Chapter 1
Adaptive Networks

Thilo Gross and Hiroki Sayama

1.1 Introduction

Over the past decades it has become clear that the metaphor of *networks* – ensembles of discrete *nodes* connected by *links* – offers a powerful conceptual framework for the description and analysis of many real world systems [1–5]. The science of networks has grown into a field which is by now firmly established in several disciplines, including mathematics, physics, biology, computer science, economics and the social sciences. Dramatic progress has been made both in the characterization of real world networks and in the study of dynamical models of networks.

It is interesting to note that the notion of *dynamical networks* has so far referred to either one of two distinct concepts. First, the nodes can be individual dynamical systems which are coupled through static links. Second, the pattern of links, i.e. the network *topology*, can evolve dynamically in time. We can thus distinguish between *dynamics ON networks* and *dynamics OF networks*. Combining the two yields an *adasptive network* – a network whose links change *adaptively* with respect to its states, resulting in a dynamical interplay between the state and the topology of the network [6].

Adaptive networks are hardly new. Almost all real world networks are adaptive to some extent. Consequently, examples of adaptive networks occur in many disciplines and can be found in a large number of applied models. What is new, however, is that only over the recent years adaptive networks have come into focus of rigorous investigations that employ simple conceptual models. These investigations have revealed a number of new mechanisms and phenomena: Adaptive networks based on simple local rules can self-organize robustly toward phase transitions [7] and highly non-trivial complex topologies [8]; distinct classes of nodes can emerge spontaneously from an initially homogeneous population [9]; and, complex dynamics can be observed as a consequence of phase transitions [10] and bifurcations [11] that involve topological as well as local degrees of freedom.

T. Gross (✉)

Max-Plack-Institute for Physics of Complex Systems, 01187 Dresden, Germany

e-mail: thilo.gross@physics.org

While the results of the recent surge of activity on adaptive networks have been predominately reported in the statistical physics literature, the investigation of adaptive networks has profited greatly from approaches and insights from other fields of science, most prominently discrete mathematics, nonlinear dynamics, game theory, sociology, and computer science. As a consequence, the questions posed by adaptive networks are currently attacked from many different directions with the tools that the different disciplines have established. Despite the different approaches, common themes frequently emerge in the investigations, and insights of broad importance begin to crystallize. A theory of adaptive networks that provides a unifying framework linking the previously isolated phenomena has come within reach.

The purpose of this book is to provide a broad overview over major insights in the functioning of adaptive networks. We aim on the one hand to provide an accessible introduction for researchers starting to work on adaptive networks, and, on the other hand, to draw the attention of experts to parallel developments outside their own disciplines.

In the remainder of this introductory chapter, we will explain basic concepts and illustrate the origin of adaptive networks in different disciplines. We will start in Sect. 1.2 by introducing some key notions, especially the interplay between state and topology that is the key to many intriguing phenomena observed on adaptive networks. In Sect. 1.3, we will specifically discuss the separation of timescales between the dynamics of states and the evolution of topologies, and several important critical phenomena discovered with models that adopt separable timescales. In Sect. 1.4, we will focus on the formation of self-organized topology in adaptive networks, and discuss some important research agendas. In Sect. 1.5, we will introduce a more complex class of adaptive network models in which the timescales of changes of states and topologies are not separable. Finally, in Sect. 1.6, we will give an overview of the other chapters in this book, where we will highlight how the approaches shed light on adaptive networks from different directions and how these different views could be combined to reveal a coherent picture of this intriguing, rapidly evolving field.

1.2 The Interplay Between State and Topology

What makes dynamics in adaptive networks so different from dynamics *ON* networks or dynamics *OF* networks? To answer this question, we begin by considering dynamics on a static network first.

It is well known that almost all dynamical processes that can be placed on a network are highly sensitive to the network topology. This means that the dynamics explores the network topology and thereby, somehow, "encodes" topological information in the time series of the states of the nodes. A simple example is an infectious disease spreading across the network. Each node is either susceptible to the disease or infected. Susceptible nodes that are linked to infected nodes become infected with a certain rate, while infected nodes recover at a different rate, coming back

to susceptible again. Whether such a disease can persist in the network depends on the rates of infection and recovery, but also on the number of connections per network node, or the *degree* of the nodes. Specifically, it depends on a quantity called *mean excess degree*, which denotes the average number of additional links that one finds connected to a node that is reached by following a random link. In a large connected network in which the excess degree exceeds a certain threshold, the disease can persist indefinitely. In this case, every node will be infected once in a while. Consequently, we can estimate whether the mean excess degree exceeds the threshold (global topological information) just by watching the behavior of a single node (time series of states of an individual node) for a sufficiently long period of time.

The example given above illustrates how topological information could be encoded in the dynamics of every single node. In certain situations, we can use the information that is thus encoded; for instance, it has been proposed to simulate the dynamics of coupled oscillators on a given network in order to detect cliques of closely connected nodes. However, on networks with static topology, the flow of information is a one-way road. The information about the states of nodes cannot feed back to the network topology.

Let us return to the example of the disease spreading, but now we allow the susceptible nodes to cut links to infected neighbors with a certain probability. As a result of this additional rule, a dynamical feedback loop is formed between topologies and states of the network. Namely, the dynamics of the prevalence of the disease depends on the network topology, but the evolution of the network topology also depends on the prevalence of the disease.

1.3 Timescale Separation and Critical Phenomena

The coupling of the topological evolution and the dynamics of states leads to an intriguing interplay. The effect of this interplay can be well understood in systems with timescale separation. A good example that can be seen in our everyday life is a road network. It is apparent that the decision where to build new roads depends on the traffic load on the existing roads, and therefore on the dynamical state of the network. Once a new road is built, it will alter the traffic patterns. However, the two processes – the road construction and the traffic shift in response to it – take place at very different timescales: years compared to days.

From the theory of dynamical systems, it is known that, in systems with timescale separation, the fast processes are governed by the slow processes. For example, if we start our road network in a random initial state, the traffic load will quickly settle down to some pattern. This pattern is a dynamical attractor of the fast system. Only if the network is observed on a much longer timescale, one can notice that the traffic pattern still changes in response to the evolving topology. However, compared to phenomena on the longer timescale, the response of the traffic load to topological changes is so fast that it can be regarded an instantaneous process.

Unless discontinuous dynamical transitions in the dynamics of the fast system take place, it will remain in this *quasi-attractor* all the time. The states visited in this quasi-attractor, i.e., the specific patterns of traffic, are then governed by the network topology.

An important consequence of timescale separation was described in a ground-breaking paper by Bornholdt and Rohlf [7]. Suppose that the dynamics on the network undergoes a *phase transition*, if the topology is altered in a certain way. In the language of statistical physics, this transition is detected by an *order parameter* – a global property of the dynamics. Bornholdt and Rohlf showed that the order parameter is encoded in the dynamics and therefore accessible to every node when the system falls in its quasi-attractor. In the meantime, the topological evolution is assumed to take place a slower timescale, and therefore the information about the order parameter can be extracted and utilized for topological changes. If the topological evolution follows appropriate rules, it is conceivable that this system self-organizes toward a critical state in which the dynamics on the network is exactly in the critical, transitional regime. While this mechanism sounds slightly artificial at first, it turns out that it arises almost automatically if the system is governed by simple natural rules.

Let us return once again to the example of epidemics, but now assume that the nodes are very reluctant to cut links so that the topological dynamics is much slower than the dynamics of the epidemic. If we start the system in the epidemic state, every node will be infected once in a while. On a longer time scale, every node will decide to cut a link once in a while. This decreases the connectivity of the network very slowly, and eventually, the epidemic threshold is reached and the disease becomes extinct, freezing the network exactly at the critical connectivity. Many more realistic examples can be found in the context of regulatory and neural networks.

Critical states have some peculiar dynamical properties. In particular they are frequently linked to the appearance of power-laws in the sizes of avalanches of events. Furthermore, they can give rise to multiple coexisting periodic and quasi-periodic attractors. Several authors have presented evidence that this form of self-organization is present in biological systems at many different scales.

1.4 Self-Organization of Non-trivial Network Topologies

A different form of topological self-organization was first observed in a model by Ito and Kaneko [9]. They describe a system that is initially homogeneous, where every node is connected to every other node and all nodes are in the same state (except for very small perturbations). In their model, the states of the nodes as well as the strength of the links change continuously. A rule is applied that increases the link strength between nodes in similar states but effectively weakens the connection between nodes in different states. This rule leads to the evolution of two distinct classes of nodes: Leaders, which exert a strong influence on other nodes, and followers, which exert little influence on others. Although the topology of the network keeps evolving indefinitely, leaders remain leaders and followers remain followers.

The work of Ito and Kaneko has inspired several subsequent investigations, in particular in the context of game theory. Although the mechanism that drives the spontaneous formation of the distinct classes of nodes is still unclear, an interesting pattern can be observed: All models in which nodes in similar states are linked preferentially lead to the formation of heterogeneous topologies, which either follow a scale-free degree distribution or split into distinct classes of nodes with different mean degrees. By contrast, if the opposite rule is applied so that nodes in different states are linked preferentially, homogeneous topologies that favor synchronization are obtained [12].

Both the self-organized formation of distinct classes of nodes and the ability to communicate topological information seem to be important ingredients for the self-organized formation of global non-random topologies based on local rules. Indeed, the formation of such topologies, although not stationary ones, was first reported by Holme and Ghoshal [8], and later, the self-assembly of adaptive networks was studied by Rosvall and Sneppen [13]. Nevertheless, the formation of distinct topologies in adaptive networks still remains largely unexplored at this point. More work in this direction would be highly desirable as it is likely to reveal the mechanisms that govern structure formation in biological networks and could, in the form of design principles, be utilized in engineering applications.

1.5 Adaptive Networks with Inseparable Timescales

So far we have mainly focused on systems in which the dynamics on the network is faster than the topological evolution. We have seen that, in these networks, the dynamics can be used to communicate topological information among nodes. However, the opposite is also possible: if the topological evolution is much faster than the dynamics on the network then topology can be used as a medium to communicate information on the state of the nodes.

Yet another, particularly interesting situation can arise if the dynamics on the network and the dynamics of the network take place at the same timescale. In such a case, the states and the topology of the network are truly coupled, and the entwined feedback loops between topological and local information may give rise to highly complex outcomes. For instance, it has been shown that the adaptive interplay can give rise to new bifurcations and phase transitions that involve changes in the dynamics of state and the topology of the network [10, 11].

An alternative approach to describing the interplay between states and topologies at the same timescale is the *graph transformation* (also called graph rewriting or graph grammars), which has been studied in discrete mathematics and theoretical computer science for the last couple of decades. While very difficult to treat analytically, the graph transformation is a convenient way to formulate the evolution of networks in both states and topologies algorithmically, using some local rules of substructure rewriting applied to labeled graphs. This kind of formulation is particularly powerful for the description of complex network growth processes, such as morphogenesis and self-replication of biological systems. It is anticipated that the combination these algorithmic approaches

and other statistical-physics-based ones will produce novel research directions on adaptive networks.

1.6 In this Book

This book contains a broad survey of the state-of-the-art in adaptive networks. The individual chapters are contributed by leading researchers from different disciplines. The chapters aim on the one hand to provide first-hand accounts focused on significant insights that have been gained, and on the other hand to provide mini-reviews of selected areas.

The following three chapters of the book focus on empirical investigations of real-world adaptive networks. In Chap. 2, Palla, Pollner, Barabási, and Viscek present an approach to the evolution of communities in social networks. This approach is illustrated by the analysis of social network data which reveals significant differences between small and large communities. Chapter 3 by Braha and Bar-Yam connects almost seamlessly to the first. Here a different approach to the analysis of real-world social networks is proposed by which the role of nodes in the network can be quantified. Again the application to social network data reveal a surprising result; while local hubs are found as expected, who is a hub and who is not changes on short timescales. This result strongly emphasizes the inherent dynamic nature of adaptive networks that has long been neglected in the literature. The theme of the investigation of real world adaptive networks continues in the fourth chapter in an entirely different context. Fricker, Boddy, Nakagaki, and Bebber present data on the growth of fungal mycelial networks. In the forest floor these networks can grow to enormous size and play an important role for forest ecosystems. In the context of this book mycelial networks are interesting as they can be cultured in the lab and thus constitute an example of real-world adaptive networks that can be studied with relative ease.

