Group formation and social evolution: a computational model

Nicholas Geard and Seth Bullock

SENSe, School of Electronics and Computer Science, University of Southampton, SO17 1BJ
nlg@ecs.soton.ac.uk

Abstract

The tendency to organise into groups is a fundamental property of human nature. Despite this, many models of social network evolution consider the emergence of community structure as a side effect of other processes, rather than as a mechanism driving social evolution. We present a model of social network evolution in which the group formation process forms the basis of the rewiring mechanism. Exploring the behaviour of our model, we find that rewiring on the basis of group membership reorganises the network structure in a way that, while initially facilitating the growth of groups, ultimately inhibits it.

Introduction

Groups, it has been argued, are a "basic process of social interaction" (Turner et al., 1987). Individuals rely on groups to achieve ends they could not achieve alone, and as a means for defining their personal identity. Groups, meanwhile, exist only so long as individuals are interested in becoming members of them. Much attention has recently been devoted to the task of identifying and understanding groups and communities in social networks.[1]

In sociology, the significance of groups as an expression of human social interaction, and their importance as an object of study, have a long history (Turner et al., 1987; Wasserman and Faust, 1994). The application of analytic tools from physics revitalised the study of social systems from a network perspective (Newman et al., 2002), and groups were recognised as a significant structural phenomenon, though one not amenable to easy characterisation (Jin et al., 2001; Davidsen et al., 2002; Girvan and Newman, 2002).

From a network perspective, a community is a subset of individuals who have more connections to other individuals within their community than to individuals from outside their community. A significant proportion of the literature on community structure focuses on the challenge of identifying the presence of communities in large data sets, such as those obtained from email records, automatic recommendation systems and social networking sites (Fortunato and Castellano, 2008, provide a recent overview of developments in this area).

A smaller fraction of the literature is concerned with the question of how communities arise, and understanding the social dynamics that influence their formation and evolution (Jin et al., 2001; Skyrms and Pemantle, 2000; Grönlund and Holme, 2004; Backstrom et al., 2006). A common feature of these models is that community structure frequently emerges as a side effect of another process, such as introductions between friends, or a desire to differentiate oneself from a population average.

In many real world contexts however, groups do not appear passively. Rather, they are the outcome of an active recruitment process, which arises in response to some perceived need that can best be met by a combined effort (Olson, 1971). For example, companies organise lobby groups in order to more effectively have their concerns heard by government, workers form unions to increase their bargaining power in negotiations with employers, and social movements arise to engage in collective action for a variety of humanitarian, environmental and other causes. We focus here on individuals and their participation in social movements (McAdam and Paulsen, 1993; Della Porta and Diani, 2005; Hedström, 2006).

Social movements are groups of people who come together to act collectively in support or opposition of some political or social issue (Tilly, 1978; Della Porta and Diani, 2005). It is widely accepted that social ties between individuals are critical to the success of social movements in recruiting new members (Snow et al., 1980; Marwell et al., 1988; McAdam and Paulsen, 1993). While some choices of group affiliation are undoubtedly a product of an individual's intrinsic preferences, the affiliations of their social contacts also exert an influence (Della Porta and Diani, 2005). Prop-

[1]The terms "group" and "community" are often used interchangeably in the literature; in the remainder of this paper, we will use "group" to refer to a subset of individuals in a population who each identify as belonging to a particular organisation, and "community" to refer to the a subset of individuals in the social network who are more densely linked to each other than to the remainder of the network.

erties of the social network, such as the number and intensity of ties between individuals, the existence of central nodes, and resource heterogeneity are therefore important determinants of how effectively a social movement can grow, and hence its ability to achieve its aims (Marwell et al., 1988; Gould, 1993; Kim and Bearman, 1997).

At the same time, an individual's participation in activities associated with a particular social movement is likely to strongly influence the people they meet, and hence on the set of individuals with whom they may form social ties (Della Porta and Diani, 2005). Thus, there is a bidirectional relationship between the short term dynamic of group formation occurring on a social network, and the longer term dynamic of the evolution of the structure of that social network (Sayama, 2007; Gross and Blasius, 2008).

