Available online at www.sciencedirect.com
ScienceDirect
Physica D 224 (2006) 20–26
PHYSICA D
www.elsevier.com/locate/physd

When are networks truly modular?

Jörg Reichardt*, Stefan Bornholdt

Institute for Theoretical Physics, University of Bremen, Otto-Hahn-Allee, 28359 Bremen, Germany

Available online 30 October 2006

Abstract

The study of cluster or community structure of complex networks contributes to the understanding of networks at a functional level. In many networks, latent classes of nodes are suspected which manifest themselves as communities, i.e. groups of nodes with a high link density among the nodes of the same class and low link density between nodes of different classes. Community detection algorithms are used to infer these classes, e.g. by finding a partition of the network which maximizes a quality function such as the network modularity Q [M. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113]. However, it is known from numerical experiments that even purely random networks display intrinsic modularity and may be partitioned yielding high values of Q. Extending on our earlier results [J. Reichardt, S. Bornholdt, Statistical mechanics of community detection, Phys. Rev. E 74 (2006) 016110], the mapping of the community detection problem onto finding the ground state of a spin glass is exploited in order to derive analytical expressions for the expected modularity in random graphs and assess the theoretical limits to community detection. The results are independent of any specific community detection algorithm and allow for differentiation between modularity arising purely due to the search process in the large configuration space of possible partitionings on the one hand, or due to the actual presence of different classes of nodes on the other hand.

Keywords: Graph clustering; Community detection; Spin models

1. Introduction

With the increasing availability and steadily increasing size of relational datasets or networks the need for appropriate methods for exploratory data analysis arises. For general statistical properties such as the degree distribution, degree correlations, clustering etc. a number of well established methods and models to explain their origin exist [1,2]. However, a standard analysis for the higher order structure in graphs has not been established so far. Currently, the problem of the cluster or community structure is the subject of intense study [3,4]. Cluster analysis is an important technique that allows for data abstraction and dimensionality reduction or aids in data visualization. It is used in the life sciences [5], over bibliometrics [6], to market research [7], and has implications for experiment planning, funding policies or marketing.

The above examples have also illustrated that such exploratory analysis is often the starting point to further work. It is therefore important to assess the statistical validity of the findings and avoid the "deception of randomness" [8], i.e. to ensure that the output of a community detection algorithm is statistically significant and not the mere result of the search process. To illustrate this, we consider the following problem. Given is an Erdös–Rényi (ER) network with average degree $\langle k\rangle = 5$. Given is further an assignment of the nodes into two types A and B with 50 nodes each. Between nodes of different type, 42 links are found, the remaining links being equally distributed among nodes of type A or B alone. If nodes are connected independently of their type, the total number of links between type A and B nodes is Poisson-distributed with a mean of $\langle k\rangle N/4 = 125$ and a standard deviation of $\sigma = 11$. Hence, finding only 42 links between A and B is statistically highly significant with a p-value of $p = 2.8\times 10^{-18}$. Now assume that the type of each node was not given and an assignment into two equal sized groups A and B was found through an exhaustive search of the $\binom{N}{N/2}$ possible assignments into two equal sized

* Corresponding address: Institute for Theoretical Physics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany. Tel.: +49 931 888 5733; fax: +49 931 888 5141.
E-mail addresses: reichardt@physik.uni-wuerzburg.de (J. Reichardt), bornholdt@itp.uni-bremen.de (S. Bornholdt).

0167-2789/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2006.09.009

groups. Applying a Bonferroni correction [9] for this number of different "experiments" would lead to the situation that only less than 22 connections between nodes of type A and B would be significant at the 5% level. Hence, the initial situation with 42 links between nodes of different type could not be called significant. As will be shown below, any ER random network with 100 nodes and $\langle k \rangle = 5$ can be partitioned into two equal sized groups such that only 42 links connect the two parts. Thus, statistical significance starts much earlier than the limit given by the Bonferroni correction. The Bonferroni correction fails here because it assumes independent experiments. The different assignments into groups produced by a search process, however, are not independent.