In the fifth chapter we launch into the theoretical investigation of adaptive networks. In this chapter Rohlf and Bornholdt review their original work on self-organized criticality in adaptive networks as well as several subsequent models inspired by the original paper. The basic mechanism proposed is probably responsible for self-organization of many systems in biology and beyond. The self-organization of adaptive networks with time-scale separation is further illustrated by the work of Caldarelli and Garlaschelli presented in Chap. 6. Here, an adaptive ecological network is considered in which Bak-Sneppen-like dynamics are coupled to topological evolution rules. This model illustrates clearly how the behavior of both the topological model and the Bak-Sneppen-model are significantly altered as they are combined.

The subsequent chapters focus on a different form of self-organization of adaptive networks. In Chap. 7 Ito and Kanako show how distinct classes of nodes can spontaneously emerge in a system of coupled oscillators. While many other models have been used to investigate the emergence of discrete classes, the models reviewed

in this chapter remain, to our knowledge, the only ones that show such an emer-
gence from an initially homogeneous population and use deterministic dynamics.
Chapter 8 by Chen and Kurths likewise focuses on systems of coupled oscillators.
In contrast to the setting studied by Ito and Kaneko, the aim is to build up not het-
erogeneous topologies but very homogeneous ones promoting synchronization. The
authors discuss different adaptive rewiring rules by which this aim can be achieved.

After networks with weighted links and continuous states have been discussed
in Chaps. 7 and 8, we return to the far simpler models with unweighted links and
discrete node states. In Chap. 9 Do and Gross review several recent investigations,
which apply such models for the study of opinion formation and epidemic spreading
on networks of social contacts. The comparison of the different model approaches
reveals several interesting parallels and differences. This chapter also introduces the
moment closure, an approximation scheme by which the dynamics of an adaptive
network can be captured in a low dimensional system of differential equations. This
approximation is then used by Shaw and Schwartz in Chap. 10 to analyze a more
complex epidemiological model. This chapter also contains an interesting discus-
sion of the effect of noise in adaptive networks.

Another class of systems that can be described by unweighted links and a finite
number of node states are discrete games on adaptive networks. In Chap. 11 we
return to the formation of complex topologies, but this time consider it from a game
theoretical perspective. Skyrms and Pemantle discuss several network formation
games of which Rosseau's stag hunt, a coordination game, is the most complex.
With great skill and mathematical rigor the authors show that, in the context of
adaptive networks, even very simple scenarios can yield interesting results. The sub-
sequent chapter, Chap. 12, connects almost seamlessly with a discussion of slightly
more complex games. Traulsen, Santos and Pacheco study three fundamental classes
of games using a powerful approximation scheme and numerical simulation. It is
shown that adaptive rewiring effectively changes the rules of the game. Thus, the
players actually behave as if they were playing a different type of game in a static or
well-mixed system. While well studied models from game theory can thus show new
types of behavior if considered on an adaptive networks, also entirely new games
become possible. In Chap. 13 Holme and Ghoshal propose a game in which the
players do not attempt to maximize an abstract payoff, but struggle for an advanta-
geous topological position on the network. The desired position is characterized by
low degree and high centrality which presents the nodes with a dilemma and results
in interesting dynamics.

The models discussed in Chaps. 9–13 show that even very simple rules can lead
to the formation of complex network topologies and non-trivial dynamics. Several
graph-transformation-based approaches for the automated discovery of such "inter-
esting" rules have been proposed in the context of computer science. In Chap. 14
Tomita, Kurokawa, and Murata propose a restricted, but very flexible class of adap-
tive networks for which an exhaustive search for rules and initial network config-
urations is possible. The authors apply this framework to construct self-replicating
networks by evolutionary optimization. Finally in Chap. 15, Sayama and Laramee
propose a modeling framework named generative network automata that extends

established methods for the investigation of cellular automata and other discrete dynamical systems to adaptive networks. This formal framework allows for an enumeration of all possible rules and therefore for an exhaustive search, which reveals several distinct classes of dynamical behavior.

References

1. R. Albert, A. Barabasi, Rev. Mod. Phys. **74**(1), 1 (2002)
2. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003)
3. M.E.J. Newman, SIAM Rev. **45**(2), 167 (2003)
4. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang, Phys. Reports **424**, 175 (2006)
5. M.E.J. Newman, A. Barabasi, D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, 2006)
6. T. Gross, B. Blasius, J. R. Soc. Interface **5**, 259 (2008)
7. S. Bornholdt, T. Rohlf, Phys. Rev. Lett. **84**(26), 6114 (2000)
8. P. Holme, G. Ghoshal, Phys. Rev. Lett. **96**, 908701 (2006).
9. J. Ito, K. Kaneko, Phys. Rev. Lett. **88**(2), 028701 (2002)
10. P. Holme, M.E.J. Newman, Phys. Rev. E **74**, 056108 (2007).
11. T. Gross, C. Dommar D'Lima, B. Blasius, Phys. Rev. Lett. **96**, 208701 (2006)
12. C.S. Zhou, J. Kurths, Phys. Rev. Lett. **96**, 164102 (2006).
13. M. Rosvall, K. Sneppen, Euro. Phys. Lett. **74**(6), 1109 (2006)

Chapter 2
Social Group Dynamics in Networks

Gergely Palla, Péter Pollner, Albert-László Barabási, and Tamás Vicsek

Abstract The rich set of interactions between individuals in the society results in complex community structure, capturing highly connected circles of friends, families, or professional cliques in a social network. Due to the frequent changes in the activity and communication patterns of individuals, the associated social and communication network is subject to constant evolution. The cohesive groups of people in such networks can grow by recruiting new members, or contract by loosing members; two (or more) groups may merge into a single community, while a large enough social group can split into several smaller ones; new communities are born and old ones may disappear. We discuss a new algorithm based on a clique percolation technique, that allows to investigate in detail the time dependence of communities on a large scale and as such, to uncover basic relationships of the statistical features of community evolution. According to the results, the behaviour of smaller collaborative or friendship circles and larger communities, e.g., institutions show significant differences. Social groups containing only a few members persist longer on average when the fluctuations of the members is small. In contrast, we find that the condition for stability for large communities is continuous changes in their membership, allowing for the possibility that after some time practically all members are exchanged.

2.1 Introduction

Mapping social relations between people onto a network has a long tradition in sociology [20, 72, 76]. The standard method for revealing the topology of the connections is to use questionnaires and personal interviews. The advantage of this approach is that it can provide very detailed information about the social ties, e.g., the type of acquaintance behind a given connection, what sort of emotions do the examined pairs of people induce in each other, whether the relation is mutual or

T. Vicsek (✉)

Statistical and Biological Physics Research Group of HAS, Department of Biological Physics, Eötvös University, 1117 Budapest, Pázmány Péter sétány 1/A, Hungary
e-mail: vicsek@angel.elte.hu

not, etc. The drawback of this data collection framework is that the typical size of the examined sample is of the order of $N \approx 10^2$ individuals and the strength associated to the links between people is subjective.

In the last decade a change of paradigm took place due to the rapid development of *complex network theory* [2, 4, 39, 75]. This new interdisciplinary field is devoted to the analysis of the statistical features of systems ranging from protein interaction networks through stock correlation graphs to the Internet. Since the size of the investigated networks can grow up to more than $N \approx 10^6$ nodes, the underlying data must be collected in an automated way, extracting the relevant information from large electronic databases. This approach has been successfully used to create large social networks as well [50, 51, 73]. E-mail databases [10, 12, 13], phone-call records [1, 50, 51] and scientific co-authorship data [5, 22, 23, 42] provide good examples for the starting point of a social network analysis on large scale. Although the range of social interactions that can be detected using data bases of this type is narrow compared to the questionnaires, in some cases the strength of the connections (e.g., the number of phone-calls between two individuals in a certain time period) may be more objectively quantifiable.

In this chapter we present a study concerning the statistical properties of two large social networks of major interest, capturing the collaborations between scientists and the calls between mobile phone users. Our focus is on the *community dynamics*, where the communities (also called as modules, clusters or cohesive groups) can correspond to families, friendship circles, work groups [63, 74], etc. These structural sub-units have no widely accepted unique definition, however we can assume that a community member is usually more tightly connected to its group than to other parts of the network, and that most people in a community know each other [15, 33, 43, 57, 64] (the groups are dense).

Although most empirical studies have focused on snapshots of these communities, thanks to frequent changes in the activity and communication patterns of individuals, the associated social and communication network is subject to constant evolution [5, 11, 31, 38, 47, 70, 78]. Our knowledge of the mechanisms governing the underlying community dynamics is limited, but is essential for a deeper understanding of the development and self-optimisation of the society as a whole [25, 28, 32, 34, 37, 56].

Typically, the communities in a complex system are not isolated from each other, instead, they have overlaps, e.g., people can be members in different social groups at the same time [72]. This observation naturally leads to the definition of the *community graph*: a network representing the connections between the communities, with the nodes referring to communities and links corresponding to shared members between the communities. Accordingly, the community degree d^{com} of a community is given by the number of other communities it overlaps with, and is equal to the degree of the corresponding node in the community graph. So far, in the networks investigated, the community degree distribution was shown to decay exponentially for low and as a power law for higher community degree values. This means that fat tailed degree distributions appear at two levels in the hierarchy of these systems:

both at the level of nodes (the underlying networks are scale free), and at the level of the communities as well.

Preferential attachment is a key concept in the field of scale-free networks. In a wide range of graph models the basic mechanism behind the emerging power law degree distribution is that the new nodes attach to the old ones with probability proportional to their degree [2, 4, 39]. Furthermore, in earlier works the occurrence of preferential attachment was directly demonstrated in several real world networks with scale free degree distribution [5, 41]. The observed fat tails in the degree distribution of the community graphs indicate that the mechanism of preferential attachment could be present *at the level of communities* as well. One of our aims in the present chapter is to examine the attachment statistics of communities in order to clarify this question.

We further develop a new algorithm based on the clique percolation method (CPM) [9, 53], that allows to investigate in detail the time dependence of overlapping communities on a large scale and as such, to uncover basic relationships of the statistical features of community evolution [52, 55]. According to our results, the behaviour of large – and small communities show an interesting difference. We find that large groups persist longer if they are capable of dynamically altering their membership, suggesting that an ability to change the composition results in better adaptability and a longer lifetime for social groups. Remarkably, the behaviour of small groups displays the opposite tendency, the condition for stability being that their composition remains unchanged. We also show that the time commitment of members to a given community can be used for estimating the community's lifetime.

This chapter is organised as follows. We begin with the construction of the investigated networks from the basic data sets in Sect. 2.2. and continue with the main aspects of the CPM in Sect. 2.3. We detail the algorithm for building evolving communities from subsequent snapshots of the community structure in Sect. 2.4. The main results are discussed in Sect. 2.5, whereas the concluding remarks are drawn in Sect. 2.6.