Existing studies of community structure in networks have typically focused on exploring how communities can emerge from individual level rules. The reciprocal influence that group formation dynamics may have on social network evolution has been hitherto neglected. We are not aware of any model that explicitly considers group formation as a process that may actively influence the evolution of social networks. However, several recent models of opinion formation and cooperation in networked systems do confront a similar issue with regard to the coevolution of network's structure and the dynamic processes occurring on that network (Guimerà et al., 2005; Holme and Grönlund, 2006; Santos et al., 2006; Kozma and Barrat, 2008).

Explicitly considering the relationship between group formation and social evolution raises two interesting questions: how does social network structure influence the effectiveness of group formation, and how does group formation influence the evolution of the social network? In this paper, we propose a simple model of group formation and social network evolution and investigate the extent to which a group formation process can bring about (or hamper) the emergence of structural conditions contributing to its success; that is, the speed and size with which a group can recruit members.

Model Description

We model a social network as a simple graph containing N vertices representing individuals, and M undirected edges representing social ties (*i.e.,* each vertex has $K = 2M/N$ neighbours on average). Each vertex is associated with a trait vector $\boldsymbol{a}$. Each component of this vector is a continuous variable in the range $[0, 1]$ reflecting some aspect of an individual's social character (Watts et al., 2002; Boguñá et al., 2004); for example, their tendency to adopt a liberal or conservative stance on a particular social or political issue. Two individuals with similar values in a particular component of their trait vector will tend to share similar opinions on a particular issue. Viewed together, the totality of an individual's views describes a vector in an abstract social space. The social distance between two individuals x and y may then be calculated either in terms of the Euclidean distance between the vectors $\boldsymbol{a}_x$ and $\boldsymbol{a}_y$, or, with respect to issue n, the absolute difference between a_{xn} and a_{yn}, where a_{xn} is the nth component of $\boldsymbol{a}_x$.

Our model is updated on two distinct time scales: a short time scale corresponding to group formation, and a longer time scale corresponding to social evolution, in which each step represents a complete iteration of the group formation process.

Group formation phase: The group formation process follows the following sequence of steps:

1. G individuals are picked uniformly at random to seed G different groups. These individuals are added to a set of *active* individuals, A.
2. A single individual i_x is randomly chosen from A. This individual issues invitations to all of their network neighbours who are not currently affiliated with any group to join their group. The individual i_x is then removed from A.
3. Each individual i_y who receives an invitation accepts it with a probability equal to $\alpha(1 - |a_{xn} - a_{yn}|)$, where α is a model parameter governing the base probability of acceptance, and n is the index of the group to which i_x belongs. Therefore, if the nth component of trait vectors associated with i_x and i_y are identical, the distance between them will be zero, and the probability of acceptance will be α. As the difference between traits increases, the probability of acceptance decreases linearly.
4. Individuals who accept invitations are added to A.
5. Steps 2–4 are repeated until there are no individuals remaining in A.

At this point, the network can be in one of two states: either all individuals are members of a group, or the group formation process has died out before spreading through the entire network because all individuals on the periphery of a group have had their initiations to join refused. The probability of this occurring will depend on the value of α and structural features of the network, such as the density of edges (Figure 1). In order to ensure some variability that can be ascribed to network structure, we typically chose values of K and α that placed the initial network in the boundary region of Figure 1, where the group formation process was able to spread some distance beyond the seed individual, but did not percolate across the entire network.

Social update phase: After group formation has concluded, individuals who have joined a group adjust their social ties. We assume that being a member of a group

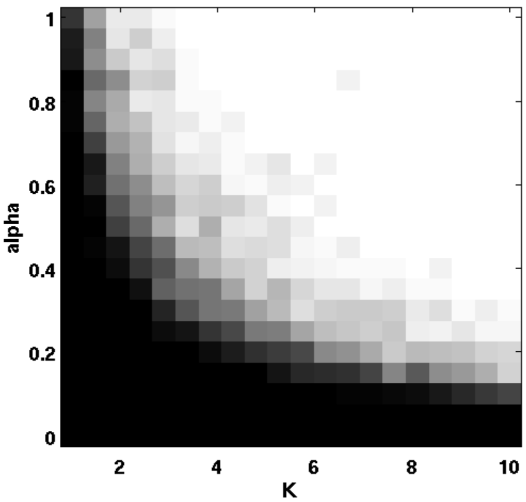

Figure 1: Proportion of network ($N = 2,000; G = 1$) becoming members of a group (black = 0%; white = 100%) for a range of values of α and K. Each parameter combination was repeated 200 times with randomly chosen seeds and the final group sizes were averaged. For low values of K and/or α, groups rarely grow beyond a few members. When both K and α are high, all individuals in the network join the group.