In addition to the problem of statistical significance, another aspect of community detection is still under intense discussion, namely the definition of the term community or cluster in a network. This article aims at contributing to both of these questions.

2. What is a community?

Despite the many applications of community detection across the sciences, it remains remarkably unclear what a community actually is. In addition to the many definitions that are given in sociology [10], the physics community has contributed a fair number as well [3,4]. All authors agree that a community should be a group of nodes that is more densely connected among each other than with the rest of the network, but still these definitions differ largely in the details. Below, we give a short overview of the different aspects that have been emphasized by different authors.

The initial work on communities by Girvan and Newman [11] gives an algorithmic definition. They design a community detection algorithm which recursively partitions the graph to produce a hierarchy of communities from the entire network down to single nodes. At each point, the nodes belonging to distinct sub-trees in the resulting dendrogram are considered as communities.

Radicchi et al. [12] tried to improve this heuristic definition by coining the term of "community in a strong sense" such that

$$k_i^{\text{in}} > k_i^{\text{out}}, \quad \forall i \in C. \tag{1}$$

This means for all nodes i in the community C, the number of connections node i has to members of its own community k_i^{in} is larger than k_i^{out}, the number of connections it has to the rest of the network. Further, they define a "community in a weak sense", such that the sum of internal connections is larger than the sum of external links $\sum_{i\in C} k_i^{\text{in}} > \sum_{i\in C} k_i^{\text{out}}$. Radicchi et al. then suggest to stop any recursive partitioning algorithm when an additional partition would not result in a community in the strong (or weak) sense.

Palla et al. [6,13] have given a definition based on reachability. They define a subgraph percolation process based on k-cliques (fully connected subgraphs with k nodes). Two k-cliques are connected, if they share a $(k-1)$-clique, e.g. two triangles (which are 3-cliques) are connected if they share an edge (a 2-clique). A community, or k-clique percolation cluster, is then defined as the group of nodes that can be reached via adjacent k-cliques. Communities may overlap, i.e. nodes may belong to more than one percolation cluster, but communities corresponding to $(k+1)$-clique percolation clusters always lie completely within k-clique clusters.

Newman and Girvan have further defined a quantitative measure of the quality of an assignment of nodes into communities. This so-called "modularity" [14] can be used to compare different assignments of nodes into communities quantitatively. The modularity is defined as:

$$Q = \sum_s (e_{ss} - a_s^2). \tag{2}$$

The sum runs over all communities s. The fraction of all links connecting nodes in group s and r is denoted by e_{sr}. Hence, e_{ss} is the fraction of all links lying within group s. The fraction of all links connecting to nodes in group s is denoted by $a_s = \sum_r e_{rs}$. One can interpret a_s^2 as the expected fraction of internal links in group s, if the network was random and the nodes were distributed randomly into the different groups. Such a measure can be used to stop recursive partitioning or agglomerative approaches when they do not lead to an improvement of Q anymore [15].

We see the diversity of definitions and approaches of which we have described only a few. Refs. [3,4] give a more comprehensive overview. Because of this controversy of opinions, we have set out from a first principles approach in the next section that will shed some light on the general properties of the problem.

3. A first principles approach to community detection

As outlined in Ref. [16] the problem of community detection can be addressed from a first principles perspective by adhering to a simple principle: to group nodes that are not linked in different communities and to put nodes which are linked in the same community. This principle is expressed in the following Hamiltonian:

$$\mathcal{H}(\sigma) = -\sum_{i<j} a_{ij} A_{ij} \delta(\sigma_i, \sigma_j) + \sum_{i<j} b_{ij}(1 - A_{ij})\delta(\sigma_i, \sigma_j). \tag{3}$$