2.2 Construction of the Networks

The data sets we consider contain the monthly roster of articles in the arXiv.org cond-mat archive spanning 142 months, with over 30,000 authors [71], and the complete record of phone-calls between the customers of a mobile phone company spanning 52 weeks (accumulated over two week long periods), and containing the communication patterns of over 4 million users [50, 51]. Both type of collaboration events (a new article or a phone-call) document the presence of social interaction between the involved individuals (nodes), and can be represented as (time-dependent) links. We assumed that in both cases the *social connection* between people had started some time before the collaboration/communication events and lasted for some time after these events as well. (E.g., the submission of an article to

the archive is usually preceded by intense collaboration and reconciliation between the authors, which is in most cases prolonged after the submission as well). Collaboration/communication events between the same people can be repeated from time to time again, and higher frequency of collaboration/communication acts usually indicates closer relationship [58]. Furthermore, weights can be assigned to the collaboration and communication events quite naturally: an article with n authors corresponds to a collaboration act of weight $1/(n-1)$ between every pair of its authors, whereas the cost of the phone-calls provide the weight in case of the phone-call network. Based on this, we define the *link weight* between two nodes a and b at time t as

$$w_{a,b}(t) = \sum_i \left[w_i \Theta(t-t_i) \exp\left(-\lambda_+(t-t_i)/w_i\right) \right.$$
$$\left. + w_i \Theta(t_i - t) \exp\left(-\lambda_-(t_i - t)/w_i\right) \right], \qquad (2.1)$$

where the summation runs over all collaboration events in which a and b are involved e.g., a phone-call between a and b, and w_i denotes the weight of the event i occurring at t_i. (The constants λ_+ and λ_- are decay time characteristic for the particular social system we study. The function $\Theta(t)$ is the step function taking 0 at negative t values and 1 for positive). Thus, in this approach the time evolution of the network is manifested in the changing of the link weights. However, if the links weaker than a certain threshold w^* are neglected, the network becomes truly restructuring in the sense that links appear only in the vicinity of the events and disappear further away in time (Fig. 2.1). The above method of weighting ties between people is very useful in capturing the continuous time dependence of the strength of connections when the information about them is available only at discrete time steps. Except for our analysis of the preferential attachment of communities (Sect. 2.3.2.) we used symmetric decay characteristics $\lambda_- = \lambda_+$, whereas in Sect. 2.3.2. we applied a special choice corresponding to a simple growing network.
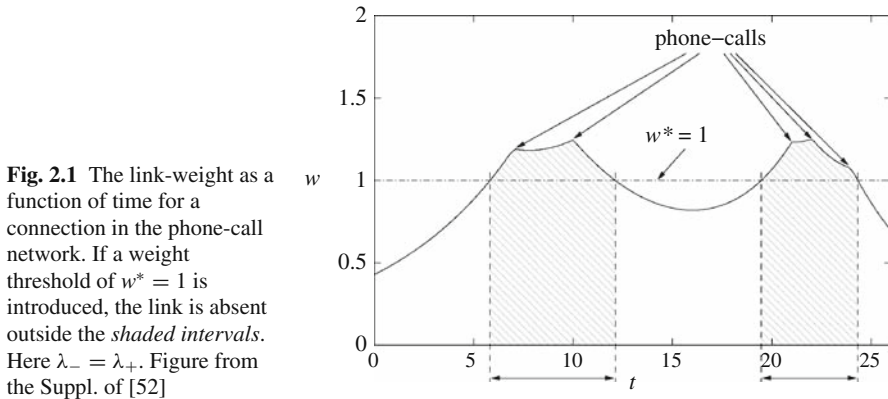

Fig. 2.1 The link-weight as a function of time for a connection in the phone-call network. If a weight threshold of $w^* = 1$ is introduced, the link is absent outside the *shaded intervals*. Here $\lambda_- = \lambda_+$. Figure from the Suppl. of [52]

2.3 Finding Communities

2.3.1 The Clique Percolation Method

The study of the *intermediate-scale* substructures in networks, made up of vertices more densely connected to each other than to the rest of the network, has become one of the most highlighted topic in complex network theory. These structural sub-units can correspond to multi-protein functional units in molecular biology [59, 65], a set of tightly coupled stocks or industrial sectors in economy [30, 49], groups of people [52, 63, 74], cooperative players [66, 67, 69], etc. The location of such building blocks can be crucial to the understanding of the structural and functional properties of the systems under investigation. Furthermore, a reliable method to pinpoint such objects has many potential industrial applications, e.g., it can help service providers (phone, banking, Internet, etc.) identify meaningful groups of customers (users), or support biomedical researchers in their search for individual target molecules and novel protein complex targets [3, 35].

Since communities have no widely accepted unique definition, the number of available methods to pinpoint them is vast [15, 16, 19, 24, 26, 27, 33, 40, 43, 53, 54, 60–64]. The majority of these algorithms classify the nodes into disjunct communities, and in most cases a global quantity called *modularity* [44, 45] is used to evaluate the quality of the partitioning. However, as pointed out in [17, 36], the modularity optimisation introduces a resolution limit in the clustering, and communities containing a smaller number of edges than $\sqrt{M}$ (where M is the total number of edges) cannot be resolved.

One of the big advantages of the *clique percolation method* (CPM) is that it provides a *local* algorithm for detecting the communities, and therefore, it does not suffer from resolution problems of this type [9, 53]. In this approach the communities are built up from k-cliques, corresponding to complete (fully connected) sub-graphs of size k. Two k-cliques are said adjacent if they share $k-1$ nodes [6, 9, 14], and a k-clique community corresponds to a set of k-cliques in which all k-cliques can reach each other through chains of k-clique adjacency. In other words, the communities defined in this way are equivalent to k-clique percolation clusters. These objects can be best visualised with the help of k-clique templates (Fig. 2.2), that are objects isomorphic to a complete graph of k vertices. Such a template can be placed onto any k-clique in the graph, and rolled to an adjacent k-clique by relocating one of its vertices and keeping its other $k-1$ vertices fixed. Thus, the k-clique percolation clusters (k-clique communities) of a graph are all those subgraphs that can be fully explored by rolling a k-clique template in them but cannot be left by this template.

The further advantages of the community definition above (beside its locality) are that it is not too restrictive, it is based on the density of the links and it allows *overlaps between the communities*: a node can be part of several k-clique percolation clusters at the same time. Revealing overlaps between communities has obtained a significant attention in the recent literature devoted to community detection [7, 18, 29, 40, 46, 60, 68, 77, 79]. Indeed, communities in real-world
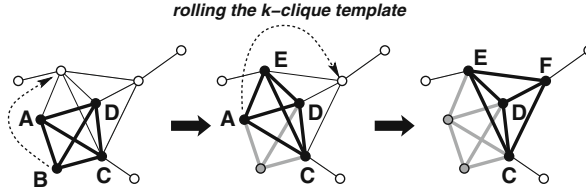
Fig. 2.2 Illustration of k-clique template rolling at $k = 4$. Initially the template is placed on A-B-C-D (*left panel*) and it is "rolled" onto the subgraph A-C-D-E (*middle panel*). The position of the k-clique template is marked with *thick black lines* and *black nodes*, whereas the already visited edges are represented by *thick gray lines* and *gray nodes*. Observe that in each step only one of the nodes is moved and the two 4-cliques (before and after rolling) share $k - 1 = 3$ nodes. At the final step (*right panel*) the template reaches the subgraph C-D-E-F, and the set of nodes visited during the process (A-B-C-D-E-F) are considered as a k-clique percolation cluster

graphs are often inherently overlapping: each person in a social web belongs usually to several groups (family, colleagues, friends, etc.), proteins in a protein interaction network may participate in multiple complexes [29] and a large portion of web-pages can be classified under multiple categories. Prohibiting overlaps during module identification strongly increases the percentage of false negative co-classified pairs. As an example, in a social web a group of colleagues might end up in different modules, each corresponding to e.g., their families. In this case, the network module corresponding to their work-group is bound to become lost.

2.3.2 Preferential Attachment at the Level of Communities

In this section we examine whether the fat tails observed earlier in the community distribution could result from preferential attachment mechanisms at the level of communities. The method presented below can be applied in general to any empirical study of an attachment process where the main goal is to decide whether the attachment is uniform or preferential with respect to a certain property (e.g., degree, size, etc.) of the attached objects (e.g., nodes, communities etc.).

2.3.2.1 Method for Detecting Preferential Attachment

If the studied process is uniform with respect to a property ρ, then objects with a given ρ are chosen at a rate given by the distribution of ρ amongst the available objects. However, if the attachment mechanism prefers high (or low) ρ values, then objects with high (or low) ρ are chosen with a higher rate compared to the ρ distribution of the available objects. To monitor this enhancement, one can construct the cumulative ρ distribution $P_t(\rho)$ of the available objects at each time step t, together with the un-normalised cumulative ρ distribution of the objects chosen by the process between t and $t + 1$, denoted by $w_{t \to t+1}(\rho)$. The value of $w_{t \to t+1}(\rho^*)$ at a given ρ^* equals to the number of objects chosen in the process between t and $t + 1$, that had a ρ value larger than ρ^* at t. To detect deviations from uniform attachment, it

is best to accumulate the ratio of $w_{t\to t+1}(\rho)$ and $P_t(\rho)$ during the time evolution to obtain

$$W(\rho) = \sum_{t=0}^{t_{max}-1} \frac{w_{t\to t+1}(\rho)}{P_t(\rho)}. \qquad (2.2)$$

If the attachment is uniform with respect to ρ, then $W(\rho)$ becomes a flat function. However, if $W(\rho)$ is an increasing function, then the objects with large ρ are favoured, if it is a decreasing function, the objects with small ρ are favoured in the attachment process. The advantage of this approach is that the rate-like variable $w_{t\to t+1}(\rho)$ associated to the time step between t and $t+1$ is always compared to the $P_t(\rho)$ distribution at t. Therefore $W(\rho)$ is able to indicate preference (or the absence of preference) even when $P_t(\rho)$ is slowly changing in time (as in the case of the community degree in the co-authorship network under investigation).

We have tested the above method on simulated graphs grown with known attachment mechanisms, (i) uniform attachment (new nodes are attached to a randomly selected old node), (ii) linear preferential attachment (new nodes are attached to old ones with a probability proportional to the degree), (iii) and anti-preferential attachment (new nodes are attached to the old ones with a probability proportional to $\exp(-d)$, where d is the degree). In these cases the degree d of the individual nodes plays the role of the parameter ρ. For each time step, we recorded the cumulative degree distribution of the nodes $P_t(d)$, together with the number of nodes gaining new links with a degree higher than a given d, labelled by $w_{t\to t+1}(d)$. By summing the ratio of these two functions along the time evolution of the system one gets $W(d) = \sum_{t=0}^{t_{max}-1} w_{t\to t+1}(d)/P_t(d)$. In Fig. 2.3a. we show the empirical results for $W(d)$ obtained for the simulated networks grown with the three different attachment rules. The curves reflect the difference between the three cases very well: for the uniform attachment probability $W(d)$ is flat, for the preferential attachment $W(d)$ is clearly increasing, and for the anti-preferential attachment $W(d)$ is decreasing.

We have also calculated the attachment statistics of the nodes in the studied co-authorship network. In this case we used extremely asymmetric decay characteristics in (2.1): $\lambda_- = \infty$ and $\lambda_+ = 0$. This results in a simply growing network, where every collaboration event gives rise to a set of links between each pair of collaborators at the very moment of the collaboration act, and the strengths of these links remain constant from then on. As it can be seen in Fig. 2.3b., the corresponding $W(d)$ curve is increasing, therefore preferential attachment is present at the level of nodes in the system.

2.3.2.2 Community Growth in the Co-authorship Network

The two properties to be substituted in place of ρ in Eq.(2.2) are the community degree d^{com} and the community size s, therefore, the cumulative community size distribution $P_t(s)$ and the cumulative community degree distribution $P_t(d^{com})$ were recorded at each time step t. To study the *establishment of the new community links*,
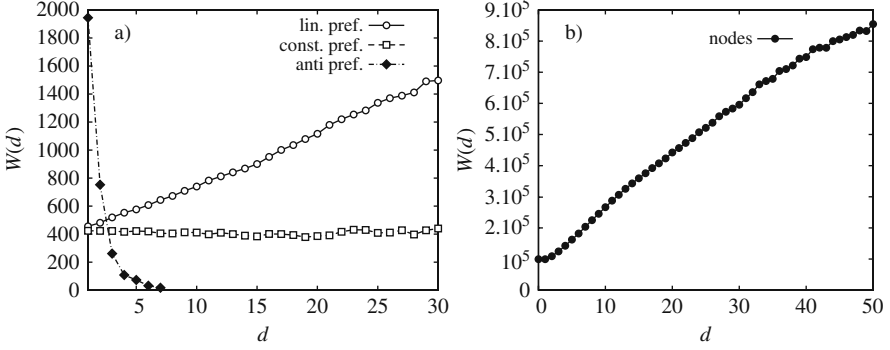
Fig. 2.3 (**a**) The $W(d)$ function for networks grown with known attachment rules: uniform probability (*squares*), linear preferential attachment (*open circles*), and anti preferential attachment (*diamonds*). (**b**) The $W(d)$ function in the co-authorship network of the Los Alamos cond-mat archive. Figure from [56]

we constructed the un-normalised cumulative size distribution $w_{t\to t+1}(s)$ and the un-normalised cumulative degree distribution $w_{t\to t+1}(d^{\text{com}})$ of the communities gaining new community links to previously unlinked communities. The value of these distributions at a given s (or given d^{com}) is equal to the number of unlinked communities at t that establish a community link between t and $t+1$ with a community larger than s (or having larger degree than d^{com}) at t. By accumulating the ratio of the rate-like variables and the corresponding distributions we obtain

$$W(s) = \sum_{t=0}^{t_{\max}-1} \frac{w_{t\to t+1}(s)}{P_t(s)}, \qquad\qquad W(d^{\text{com}}) = \sum_{t=0}^{t_{\max}-1} \frac{w_{t\to t+1}(d^{\text{com}})}{P_t(d^{\text{com}})}. \quad (2.3)$$