entails involvement in group-related activities that will result in an individual spending more time with members of their group (irrespective of whether they were previously known to them) and hence, given finite time, less time with current acquaintances who are not members of their group. We make the further assumption that individuals involved in groups are likely to update thier social neighbourhoods more frequently than unaffiliated individuals.

Each individual who is a member of a group therefore drops the edge connecting them to their least similar neighbour (irrespective of whether that neighbour is in their group or not) and creates a new edge connecting them to the member of their group, who is not currently a neighbour, to whom they are most similar. Similarity is again measured as the difference between the nth component of the respective trait vectors (*i.e.,* that corresponding to the group of which they are members).

After the social update phase has occurred, all groups are cleared, and the next iteration of the group formation phase begins on the new social network, with a new set of randomly chosen group seeds. Social movements often form in response to a particular issue, and either break apart or evolve into a new form as that issue becomes less relevant (Della Porta and Diani, 2005; Fuchs, 2006). Our decision to break apart all groups between each iteration of group formation is clearly a coarse approximation of this situation, but was chosen for initial simplicity.

Model Behaviour

To begin, we consider an initially random network with $N = 2,000$ and $K = 6$ in which only one group is formed during each iteration ($G = 1$) and trait values are drawn at random from a uniform distribution. In the simulations described here, these edges are initially randomly distributed between vertices following the Eŕdos-Rényi random graph model; however, other initial configurations are possible. This section describes the behaviour in an individual simulation run in detail, before exploring the sensitivity of this behaviour to K, α and the initial network structure.

The behaviour we are interested in observing is how the size of groups formed changes as the social network evolves. The size of the group formed depends not only on the global structure of the network, but also on the local neighbourhood of the seed individual. To obtain an indication of the general propensity of a particular network structure to facilitate group formation, we measure the average size of the group formed across fifty random seedings, with groups being erased after each. The social update phase is then carried out based on the group formed from the final of these seedings.

The structure of the social network passes through three distinct periods of evolution (Figures 2 and 3). Initially, the network is well connected but disordered (Figure 2, top panel), and the mean trait difference between neighbours is high (0.329). As a consequence, invitations have a low probability of being accepted and the resulting groups are small (12.42 members—0.62% of the population—on average over the first 10 iterations of the simulation). Mean clustering coefficient and path length both remain low (approximately 0.04 and 7.2 respectively), as is typical of a random graph.

However, the groups that do form enable their local regions of the network to become more ordered, by allowing individuals with similar trait vectors to increase the density of their interconnections, (Figure 2, middle panel). By doing so, they increase the probability of future invitations between individuals in this region being accepted and so assist the formation of groups in subsequent iterations.

Surprisingly, rather than produce a steady increase in the average size of groups as the social network becomes more ordered, a phase transition occurs at the point where a large proportion of the network simultaneously becomes well organised (Figure 2). Mean group size increases dramatically, peaking at 438.12 members (21.9% of the population) in iteration 75 (Figure 3). Mean clustering coefficient increases by an order of magnitude to approximately 0.5 by iteration 80, while mean path length remains relatively low and the degree distribution becomes more skewed, properties indicative of small world structure.

A side effect of larger groups forming is that the rate of network reorganisation increases (as each individual who is in a group updates one of their social ties). Furthermore,

