition and handling need computer and information scientists, if the data is gathered from the Internet. Ideas might also be found in other fields like biology or chemistry.

There are a number of benefits of such a research. If for instance, one is the administrator of a collaboration platform for scientists, another one could monitor how the community as a whole works, but also how parts of the community are evolving. Then another one could intervene if the systems shows signs of failure. Another application could be a “recommender service” for researchers. This system will give results such as relevant papers, relevant actors, or the change of research topics within a given field.

## References

1. Freeman, L.C.: The impact of computer based communication of the social structure of an emerging scientific speciality. *Social Networks* **3** (1984) 201–221
2. Newman, M.E.: The structure of scientific collaboration communities. *Proceedings of the National Academy of Science (PNAS)* **98** (2001) 404–409
3. Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. *Physica* **311** (2002) 590–614
4. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Physical Review E* **69** (2004)
5. Batagelj, V.: Wos2pajek – networks from web of science (<http://pajek.imfm.si/doku.php?id=wos2pajek>) (2007)
6. Wasserman, S., Faust, K.: *Social network analysis: methods and applications*. Cambridge Univ. Press, Cambridge (1997)
7. Newman, M.E.: The structure and function of complex networks. *SIAM Review* **45** (2003)
8. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence Magazine* (1996) 37–54
9. McPherson, M., Lovin, L.S., Cook, J.M.: Birds of a feather: Homophily in social networks. *Annual Review of Sociology* (<http://dx.doi.org/10.1146/annurev.soc.27.1.415>) **27** (2001) 415–444
10. Zaveršnik, M., Batagelj, V.: Islands (<http://vlado.fmf.uni-lj.si/pub/networks/Doc/Sunbelt/islands.pdf>). In: XXIV International Sunbelt Social Network Conference, Portorož, Slovenia (2004)
11. Batagelj, V., Ferligoj, A., Mrvar, A.: Hierarchical clustering in large networks ([pajek.imfm.si/lib/exe/fetch.php?id=presentations&cache=cache&media=slides:clurelcon.pdf](http://pajek.imfm.si/lib/exe/fetch.php?id=presentations&cache=cache&media=slides:clurelcon.pdf)). In: Workshop on complex networks, Louvain-la-Neuve (2008)
12. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435** (2005) 814
13. Batagelj, V., Zaveršnik, M.: Generalized cores. <http://arxiv.org/abs/cs.DS/0202039> (2002)
14. Palla, G., Barabasi, A.L., Vicsek, T.: Quantifying social group evolution. *Nature* (<http://www.nature.com/nature/journal/v446/n7136/pdf/nature05670.pdf>) **446** (2007) 664