For the investigation of the *appearance of new members* in the communities, we recorded the un-normalised community size distribution $\widehat{w}_{t\to t+1}(s)$ and the un-normalised community degree distribution $\widehat{w}_{t\to t+1}(d^{\text{com}})$ of the communities gaining new members (belonging previously to none of the communities) between t and $t+1$. The corresponding distributions that can be used to detect deviations from the uniform attachment are

$$\widehat{W}(s) = \sum_{t=0}^{t_{\max}-1} \frac{\widehat{w}_{t\to t+1}(s)}{P_t(s)}, \qquad\qquad \widehat{W}(d^{\text{com}}) = \sum_{t=0}^{t_{\max}-1} \frac{\widehat{w}_{t\to t+1}(d^{\text{com}})}{P_t(d^{\text{com}})}. \quad (2.4)$$

In Fig. 2.4a. we show the empirical $W(s)$ and $\widehat{W}(s)$ functions, whereas in Fig. 2.4b. the empirical $W(d^{\text{com}})$ and $\widehat{W}(d^{\text{com}})$ are displayed. All four functions are clearly increasing, therefore we can draw the following important conclusions:
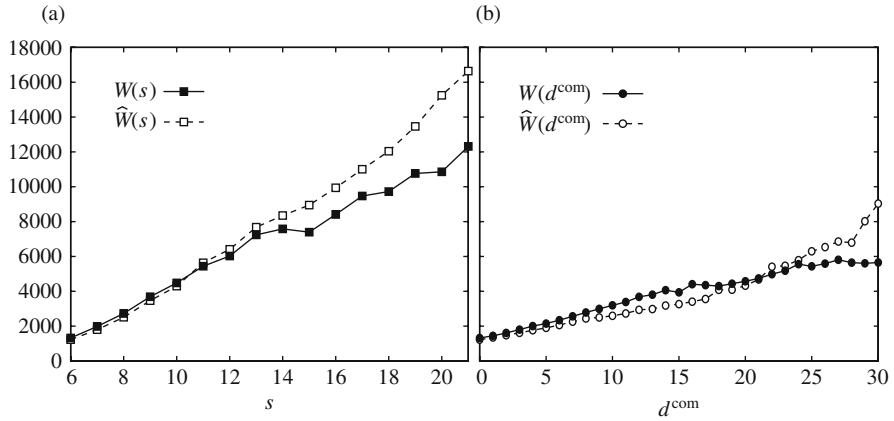
Fig. 2.4 (**a**) The $W(s)$ and $\widehat{W}(s)$ functions for the communities of the co-authorship network of the Los Alamos cond-mat e-print archive. (**b**) The $W(d^{\text{com}})$ and $\widehat{W}(d^{\text{com}})$ functions of the same network. The increasing nature of these functions indicates preferential attachment at the level of communities in the system. Figure from [56]

- When a previously unlinked community establishes a new community link, communities with large size and large degree are selected with enhanced probability from the available other communities.
- When a node previously belonging to none of the communities joins a community, communities with large size and large degree are selected with enhanced probability from the available communities.

2.3.2.3 Model for Growth of Community Network

In this section we outline a simple model for the growth of overlapping communities. Our goal is to demonstrate that preferential attachment of the nodes to communities with the community size, together with minor additional assumptions are enough for the emergence of a community system with a scaling community size and community degree distribution.

In our model the underlying network between the nodes is left unspecified, the focus is on the content of the communities. During the time evolution, similarly to the models published in [28, 48, 58], new members may join the already existing communities, and new communities may emerge as well. The new nodes introduced to the system choose their community preferentially with the community size, therefore the size distribution of the communities is expected to develop into a power-law. The appearance of the new community links originates in new nodes joining several communities at the same time. The detailed rules of the model are the following:

- The initial state of the model is a small set of communities with random size.
- The new nodes are added to the system separately. For each new node i, a membership m_i is drawn from a Poissonean distribution with an expectation value of μ.
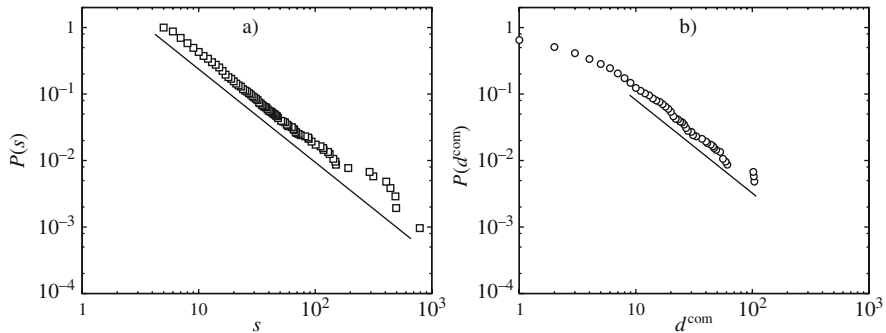
Fig. 2.5 (a) the cumulative community size distribution $P(s)$ (*open circles*) in our model at $\mu = 0.6$ follows a power-law with an exponent of -1.4 (*straight line*) (**b**) the cumulative community degree distribution $P(d^{\mathrm{com}})$ (*filled circles*) in our model at the same μ. The tail of this distribution follows the same power-law as the community size distribution (*straight line*), similarly to the communities found in the co-authorship network [53]. Figure from [56]

- If $m_i \geq 1$, communities are subsequently chosen with probabilities proportional to their sizes, until m_i is reached, and the node i joins the chosen communities simultaneously.
- If $m_i = 0$, the node i joins the group of unclassified vertices.
- When the ratio r of the group of unclassified nodes compared to the total number of nodes N exceeds a certain limit r^*, a number of q vertices from the group establish a new community. (Obviously, q must be smaller than Nr even in the initial state).

To be able to compare the results of the model with the community structure of the co-authorship network, the runs were stopped when the number of nodes in the model reached the size of the co-authorship network.

Our experience showed that the model is quite insensitive to changes in r or q, and μ is the only important parameter. For small values ($\mu < 0.3$) the resulting community degree distribution is truncated, whereas when μ is too large ($\mu > 1$), a giant community with abnormally large community degree appears. For intermediate μ values ($0.3 < \mu < 1$), the community size – and community degree distributions become fat tailed, similarly to the co-authorship network. In Fig. 2.5. we show the cumulative community size distribution $P(s)$ and the cumulative community degree distribution $P(d^{\mathrm{com}})$ of the communities obtained in our model at $\mu = 0.6$. (Changes in the parameters r and q only shifts these curves, their shape remains unchanged). Our model grasps the relevant statistical properties of the community structure in the co-authorship network [53] quite well: the community size distribution and the tail of the community degree distribution follow a power-law with the same exponent.

2.3.3 The Static Communities

Turning back to the study of the community evolution (where links corresponding to abandoned social connections may disappear with time, $0 < \lambda_- = \lambda_+ < \infty$),

the communities at each time step were extracted with the CPM for both the co-authorship and the phone-call networks. When applied to weighted networks, the CPM has two parameters: the k-clique size k, (in Fig. 2.6a, b we show the communities for $k = 4$), and the weight threshold w^* (links weaker than w^* are ignored). By increasing k or w^*, the communities start to shrink and fall apart, but at the same time they become also more cohesive. In the opposite case, at low k there is a critical w^*, under which a giant community appears in the system that smears out the details of the community structure by merging (and making invisible) many smaller communities. The criterion used to fix these parameters is based on finding a community structure as highly structured as possible: at the highest k value for which a giant community may emerge, the w^* is decreased just below the critical point. The actual values of these parameters in our studies were $k = 3$, $w^* = 0.1$ in case of the co-authorship network, and $k = 4$, $w^* = 1.0$ in case of the phone-call network.

In Fig. 2.6a, b we show the local structure at a given time step in the two networks in the vicinity of a randomly chosen individual (marked by a black frame). The communities (social groups represented by more densely interconnected parts within a network of social links) are coloured with different shades of gray, so that white nodes (and dashed edges) do not belong to any community, and those that simultaneously belong to two or more communities are shown in black. The two networks have rather different local structure: due to its bipartite nature, the collaboration network is quite dense and the overlap between communities is very significant, whereas in the phone-call network the communities are less interconnected and are often separated by one or more inter-community nodes/edges. Indeed, while the phone record captures the communication between two people, the publication record assigns to all individuals that contribute to a paper a fully connected clique. As a result, the phone data is dominated by single links, while the co-authorship
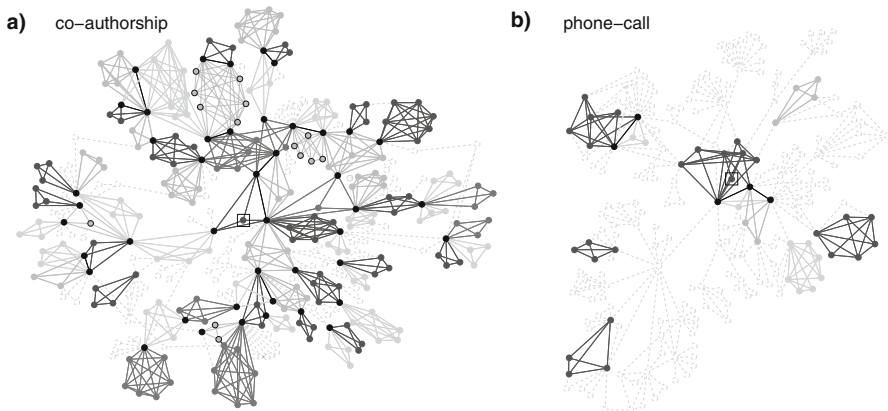

Fig. 2.6 (**a**) The local community structure at a given time step in the vicinity of a randomly selected node in case of the co-authorship network. (**b**) The same picture in the phone-call network. Figure from [52]

data has many dense, highly connected neighbourhoods. Furthermore, the links in the phone network correspond to instant communication events, capturing a relationship as is happens. In contrast, the co-authorship data records the results of a long term collaboration process. These fundamental differences suggest that any potential common features of the community evolution in the two networks potentially, represent generic characteristics of community formation, rather than being rooted in the details of the network representation or data collection process.

2.3.4 Validating the Communities

When validating the found communities, as a first step, it is important to check if the uncovered communities correspond to groups of individuals with a shared common activity pattern. For this purpose we compared the average weight of the links inside communities, w_c, to the average weight of the inter-community links, w_{ic}. For the co-authorship network w_c/w_{ic} is about 2.9, while for the phone-call network the difference is even more significant, since $w_c/w_{ic} \simeq 5.9$, indicating that the intensity of collaboration/communication within a group is significantly higher than with contacts belonging to a different group [8, 21, 50, 51].