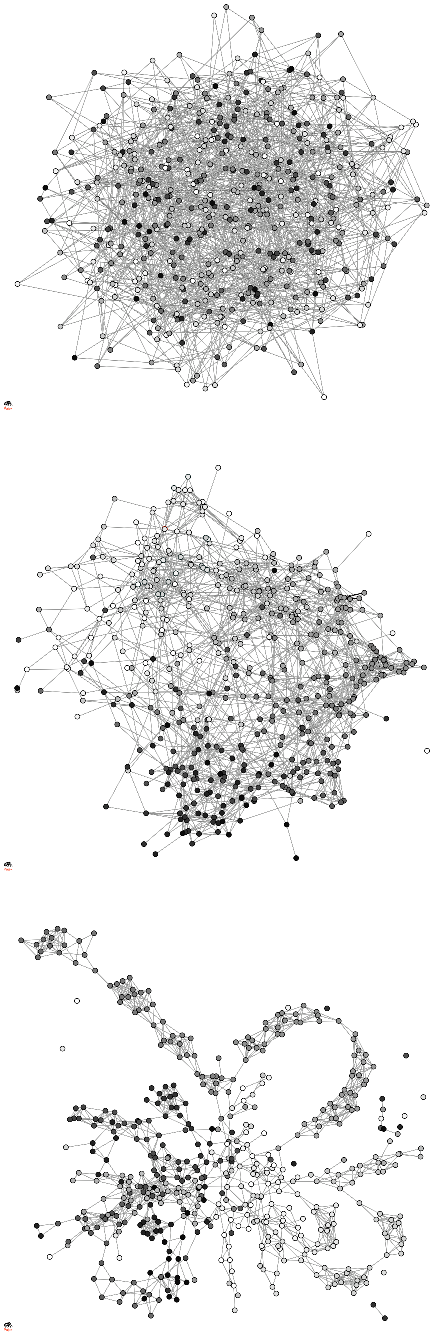

Figure 2: Network structure observed at different points in evolution: (a) the initial random network, with trait values dispersed throughout; (b) the network at the point of phase transition, when group formation spreads rapidly between neighbours with similar trait values; and (c) the network at the end of a run, with individuals clustered into weakly connected communities. Note that smaller networks ($N = 500$) are shown for clarity; however, their qualitative features are otherwise similar.

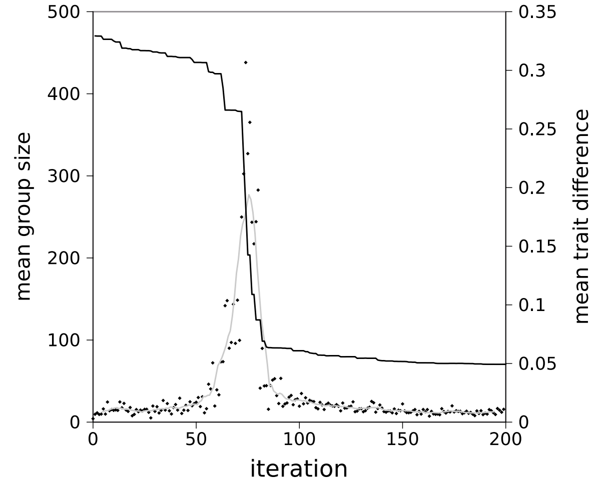

Figure 3: A representative run of the model initialised with a random network ($N = 2000, K = 6, \alpha = 0.25$). Each symbol represents the mean group size observed over 20 random seedings (as described in the text), together with a moving average calculated over 10 iterations (gray line) and mean trait difference between neighbouring nodes (black line). The three networks in Figure 2 correspond to networks observed prior to, during, and after, the spike in mean group size.

each individual is now able to select their new neighbour from a wider pool of potential candidates (their fellow group members). The mean trait difference between neighbours drops (to 0.049, Figure 3) and the network begins to partition into a number of weakly connected communities (Figure 2, bottom panel, and Figure 3). Around iteration 90, mean path length begins to increase steadily, reaching approximately 14 by iteration 200.

In the extreme case, the network may disintegrate completely into a set of disconnected components; however, this is not required in order for group size to fall: by iteration 200, 94.4% of individuals still belonged to a single connected component. The appearance of community structure sufficient to hamper the formation of groups by creating bottlenecks that impede the spread of invitations. If there is only a single link between two communities, then, even if it is between two very similar individuals, group membership has a chance of spreading at best equal to α.