While for coauthors the quality of the clustering can be directly tested by studying their publication records in more detail, in the phone-call network personal information is not available. In this case the zip-code and the age of the users provides additional information for checking the homogeneity of the communities. In Fig. 2.7a we show the size of the largest subset of people having the same zip code in the communities, $\langle n_{\text{real}} \rangle$, averaged over the time steps, as the function of the community size s, divided by $\langle n_{\text{rand}} \rangle$, representing the average over random sets of users. The significantly higher number of people with the same zip-code in the CPM
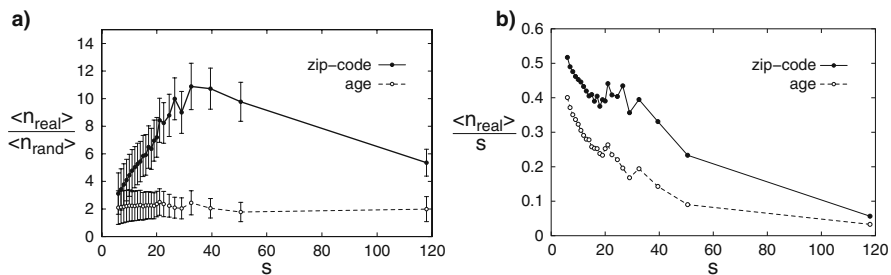

Fig. 2.7 (**a**) The *black symbols* correspond to the average size of the largest subset of members with the same zip-code, $\langle n_{\text{real}} \rangle$, in the phone-call communities divided by the same quantity found in random sets, $\langle n_{\text{rand}} \rangle$, as the function of the community size s. Similarly, the *white symbols* show the average size of the largest subset of community members with an age falling in a three year time window, divided by the same quantity in random sets. The error-bars in both cases correspond to $\langle n_{\text{real}} \rangle / (\langle n_{\text{rand}} \rangle + \sigma_{\text{rand}})$ and $\langle n_{\text{real}} \rangle / (\langle n_{\text{rand}} \rangle - \sigma_{\text{rand}})$, where σ_{rand} is the standard deviation in case of the random sets (**b**) The $\langle n_{\text{real}} \rangle / s$ as a function of s, for both the zip-code (*black symbols*) and the age (*white symbols*). Figure from [52]

communities as compared to random sets indicates that the communities usually correspond to individuals living relatively close to each other. It is of specific interest that $\langle n_{real}\rangle / \langle n_{rand}\rangle$ has a prominent peak at $s \simeq 35$, suggesting that communities of this size are geographically the most homogeneous ones. However, as Fig. 2.7b shows, the situation is more complex: on average, the smaller communities are more homogeneous, but there is still a noticeable peak at $s \simeq 30 - 35$. In Fig. 2.7a we also show the average size of the largest subset of members with an age falling into a three years wide time window, divided by the same quantity obtained for randomly selected groups of individuals. The fact that the ratio is larger than one indicates that communities have a tendency to contain people from the same generation, and the $\langle n_{rand}\rangle / s$ plot indicates that the homogeneity of small groups is on average larger than that of the big groups.

Another interesting feature of Fig. 2.7 is that the difference in the homogeneity of the age is less pronounced than in case of the zip-code. A plausible reason for this effect is that due to the strong social relation between parents and children, many communities contain members coming from different generations. This is supported by the distribution of the age difference in communities, shown in Fig. 2.8a: there is a major peak at zero corresponding to members with the same age, however there is also another peak at 25, corresponding to the typical age difference between parents and children.

Beside the zip-code and the age, the statistics of the *service usage* of the customers supports the validity of the communities as well. In our primary data, the number of times people have used a certain service in one of the two weeks long periods was also available. (There were altogether 34 available services for the customers). However, for most services, the probability for a randomly selected customer using the service at all is very low. For this reason, instead of comparing the average number of members using the same service in communities and random
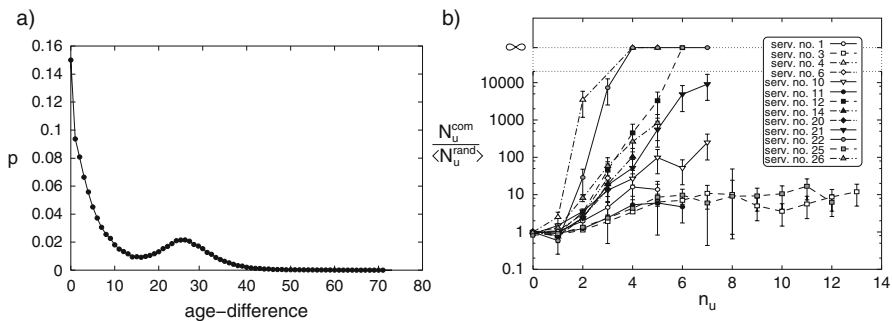

Fig. 2.8 (**a**) The probability distribution of the age difference between community members in the phone-call network. The most probable values are zero and 25, indicating that a pair of members from a community are most likely to be of the same age, or to be a generation apart from each other. (**b**) The number of communities divided by the average number of random sets containing the same n_u number of people using a given service. Each sample of the random sets was prepared with size distribution of the communities determined for the phone-call network. Figure from the Suppl. of [52]

sets, we compared the $N_u^{com}(n_u)$ number of communities having n_u members using the same service to the same quantity in random sets, denoted by $N_u^{rand}(n_u)$. For each service, random sets with the same size distribution as the communities were constructed 10,000 times, and $N_u^{rand}(n_u)$ was averaged over the samples. As it can be seen from Fig. 2.8b, for 13 services the $N_u^{com}(n_u)$ number of communities having n_u members using the service is significantly larger than in case of random sets. In fact, the $N_u^{com}(n_u)/N_u^{rand}(n_u)$ ratio in some cases reaches infinity, indicating that there were no random sets at all containing such high number of service users as some communities.

In summary, the phone-call communities uncovered by the CPM tend to contain individuals living in the same neighbourhood, and with comparable age, a homogeneity that supports the validity of the uncovered community structure.

2.4 Evolving Communities

Our focus is on the statistical properties of evolving communities, therefore, we need a reliable method for matching the static "snap-shots" of the community structure at subsequent time steps. The basic events that may occur in the life of a community are shown in Fig. 2.9a: a community can grow by recruiting new members, or contract by loosing members; two (or more) groups may merge into a single community, while a large enough social group can split into several smaller ones; new communities are born and old ones may disappear.

Given the huge number of groups present at each time step, it is a significant algorithmic and computational challenge to match communities uncovered at different time steps. The fact that the communities obtained by the CPM can have overlaps makes the problem even more complicated.

A simple approach would be to match communities from consecutive time steps in descending order of their relative overlap. The relative overlap between communities A and B can be defined as
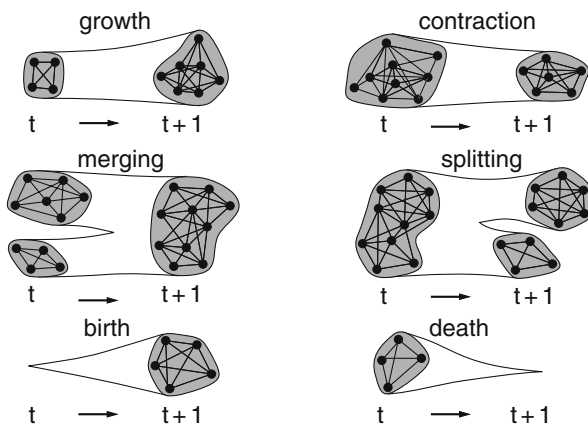

Fig. 2.9 Possible events in the community evolution. When new members are introduced, the community grows, whereas leaving members cause decay in the size. Communities can merge and split, new groups may emerge and old ones can disappear. Figure from [52]

$$C(A, B) \equiv \frac{|A \cap B|}{|A \cup B|}, \tag{2.5}$$

where $|A \cap B|$ is the number of common nodes in A and B, and $|A \cup B|$ is the number of nodes in the union of the two communities. However, the nodes shared between the communities can undermine this type of community conjugation between consecutive time steps: In case a small community A is inflated by large magnitude between time steps t and $t + 1$, and at $t + 1$ it overlaps with a small static community $B = B_t = B_{t+1}$, then the relative overlap (2.5) between A_{t+1} and B_t can be larger than the relative overlap between A_{t+1} and A_t.

To overcome this difficulty, we refine the identification of communities as shown in Fig. 2.10. For each consecutive time steps t and $t + 1$ we construct a joint graph consisting of the union of links from the corresponding two networks, and extract the CPM community structure of this joint network (we thank I. Derényi for pointing out this possibility). When new links are introduced in a network, the CPM communities may remain unchanged, they may grow, or a group of CPM communities may become joined into a single community, however no CPM community may decay by loosing members. From this it follows that if we merge two networks, any CPM community in any of the original networks will be contained in exactly one community in the joined network.

Let us denote the set of communities from t by $\mathbf{A}$, the set of communities from $t + 1$ by $\mathbf{B}$, and the set of communities from the joint network by $\mathbf{V}$. For any community
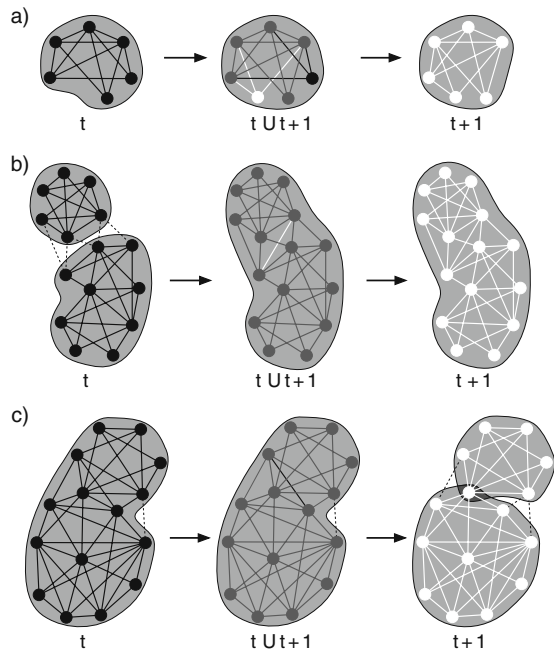

Fig. 2.10 Simple scenarios in the community evolution of the phone-call network for $k = 4$. The communities at t are coloured *black on light gray*, the communities at $t + 1$ are coloured *white on light gray*, and the communities in the joint network are coloured *dark gray on light gray*. (**a**) a community simply "propagates", (**b**) the larger community swallows the smaller one, (**c**) a small community is detached from a larger one. Figure from the Suppl. of [52]

$A_i \in \mathbf{A}$ or $B_j \in \mathbf{B}$ we can find exactly one community $V_k \in \mathbf{V}$ containing it. When matching the communities in $\mathbf{A}$ and in $\mathbf{B}$, first for every community $V_k \in \mathbf{V}$ in the joint system we extract the list of communities $A_i^k \in \mathbf{A}$ and $B_j^k \in \mathbf{B}$ that are contained in V_k (this means $A_i^k \subseteq V_k$ and $B_j^k \subseteq V_k$). (Note that either of the lists may be empty). Then the relative overlap between every possible (A_i^k, B_j^k) pairs can be obtained as

$$C_{ij}^k = \frac{\left| A_i^k \cap B_j^k \right|}{\left| A_i^k \cup B_j^k \right|}, \tag{2.6}$$

and we match the pairs of communities in descending order of their relative overlap.

As an illustration of the above process, in Fig. 2.10 we show three simple scenarios occurring in the community evolution of the phone-call network. In Fig. 2.10a both lists A_i^k and B_j^k consist of only a single community, therefore these can be matched right away. However, in Fig. 2.10b the A_i^k list contains two elements, let us denote the smaller community of size $s = 6$ at t by A_1^k and the larger community consisting of nine nodes at t by A_2^k. The corresponding B_j^k list contains a single community B_1^k having 15 members. The relative overlaps between the communities are given as $C_{1,1}^k = 2/5$ and $C_{2,1}^k = 3/5$. Since the $C_{2,1}^k$ relative overlap of the B_1^k community with A_2^k community is larger than the $C_{1,1}^k$ relative overlap with A_1^k, we assign B_1^k to A_2^k. As a consequence the A_1^k community comes to the end of its life at t, and it is swallowed by A_2^k. The opposite process is shown in Fig. 2.10c: in this case the A_i^k list consists of a single community A_1^k of size $s = 15$, whereas the B_j^k list has two elements, the community with six members labelled by B_1^k, whereas and the community containing ten nodes labelled by B_2^k. The relative overlaps are $C_{1,1}^k = 2/5$ and $C_{1,2}^k = 2/3$, therefore the A_1^k is matched to B_2^k, and B_1^k is treated as a new born community. In general, whenever the community V_k contains more communities from $\mathbf{A}$ than from $\mathbf{B}$, the communities A_i^k left with no counterpart from B_j^k finish their life's at t, and when V_k contains more communities from $\mathbf{B}$ than from $\mathbf{A}$, the communities B_j^k left with no counterpart from A_i^k are considered as new born communities.