This social evolution dynamic was observed across a range of parameter settings, with the primary differences being the time required for the network to organise, and the maximum size to which groups are able to grow (Figures 4 and 5). As K and/or α increase, the size of groups that form throughout each simulation run also increases, in line with the trend illustrated in Figure 1. For all combinations of K and α, the peak group size achieved is substantially greater

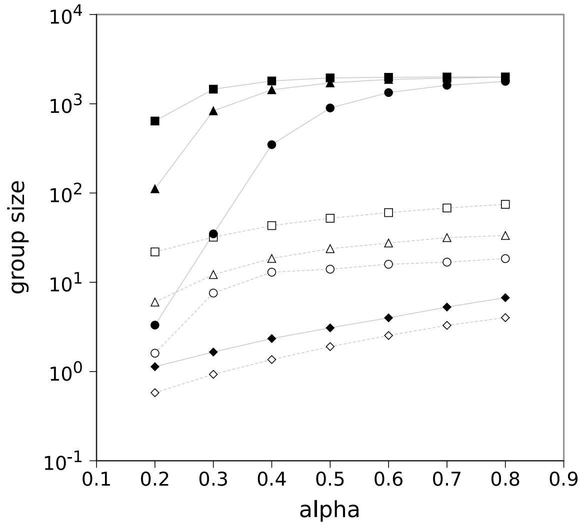

Figure 4: Peak (filled) and average (hollow) group sizes for various values of α and K (diamond: 2; circle: 4; triangle: 6; square: 8). Note that the Y-axis is log-scaled. Other model parameters: $N = 2,000; G = 1$. Each data point is averaged over 20 runs. The peak group size is that obtained during the phase transition in network structure. Average group size is that resulting after the phase transition has occurred.

than the average. Furthermore, increasing K and α results in the peak group size being obtained earlier in the simulation run.

We investigated the effect of the initial network configuration (Figure 6) on social evolution by varying the rewiring probability p used to create the initial network. In comparison to random graphs ($p = 1.0$), regular lattices ($p = 0.0$) with comparable N and M take considerably longer to organise and, at their peak, result in smaller groups. In many simulation runs (such as that shown in Figure 6), no peak phase occurs, and the network transitions directly to the disjoint community phase. Small world networks ($p = 0.05$) organise more slowly than random graphs, but otherwise behave similarly, and the occurrence of a peak phase is more reliable than in lattices.

We have also carried out preliminary investigations into the behaviour of the model when there is more than one group forming in each iteration ($G > 1$). In this case, competition between groups appears to lead to a "rich get richer" process, whereby large groups tend to increase in size, at the expense of smaller groups. The mechanism responsible for this is straightforward: once one group begins to increase in size with respect to the others, its members dominate the set of active individuals (A), and hence benefit from more frequent opportunities to recruit unaffiliated individuals.

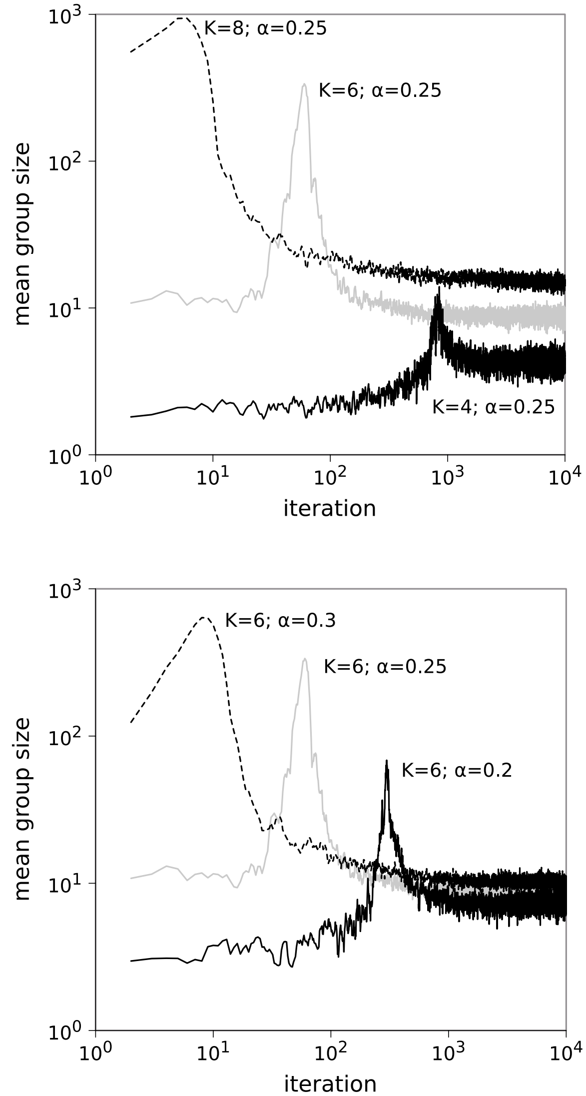

Figure 5: Mean group size trends for various values of K (top) and α (bottom). For clarity, a moving average over 5 iterations is shown, rather than individual data points. Note that both axes are log-scaled. Increasing either the density of connections (K) or the base probability of invitations being accepted (α) increases both the speed with which the network organises, and the peak group size that can be achieved.