In some cases we can observe that although a community was disintegrated, after a few steps it suddenly reappears in the network. Our conjecture is that this is more likely to be the consequence of a temporally lower publishing-rate/calling-rate of the people in question than of the real disassembly and re-assembly of the corresponding social community between the people. Therefore, whenever a newborn community includes a formerly disintegrated one, then the last state of the old community is elongated to fill the gap before the reappearance, and the newborn community is treated as the continuation of the old one, as shown in Fig. 2.11.
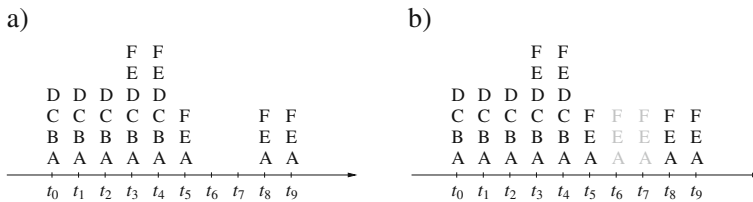
a)

```
              F  F
              E  E
    D  D  D  D
    C  C  C  C  C  F              F  F
    B  B  B  B  B  E              E  E
    A  A  A  A  A  A           A  A
   ───────────────────────────────────────────►
    t₀ t₁ t₂ t₃ t₄ t₅ t₆ t₇ t₈ t₉
```

b)

```
              F  F
              E  E
    D  D  D  D
    C  C  C  C  C  F  F  F  F  F
    B  B  B  B  B  E  E  E  E  E
    A  A  A  A  A  A  A  A  A  A
   ───────────────────────────────────────────►
    t₀ t₁ t₂ t₃ t₄ t₅ t₆ t₇ t₈ t₉
```

Fig. 2.11 (**a**) A community is disintegrated after step t_5, and it is reborn at step t_8. (**b**) We treat the community as if it was alive at steps t_6 and t_7 too, with the same nodes as at step t_5. Figure from the Suppl. of [52]

2.5 Statistical Properties of the Community Dynamics

2.5.1 Basic Statistics

One of the most basic properties characterising the partitioning of a network is the overall coverage of the community structure, i.e. the ratio of nodes contained in at least one community. In case of the co-authorship network the average value of this ratio was above 59%, which is a reasonable coverage for the CPM. In contrast, we could only achieve a significantly smaller ratio for the phone-call network. At such a large system size, in order to be able to match the communities at subsequent time steps in reasonable time we had to decrease the number of communities by choosing a higher k and w^* parameter ($k = 4$ and $w^* = 1.0$), and keeping only the communities having a size larger or equal to $s = 6$. Therefore, in the end the ratio of nodes contained in at least one community was reduced to 11%. However, this still means more than 400,000 customers in the communities on average, providing a representative sampling of the system. By lowering the k to $k = 3$, the fraction of nodes included in the communities is raised to 43%. Furthermore, a significant number of additional nodes can be also classified into the discovered communities. For example, if a node not yet classified has link(s) only to a single community (and, if it has no links connecting to nodes in any other community) it can be safely added to that community. Carrying out this process iteratively, the fraction of nodes that can be classified into communities increases to 72% for the k = 3 co-authorship network, and to 72% (61%) for the k = 3 (k = 4) mobile phone network, which, in principle, allows us to classify over 2.4 million users into communities.

Another important statistics describing the community system is the community size distribution. In Fig. 2.12a we show the community size distribution in the phone-call network at different time steps. They all resemble to a power-law with a high exponent. In case of $t = 0$, the largest communities are somewhat smaller than in the later time steps. This is due to the fact that the events before the actual time step cannot contribute to the link-weights in case of $t = 0$, whereas they can if $t > 0$. In Fig. 2.12b we can follow the time evolution of the community size distribution in the co-authorship network. In this case $t = 0$ corresponds to the birth of the system itself as well (whereas in case of the phone-calls it does not), therefore the network
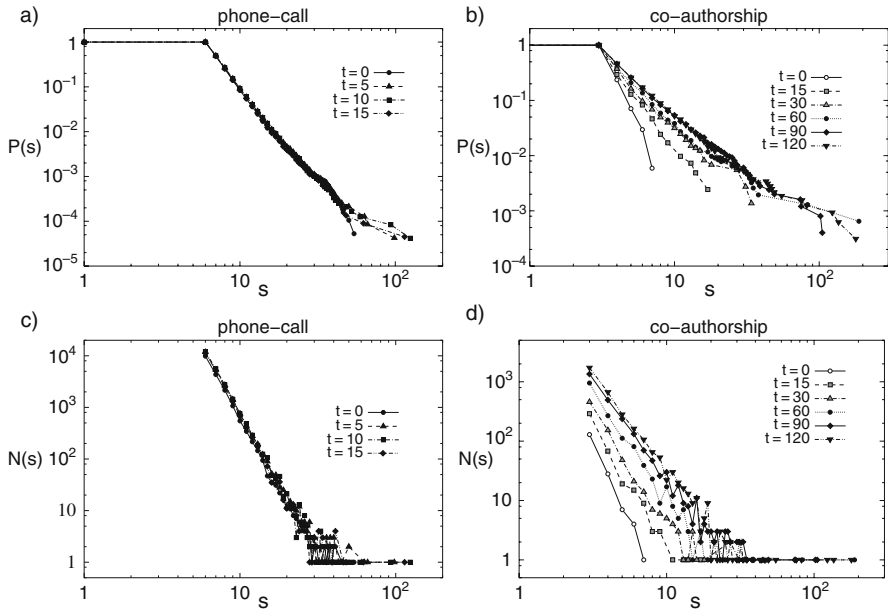
Fig. 2.12 (**a**) The cumulative community size distribution in the phone-call network at different time steps. (**b**) The time evolution of the cumulative community size distribution in the co-authorship network. (**c**) The number of communities of a given size at different time steps in the phone-call network. (**d**) The time evolution of the number of communities with a given size in the co-authorship network. Figure from [55]

and the communities in the network are small in the first few time steps. Later on, the system is enlarged, and the community size distribution is stabilised close to a power-law. In Fig. 2.12c, d we show the number of communities as a function of the community size at different time steps in the examined systems. For the phone-call network (Fig. 2.12c), this distribution is more or less constant in time. In contrast, (due to the growth of the underlying network) we can see an overall growth in the number of communities with time in the co-authorship network (Fig. 2.12d). Since the number of communities drops down to only a few at large community sizes in both systems, we used size binning when calculating the statistics shown in Figs. 2.13, 2.14, and 2.17.

As for evolving communities, we first consider two basic quantities characterising a community: its size s and its age τ, representing the time passed since its birth. s and τ are positively correlated: larger communities are on average older (Fig. 2.13a), which is quite natural, as communities are usually born small, and it takes time to recruit new members to reach a large size.

Next we used the auto-correlation function, $C(t)$, to quantify the relative overlap between two states of the same community $A(t)$ at t time steps apart:

$$C_A(t) \equiv \frac{|A(t_0) \cap A(t_0 + t)|}{|A(t_0) \cup A(t_0 + t)|}, \tag{2.7}$$
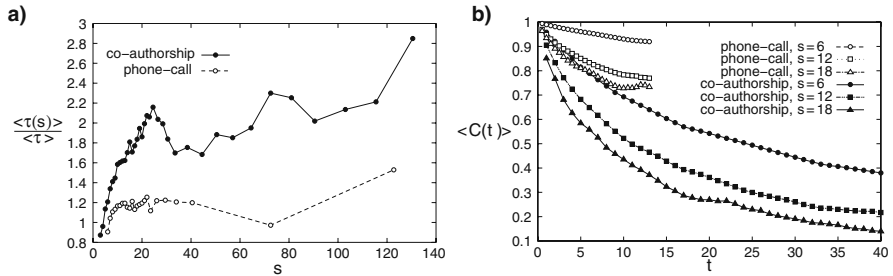
Fig. 2.13 (**a**) The average age τ of communities with a given size (number of people) s, divided by the average age of all communities $\langle\tau\rangle$, as the function of s, indicating that larger communities are on average older. (**b**) The average auto-correlation function $C(t)$ of communities with different sizes (the unit of time, t, is one month). The $C(t)$ of larger communities decays faster. Figure from [52]
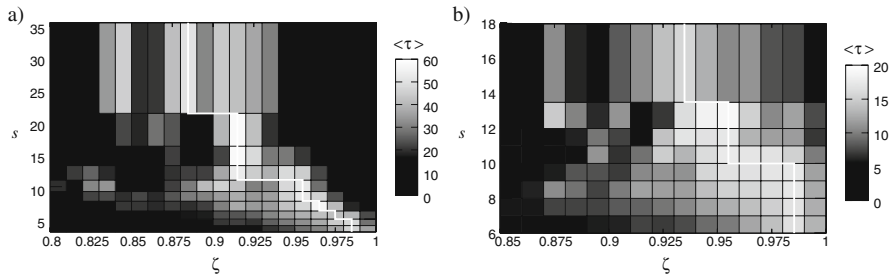

Fig. 2.14 (**a**) The average life-span $\langle\tau^*\rangle$ of the communities as the function of the stationarity ζ and the community size s for the co-authorship network. The peak in $\langle\tau^*\rangle$ is close to $\zeta = 1$ for small sizes, whereas it is shifted towards lower ζ values for large sizes. (**b**) Similar results found in the phone-call network. Figure from [52]

where $|A(t_0) \cap A(t_0 + t)|$ is the number of common nodes (members) in $A(t_0)$ and $A(t_0 + t)$, and $|A(t_0) \cup A(t_0 + t)|$ is the number of nodes in the union of $A(t_0)$ and $A(t_0 + t)$. Figure 2.13b shows the average time dependent auto-correlation function for communities born with different sizes. We find that in both networks, the auto-correlation function decays faster for the larger communities, indicating that the membership of the larger communities is changing at a higher rate. On the contrary, small communities change at a smaller rate, their composition being more or less static.

2.5.2 Stationarity and Lifetime

According to the results of Sect.2.5.1 a difference can be observed in the versatility of small and large communities. To quantify this aspect of community evolution, we define the *stationarity* ζ of a community as the average correlation between subsequent states:

$$\zeta \equiv \frac{\sum_{t=t_0}^{t_{max}-1} C(t, t+1)}{t_{max} - t_0 - 1}, \tag{2.8}$$

where t_0 denotes the birth of the community, and t_{max} is the last step before the extinction of the community. In other words, $1 - \zeta$ represents the average ratio of members changed in one step; larger ζ corresponds to smaller change (more stationary membership).

We observe a very interesting effect when we investigate the relationship between the lifetime τ^* (the number of steps between the birth and disintegration of a community), the stationarity and the community size. The lifetime can be viewed as a simple measure of "fitness": communities having higher fitness have an extended life, while the ones with small fitness quickly disintegrate, or are swallowed by another community. In Fig. 2.14a, b we show the average life-span $\langle \tau^* \rangle$ as a function of the stationarity ζ and the community size s (both s and ζ were binned). In both networks, for small community sizes the highest average life-span is at a stationarity value very close to one, indicating that for small communities it is optimal to have static, time independent membership. On the other hand, the peak in $\langle \tau^* \rangle$ is shifted towards low ζ values for large communities, suggesting that for these the optimal regime is to be dynamic, i.e., a continually changing membership. In fact, large communities with a ζ value equal to the optimal ζ for small communities have a very short life, and similarly, small communities with a low ζ (being optimal at large sizes) are disappearing quickly as well.

To illustrate the difference in the optimal behaviour (a pattern of membership dynamics leading to extended lifetime) of small and large communities, in Fig. 2.15. we show the time evolution of four communities from the co-authorship network. As Fig. 2.15. indicates, a typical small and stationary community undergoes minor changes, but lives for a long time. This is well illustrated by the snapshots of the community structure, showing that the community's stability is conferred by a core of three individuals representing a collaborative group spanning over 52 months. While new co-authors are added occasionally to the group, they come and go. In contrast, a small community with high turnover of its members, (several members abandon the community at the second time step, followed by three new members joining in at time step three) has a lifetime of nine time steps only (Fig. 2.15b). The opposite is seen for large communities: a large stationary community disintegrates after four time steps (Fig. 2.15c). In contrast, a large non-stationary community whose members change dynamically, resulting in significant fluctuations in both size and the composition, has quite extended lifetime (Fig. 2.15d). Indeed, while the community undergoes dramatic changes, gaining (Fig. 2.15e) or loosing a high fraction of its membership, it can easily withstand these changes.

2.5.3 Predicting Community Break Up

The quite different stability rules followed by the small and large communities raise an important question: could an inspection of the community itself predict its future?
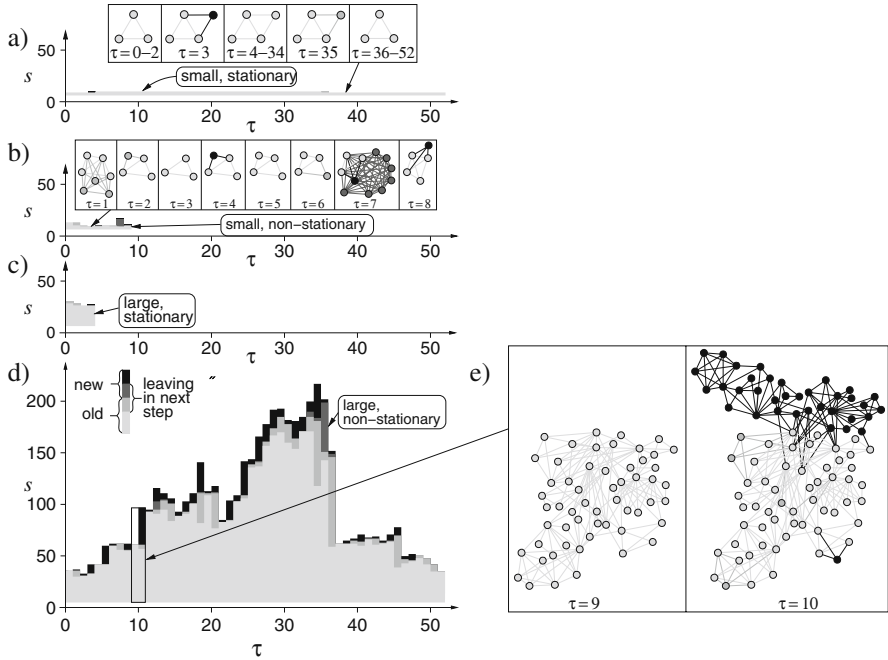
Fig. 2.15 Time evolution of four communities in the co-authorship network. The height of the columns corresponds to the actual community size, and within one column the *light gray colour* indicates the number of "old" nodes (that have been present in the community at least in the previous time step as well), while newcomers are shown with *black*. The members abandoning the community in the next time step are shown with *mid gray colours*, the shade depending on whether they are old or new. (This latter type of member joins the community for only one time step). From *top to bottom*, we show a small and stationary community (**a**), a small and non-stationary community (**b**), a large and stationary community (**c**) and, finally, a large and non-stationary community (**d**). A mainly growing stage (two time steps) in the evolution of the latter community is detailed in panel (**e**). Figure from [52]

To address this question, for each member in a community we measured the total weight of this member's connections to outside of the community (w_{out}) as well as to members belonging to the same community (w_{in}). We then calculated the probability that the member will abandon the community as a function of the $w_{out}/(w_{in} + w_{out})$ ratio. As Fig. 2.16a shows, for both networks this probability increases monotonically, suggesting that if the relative commitment of a user is to individuals outside a given community is higher, then it is more likely that he/she will leave the community.

In parallel, the average time spent in the community by the nodes, $\langle \tau_n \rangle$, is a decreasing function of the above ratio (Fig. 2.16a inset). Individuals that are the most likely to stay are those that commit most of their time to community members, an effect that is particularly prominent for the phone network. As Fig. 2.16a shows, those with the least commitment have a quickly growing likelihood of leaving the community.
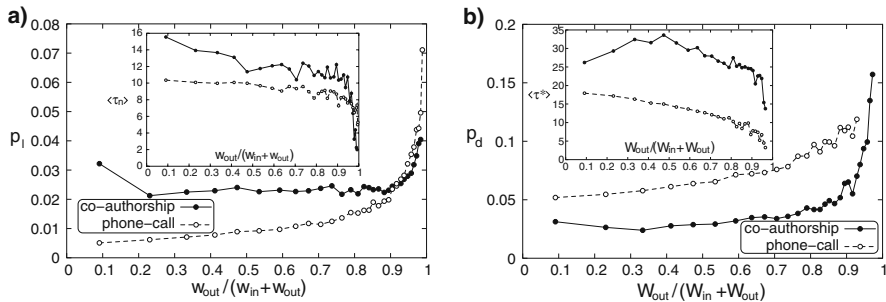
Fig. 2.16 (**a**) The probability p_ℓ for a member to abandon its community in the next step as a function of the ratio of its aggregated link weights to other parts of the network (w_{out}) and its total aggregated link weight ($w_{in} + w_{out}$). The *inset* shows the average time spent in the community by the nodes, $\langle \tau_n \rangle$, in function of $w_{out}/(w_{in} + w_{out})$. (**b**) The probability p_d for a community to disintegrate in the next step in function of the ratio of the aggregated weights of links from the community to other parts of the network (W_{out}) and the aggregated weights of all links starting from the community ($W_{in} + W_{out}$). The *inset* shows the average life time $\langle \tau^* \rangle$ of communities as a function of $W_{out}/(W_{in} + W_{out})$. Figure from [52]

Taking this idea from individuals to communities, we measured for each community the total weight of links (a measure of how much a member is committed) from the members to others, outside of the community (W_{out}), as well as the aggregated link weight inside the community (W_{in}). We find that the probability for a community to disintegrate in the next step increases as a function of $W_{out}/(W_{in} + W_{out})$ (Fig. 2.16b), and the lifetime of a community decreases with the $W_{out}/(W_{in} + W_{out})$ ratio (Fig. 2.16b inset). This indicates that self-focused communities have a significantly longer lifetime than those that are open to the outside world. However, an interesting observation is that, while the lifetime of the phone-call communities for moderate levels is relatively insensitive to outside commitments, the lifetime of the collaboration communities possesses a maximum at intermediate levels of inter-collaborations (collaboration between colleagues who belong to different communities). These results suggest that a tracking of the individual's as well as the community's relative commitment to the other members of the community provides a clue for predicting the community's fate.

2.5.4 Merging of Communities

Finally, we investigate a special aspect of the merging process between communities. During such event, a pair (or a larger group) of initially distinct communities join together and form a single community. A very interesting question connected to this is that can we find a simple relation between the size of a community and the likelihood that it will take part in such process?

To investigate this issue we carried out measurements similar to those in [56] and presented in Sect. 2.3.2.1. The basic idea is that if the merging process is uniform

with respect to the size of the communities s, then communities with a given s are chosen at a rate given by the size distribution of the available communities. However, if the merging mechanism prefers large (or small) sizes, then communities with large (or small) s are chosen with a higher rate compared to the size distribution of the available communities. To monitor this enhancement we used the indicator function, defined in Eq. (2.2), substituting the $\rho = (s_1, s_2)$ size-pair object. At each time step t the cumulative size-pair distribution $P_t(s_1, s_2)$ was recorded. Simultaneously, the un-normalised cumulative size-pair distribution of the communities merging between t and $t+1$ was constructed; we shall denote this distribution by $w_{t \to t+1}(s_1, s_2)$. The value of this rate-like variable $w_{t \to t+1}(s_1^*, s_2^*)$ at a given value of s_1^* and s_2^* is equal to the number of pairs of communities that merged between t and $t+1$ and had sizes $s_1 > s_1^*$ and $s_2 > s_2^*$. Here the resulting indicator function

$$W(s_1, s_2) \equiv \sum_{t=0}^{t_{\max}-1} \frac{w_{t \to t+1}(s_1, s_2)}{P_t(s_1, s_2)} \tag{2.9}$$

is defined on a two dimensional plane. When the merging process is uniform with respect to the community size the $W(s_1, s_2)$ becomes a flat function: on average we see pairs of communities merging with sizes s_1 and s_2 at a rate equal to the probability of finding a pair of communities of these sizes. However, if the merging process prefers large (or small) communities, than pairs with large (or small) sizes merge at a higher rate than the probability of finding such pairs, and $W(s_1, s_2)$ becomes increasing (or decreasing) with the size.

The reason for using un-normalised $w_{t \to t+1}(s_1, s_2)$ distributions is that in this way each merging event contributes to $W(s_1, s_2)$ with equal weight, and the time steps with a lot of merging events count more than those with only a few events. In the opposite case (when $w_{t \to t+1}(s_1, s_2)$ is normalised for each pairs of subsequent time steps $t, t+1$), the merging events occurring between time steps with a lot of other merging events are suppressed compared to the events with only a few other parallel events, as each pairs of consecutive time steps $t, t+1$ contribute to the $W(s_1, s_2)$ function with equal weights. This difference between normalised and un-normalised $w_{t \to t+1}(s_1, s_2)$ becomes important in case of the co-authorship network, where in the beginning the system is small and merging is rare, and later on as the system is developing, merging between communities becomes a regular event.

In Fig. 2.17. we show $W(s_1, s_2)$ for both networks, and the picture suggests that large sizes are preferred in the merging process. This is consistent with our findings that the content of large communities is changing at a faster rate compared to the small ones. Swallowing other communities is an efficient way to bring numerous new members into the community in just one step, therefore taking part in merging is beneficial for large communities following a survival strategy based on constantly changing their members.

Another interesting aspect of the results shown in Fig. 2.17. is that they are analogous to the attachment mechanism of links between already existing nodes in collaboration networks [5]: the probability for a new link to appear between two
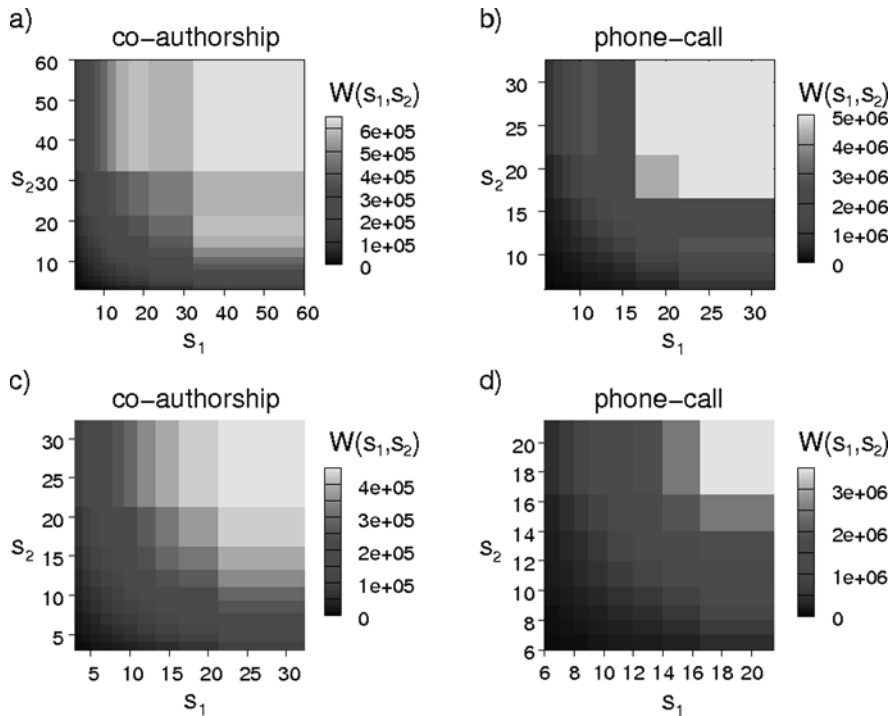
Fig. 2.17 The merging of communities. (**a**) the $W(s_1, s_2)$ function for the co-authorship network, (**b**) the $W(s_1, s_2)$ function for the phone-call network, (**c**) the region with smaller $W(s_1, s_2)$ in (**a**) enlarged, (**d**) the region with smaller $W(s_1, s_2)$ in (**b**) enlarged. Figure from the Suppl. of [52]

nodes with degree d_1 and d_2 is roughly proportional to $d_1 \times d_2$. Similarly, the probability that two communities of sizes s_1 and s_2 will merge is proportional to $s_1 \times s_2$, therefore the large communities attract each other in a similar manner to hubs in collaboration networks.

2.6 Conclusion

In this chapter we investigated the statistical properties of community dynamics in two large social networks. Due to the frequent changes in the communication/collaboration patterns between individuals, the communities corresponding to groups of mobile phone users or collaborating scientists are under constant evolution. In case of a simple growing scenario, we found that similar processes control the development of the system at different levels in the hierarchy, as the growth of the communities, the development of the community graph and the growth of the underlying network are all driven by preferential attachment.