Discussion

How can we interpret the pattern of social evolution observed in our model? As communities begin to emerge, the ability of groups to recruit large numbers of people initially improves. However, as these communities become stronger, they also become more homogeneous and detached from the wider social context in which they exist. This social isola-

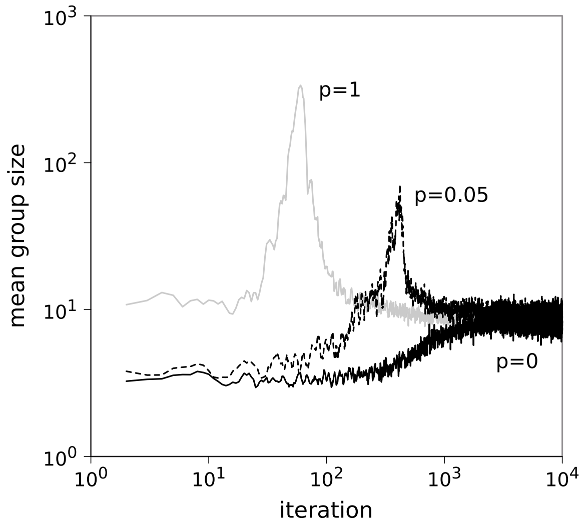

Figure 6: Mean group size trends for various initial network configurations (rewiring probabilities, $p = 0, 0.05$ and 1). For clarity, a moving average over 5 iterations is shown, rather than individual data points. Note that both axes are log-scaled. See text for discussion of trends.

tion severely limits the ability of groups to recruit new members and hence, potentially, to achieve their aims in an effective fashion (Snow et al., 1980). Networks with a higher density of social ties are more rapidly reorganised to facilitate, and later inhibit, group formation. Similarly, populations in which people have a strong predisposition toward joining groups reorganise more rapidly.

There is general evidence that segregation of social networks can arise despite the absence of any explicit preference for such an outcome (Schelling, 1971). Even when interaction structures are externally imposed, such as the hierarchical reporting relationships of a large organisation, there is evidence to suggest that the existence of communities can have a negative effect on global integration (Kilduff and Tsai, 2003). In the context of organisations, such an effect has led to the value placed upon *bridges*—individuals who fill structural holes in a network by linking otherwise disconnected components. Individuals in such positions often gain social capital from the role they play in mediating between different interest groups (Burt, 2002).

Social movements, too, can benefit from being linked together. Della Porta and Diani (2005) summarise extensive evidence suggesting that linkages between social movements allow sharing of information and resources, and facilitate cooperation and coordination of the aims of different movements. A key factor in linking movements is overlapping memberships—the existence of individuals who are members of two or more groups (Palla et al., 2005). This suggests that our assumption of exclusive group membership will require reevaluation. One promising direction for future work is to allow individuals to belong to multiple groups at the same time, and to explore the extent to which this enables the social network to organise in such a way that it facilitates the formation of groups without disintegrating into weakly connected components.

In summary, this paper has presented a novel model of group formation and social evolution that takes as its starting point two main ideas: first, that group formation is a process in which individuals actively seek to engage, and second, that this tendency has repercussions for the evolution of social network structure. The investigations reported here indicate that rewiring on the basis of group membership reorganises the network structure in a way that, while it initially benefits the growth of groups, ultimately inhibits it.