When deletion of the links is taken into account as well, the picture gets more complex. In order to be able to track the intricate merging, splitting, growth, decay,

etc. of the investigated social groups, we developed an algorithm based on the CPM for matching the communities extracted at subsequent time steps. According to our results, a significant difference can be observed between smaller collaborative or friendship circles and institutions when subjected to the processes above. At the heart of small cliques are a few strong relationships, and as long as these persist, the community around them is stable. In other words, small groups can persist for a long time if their membership is constant. It appears to be almost impossible to maintain this strategy for large communities, however. Thus we find that the condition for stability for large communities is continuous changes in their membership, allowing for the possibility that after some time practically all members are exchanged. Such loose, rapidly changing communities are reminiscent of institutions, that can continue to exist even after all members have been replaced by new members. For example, in a few years most members of a school or a company could change, yet the school and the company will be detectable as a distinct community at any time step during its existence.

We also showed that the knowledge of the time commitment of the members to a given community can be used for predicting the community's lifetime. Furthermore, we found that the likelihood of merging between communities is increasing with the community size. These findings offer a new view on the fundamental differences between the dynamics of small groups and large institutions.

References

1. Aiello, W., Chung, F., Lu, L.: A random graph model for massive graphs. In: Proc. 32nd ACM Symp. on the Theory Comput., pp. 171–180, ACM, New York (2000)
2. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**, 47–97 (2002)
3. Antonov, A.V., Mewes, H.W.: Complex functionality of gene groups identified from high-throughput data. J. Mol. Biol. **363(1)**, 289–296 (2006)
4. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
5. Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Physica A **311**, 590–614 (2002)
6. Batagelj, V., Zaveršnik, M.: Short cycle connectivity. Discrete Math. **307**, 310–318 (2007)
7. Baumes, J., Goldberg, M., Magdon-Ismail, M.: Efficient identification of overlapping communities. Lect. Notes Comput. Sci. **3495**, 27–36 (2005)
8. Csermely, P.: Weak Links. Springer Verlag, Heidelberg, Germany (2006)
9. Derényi, I., Palla, G., Vicsek, T.: Clique percolation in random networks. Phys. Rev. Lett. **94**, 160202 (2005)
10. Dodds, P.S., Muhamad, R., Watts, D.J.: An experimental study of search in global social networks. Science **301**, 827–829 (2003)
11. Ebel, H., Davidsen, J., Bornholdt, S.: Dynamics of social networks. Complexity **8**, 24–27 (2002)
12. Ebel, H., Mielsch, L.I., Bornholdt, S.: Scale-free topology of e-mail networks. Phys. Rev. E **66**, 35103(R) (2002)
13. Eckmann, J.P., Moses, E., Sergi, D.: Entropy of dialogues creates coherent structures in e-mail traffic. Proc. Natl. Acad. Sci. USA **101**, 14333–14337 (2004)
14. Everett, M.G., Borgatti, S.P.: Analyzing clique overlap. Connections **21**, 49–61 (1998)

15. Everitt, B.S.: Cluster Analysis, 3th edn. Edward Arnold, London (1993)
16. Farkas, I.J., Ábel, D., Palla, G., Vicsek, T.: Weighted network modules. New J. Phys. **9**, 180 (2007)
17. Fortunato, S., Barthelemy, M.: Resolution limit in community detection. Proc. Natl. Acad. Sci. USA **104**, 36–41 (2007)
18. Gfeler, D., Chappelier, J.C., Rios, P.D.L.: Finding instabilities in the community structure of complex networks. Phys. Rev. E **72**, 056135 (2005)
19. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA **99**, 7821–7826 (2002)
20. Granovetter, M.: Decision Making: Alternatives to Rational Choice Models Economic Action and Social Structure: The Problem of Embeddedness. SAGE, Newbury Park, CA (1992)
21. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. **78**, 1360–1380 (1973)
22. Grossman, J.W.: The evolution of the mathematical research collaboration graph. Congressus Numerantium **158**, 202–212 (2002)
23. Grossman, J.W., Ion, P.D.F.: On a Portion of the well-known collaboration graph. Congressus Numerantium **108**, 129–131 (1995)
24. Guimerà, R., Amaral, L.A.N.: Functional cartography of complex metabolic networks. Nature **433**, 895–900 (2005)
25. Guimerá, R., Danon, L., Diaz-Guilera, A., Giralt, F., Arenas, A.: Self-similar community structure in organisations. Phys. Rev. E **68**, 065103 (2003)
26. Guimerà, R., Mossa, S., Turtschi, A., Amaral, L.A.N.: The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proc. Natl. Acad. Sci. USA **102**, 7794–7799 (2005)
27. Guimerà, R., Sales-Pardo, M., Amaral, L.A.N.: Module identification in bipartite and directed networks. Phys. Rev. E **76**, 036102 (2007)
28. Guimerá, R., Uzzi, B., Spiro, J., Amaral, L.A.N.: Team assembly mechanisms determine collaboration network structure and team performance. Science **308**, 697–702 (2005)
29. Guldener, U., Munsterkotter, M., Kastenmuller, G., Strack, N., van Helden, J.: CYGD: The comprehensive yeast genome database. Nucl. Acad. Res. **33**, D364–D368 (2005)
30. Heimo, T., Saramäki, J., Onnela, J.P., Kaski, K.: Spectral and network methods in the analysis of correlation matrices of stock returns. Physica A-Statist. Mech. Appl. **383**, 147–151 (2007)
31. Holme, P., Edling, C.R., Liljeros, F.: Structure and time-evolution of an internet dating community. Soc. Networks **26**, 155–174 (2004)
32. Hopcroft, J., Khan, O., Kulis, B., Selman, B.: Tracking evolving communities in large linked networks. Proc. Natl. Acad. Sci. USA **101**, 5249–5253 (2004)
33. Knudsen, S.: A Guide to Analysis of DNA Microarray Data, 2nd edn. Wiley-Liss, New York (2004)
34. Kossinets, G., Watts, D.J.: Empirical analysis of an evolving social network. Science **311**, 88–90 (2006)
35. Krogan, N.J., Cagney, G., Yu, H.Y., Zhong, G.Q., Guo, X.H., Ignatchenko, A., Li, J., Pu, S.Y., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Orsi, S.W.C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., Onge, P.S., Ghanny, S., Lam, M.H.Y., Butland, G., Altaf-Ui, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., Greenblatt, J.F.: Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature **440**, 637–643 (2006)
36. Kumpula, J.M., Saramäki, J., Kaski, K., Kertész, J.: Limited resolution in complex network community detection with Potts model approach. Eur. Phys. J. B **56**, 41–45 (2007)
37. Li, C., Maini, P.K.: An evolving network model with community structure. J. Phys. A: Math. Gen. **38**, 9741–9749 (2005)

38. Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Aberg, Y.: The web of human sexual contacts. Nature **411**, 907–908 (2001)
39. Mendes, J.F.F., Dorogovtsev, S.N.: Evolution of networks: From biological nets to the Internet and WWW. Oxford University Press, Oxford (2003)
40. Nepusz, T., Petróczi, A., Négyessy, L., Bazsó, F.: Fuzzy communities and the concept of bridgeness in complex networks. Phys. Rev. E **77**, 016107 (2008)
41. Newman, M.E.J.: Phys. Rev. E **64**, 025102 (2001)
42. Newman, M.E.J.: From the cover: The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. USA **98**, 404–409 (2001)
43. Newman, M.E.J.: Detecting community structure in networks. Eur. Phys. J. B **38**, 321–330 (2004)
44. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Phys. Rev. E **69**, 066133 (2004)
45. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**, 026113 (2004)
46. Newman, M.E.J., Leicht, E.A.: Mixture models and exploratory analysis in networks. Proc. Natl. Acad. Sci. USA **104**, 9564–9569 (2007)
47. Newman, M.E.J., Park, J.: Why social networks are different from other types of networks. Phys. Rev. E **68**, 036122 (2003)
48. Noh, J.D., Jeong, H.C., Ahn, Y.Y., Jeong, H.: Growing network model for community with group structure. Phys. Rev. E **71**, 036131 (2005)
49. Onnela, J.P., Chakraborti, A., Kaski, K., Kertész, J., Kanto, A.: Dynamics of market correlations: Taxonomy and portfolio analysis. Phys. Rev. E **68**, 056110 (2003)
50. Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.L.: Structure and tie strengths in mobile communication networks. Proc. Natl. Acad. Sci. USA **104**, 7332–7336 (2007)
51. Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., de Menezes, M.A., Kaski, K., Barabási, A.L., Kertész, J.: Analysis of a large-scale weighted network of one-to-one human communication. New J. Phys. **9**, 179 (2007)
52. Palla, G., Barabási, A.L., Vicsek, T.: Quantifying social group evolution. Nature **446**, 664–667 (2007)
53. 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)
54. Palla, G., Farkas, I.J., Pollner, P., Derényi, I., Vicsek, T.: Directed network modules. New J. Phys. **9**, 186 (2007)
55. Palla, G., Vicsek, T., Barabási, A.L.: Community dynamics in social networks. Fluctuation and Noise Letters **7**, L273–L287 (2007)
56. Pollner, P., Palla, G., Vicsek, T.: Preferential attachment of communities: The same principle, but a higher level. Europhys. Lett. **73**, 478–484 (2006)
57. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., parisi, D.: Defining and identifying communities in networks. Proc. Natl. Acad. Sci. USA **101**, 2658–2663 (2004)
58. Ramasco, J.J., Morris, S.A.: Social inertia in collaboration networks. Phys. Rev. E **73**, 016122 (2006)
59. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L.: Hierarchical organization of modularity in metabolic networks. Science **297**, 1551–1555 (2002)
60. Reichardt, J., Bornholdt, S.: Detecting fuzzy community structures in complex networks with a Potts Model. Phys. Rev. Lett. **93**, 218701 (2004)
61. Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. Phys. Rev. E **74**, 016110 (2006)
62. Rives, A.W., Galitski, T.: Modular organization of cellular networks. Proc. Natl. Acad. Sci. USA **100**, 1128–1133 (2003)
63. Scott, J.: Social Network Analysis: A Handbook, 2nd edn. Sage Publications, London (2000)

64. Shiffrin, R.M., Börner, K.: Mapping knowledge domains. Proc. Natl. Acad. Sci. USA **101**, 5183–5185 (2004)
65. Spirin, V., Mirny, K.A.: Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci. USA **100**, 12123–12128 (2003)
66. Szabó, G., Fáth, G.: Evolutionary games on graphs. Phys. Rep.-Rev. Section Phys. Lett. **446**, 97–216 (2007)
67. Szabó, G., Vukov, J., Szolnoki, A.: Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices. Phys. Rev. E **72**, 047107 (2005)
68. Vicsek, T.: Phase transitions and overlapping modules in complex networks. Physica A-Statist. Mech. Appl. **378**, 20–32 (2007)
69. Vukov, J., Szabó, G., Szolnoki, A.: Cooperation in the noisy case: Prisoner's dilemma game on two types of regular random graphs. Phys. Rev. E **73**, 067103 (2006)
70. Wagner, C.S., Leydesdorff, L.: Network structure, self-organization, and the growth of international collaboration in science. Res. Policy **34**, 1608–1618 (2005)
71. Warner, S.: E-prints and the open archives initiative. Library Hi Tech **21**, 151–158 (2003)
72. Wasserman, S., Faust, K.: Social network analysis: methods and applications structural analysis in the social sciences. Cambridge University Press, Cambridge (1994)
73. Watts, D.J.: A twenty-first century science. Nature **445**, 489 (2007)
74. Watts, D.J., Dodds, P.S., Newman, M.E.J.: Identity and search in social networks. Science **296**, 1302–1305 (2002)
75. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
76. White, H.C., Boorman, S.A., Breiger, R.R.: Social structure from multiple networks. I. Blockmodels of roles and positions. Am. J. Sociol. **81**, 730–780 (1976)
77. Wilkinson, D.M., Huberman, B.A.: A method for finding communities of related genes. Proc. Natl. Acad. Sci. USA **101**, 5241–5248 (2004)
78. Yeung, Y.Y., Liu, T.C.Y., Ng, P.H.: A social network analysis of research collaboration in physics education. Am. J. Phys. **73**, 145–150 (2005)
79. Zhang, S., Wang, R.S., Zhang, X.S.: Uncovering fuzzy community structure in complex networks. Phys. Rev. E **76**, 046103 (2007)