It is worth noting that, in order to remain as simple as possible, the model described here makes several assumptions that may limit its general applicability. For example, an individual's decision to join a particular group is based purely upon their similarity with the individual who has invited them, and does not take into account factors such as the alignment of their values with those of the group, or the opinions of their social neighbours (McAdam and Paulsen, 1993). Group membership is exclusive; that is, it is not possible for an individual to simultaneously be a member of more than one group which, as discussed above, is likely to be play a role in ensuring social cohesion (Della Porta and Diani, 2005; Palla et al., 2005). Despite these limitations of the model in its current formulation, we believe it to be a fruitful starting point for further exploration into the the co-evolution of topology and dynamics in social networks.

References

Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54.

Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., and Arenas, A. (2004). Models of social networks based on social distance attachment. *Phys. Rev. E*, 70(5):056122.

Burt, R. S. (2002). The social capital of structural holes. In Guillén, M. F., editor, *The New Economic Sociology: Developments in an Emerging Field*, pages 148–190. Russell Sage Foundation, New York, NY.

Davidsen, J., Ebel, H., and Bornholdt (2002). Emergence of a small world from local interaction: Modeling acquaintance networks. *Phys. Rev. Lett.*, 88:128701.

Della Porta, D. and Diani, M. (2005). *Social Movements: An Introduction*. Wiley Blackwell, Oxford, 2nd edition.

Fortunato, S. and Castellano, C. (2008). Community structure in graphs. To appear in Springer's *Encyclopedia of Complexity and System Science*. arXiv:0712.2716v1.

Fuchs, C. (2006). The self-organization of social movements. *Systemic Practice and Action Research*, 19(1):101–137.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Nat. Acad. Sci. USA*, 99:7821–7826.

Gould, R. V. (1993). Collective action and network structure. *American Sociological Review*, 58(2):182–196.

Grönlund, A. and Holme, P. (2004). Networking the seceder model: Group formation in social and economic systems. *Phys. Rev. E*, 70(3):036108.

Gross, T. and Blasius, B. (2008). Adaptive coevolutionary networks: a review. *Journal of the Royal Society Interface*, 5:259–271.

Guimerà, R., Uzzi, B., Spiro, J., and Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308:697–703.

Hedström, P. (2006). Explaining the growth patterns of social movements. In Elster, J., Gjelsvik, O., Hylland, A., and Moene, K. O., editors, *Understanding Choice, Explaining Behaviour*. Oslo Academic Press, Oslo.

Holme, P. and Grönlund, A. (2006). Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E*, 74:056108.

Jin, E. M., Girvan, M., and Newman, M. E. J. (2001). The structure of growing social networks. *Phys. Rev. E*, 64:046132.

Kilduff, M. and Tsai, W. (2003). *Social Networks and Organizations*. Sage Publications, London.

Kim, H. and Bearman, P. S. (1997). The structure and dynamics of movement participation. *American Sociological Review*, 62(1):70–93.

Kozma, B. and Barrat, A. (2008). Consensus formation on coevolving networks: groups' formation and structure. Submitted to *J. Phys. A: Math. Gen.* arXiv:0801.4860v1.

Marwell, G., Oliver, P. E., and Prahl, R. (1988). Social networks and collective action: A theory of the critical mass III. *The American Journal of Sociology*, 94(3):502–534.

McAdam, D. and Paulsen, R. (1993). Specifying the relationship between social ties and activism. *The American Journal of Sociology*, 99(3):640–667.

Newman, M. E. J., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proc. Nat. Acad. Sci. USA*, 99:2566–2572.

Olson, M. (1971). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA, 2nd edition.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818.

Santos, F. C., Pacheco, J. M., and Lenaerts, T. (2006). Cooperation prevails when individuals adjust their social ties. *PLoS Comput. Biol.*, 2(10):e140.

Sayama, H. (2007). Generative network automata: A generalized framework for modeling complex dynamical systems with autonomously varying topologies. In *Proceedings of the First IEEE Symposium on Artificial Life*, pages 214–221. IEEE Press.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143–186.

Skyrms, B. and Pemantle, R. (2000). A dynamic model of social network formation. 97(16):9340–9346.

Snow, D. A., Zurcher, L. A., and Ekeland-Olson, S. (1980). Social networks and social movements: A microstructural approach to differential recruitment. *American Sociological Review*, 45:787–801.

Tilly, C. (1978). *From Mobilization to Revolution*. Addison-Wesley, Reading, MA.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S., and Wetherell, M. S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Blackwell Publishing, Oxford.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis*. Cambridge University Press, Cambridge.

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