Here, σ_i denotes the group index of node i, $\delta(\sigma_i, \sigma_j)$ is the Kronecker delta, A_{ij} is the adjacency matrix of the network with $A_{ij} = 1$ if nodes i and j are connected and zero otherwise. Hence, the first sum runs over all pairs of connected nodes, while the second sum runs over all pairs of unconnected nodes. Our Hamiltonian rewards every pair of connected nodes (i, j) in the same group with a_{ij} and penalizes every pair of unconnected nodes (i, j) in the same community with b_{ij}. It implements just the principle we started out from. Any spin configuration that will minimize (3) is hence optimal in the sense of this first principle. It is now important to define the weights a_{ij} and b_{ij} in a sensible way. A particularly good choice is to balance them, such that all existing connections in the network are equally important to our optimality criterion as are all missing connections [16]. One way of achieving this is to

set $a_{ij} = 1 - \gamma p_{ij}$ and $b_{ij} = \gamma p_{ij}$ which also reduces the need for two different weights to only one. We have introduced an additional parameter γ that will allow us to adjust the balance of missing and existing links. The only constraint we have to impose on p_{ij} is that $\sum_{i<j} p_{ij} = M$ with M being the total number of links in the network. With this choice, Eq. (3) is written in the much simpler form:

$$\mathcal{H}(\sigma) = -\sum_{i<j} \left(A_{ij} - \gamma p_{ij}\right) \delta(\sigma_i, \sigma_j). \tag{4}$$

Eq. (4) is formally identical to the Hamiltonian for a q-state Potts spin glass, with q being the number of possible group indices. The coupling matrix is then defined as $J_{ij} = A_{ij} - \gamma p_{ij}$. We identify p_{ij} with the connection probability between nodes i and j in the network. Depending on the network under study, this can be $p_{ij} = p$, if the links are assumed to connect nodes with constant probability $p = 2M/N(N-1)$. Another possible choice is $p_{ij} = \frac{k_i k_j}{2M}$, if the degree distribution of the nodes is to be taken into account and there are no degree–degree correlations. Here k_i denotes the degree of node i and M represents the number of links in the network as before.

Both of these choices render p_{ij} positive and smaller than one, hence we are dealing with a spin glass which has ferromagnetic couplings between connected nodes and anti-ferromagnetic couplings between unconnected nodes. The ground state of this spin glass defines the optimal assignment of nodes into communities. For $\gamma = 1$ and $p_{ij} = k_i k_j / 2M$, we recover the modularity Q defined by Newman and Girvan [14] from (4) via $Q = -\frac{1}{M}\mathcal{H}$ [17].

It is instructive to rewrite (4) as a sum over spin states s:

$$\mathcal{H} = -\sum_{s} \underbrace{\left(m_{ss} - \gamma [m_{ss}]_{p_{ij}}\right)}_{c_{ss}} = \sum_{s<r} \underbrace{\left(m_{rs} - \gamma [m_{rs}]_{p_{ij}}\right)}_{a_{rs}}. \tag{5}$$

We denote the number of links within group s by m_{ss} and between groups r and s by m_{rs}. Further, we denote the expectation values of these quantities under the model of connection probability p_{ij} and assuming a random assignment of spins into groups by $[\cdot]_{p_{ij}}$. In (5) we have introduced two new terms c_{ss} and a_{rs} which measure within group "cohesion" and between group "adhesion", respectively. Maximizing cohesion and minimizing adhesion are in fact equivalent and will hence always be extremal at the same time, i.e. any configuration of spins that minimizes $\mathcal{H}$ will automatically maximize cohesion and minimize adhesion.

With our mapping couplings between all pairs of nodes exist. Fortunately, the particular choice of p_{ij} allows us to implement efficient optimization routines [16] that only need to consider interactions along the links and treat the anti-ferromagnetic interactions along the non-existing links in a mean field manner, which is, however, not an approximation but accounts exactly for the repulsive interactions. One only needs to keep track of the occupation numbers of the spin states or the total sum of degrees in each group.

A definition of community follows directly from the properties of the ground state as a global minimum of (4) [16]:

(1) Every proper subset n_1 of a community n_s has a maximum coefficient of adhesion with its complement in the community compared to the coefficient of adhesion with any other community ($a_{1,s\setminus 1} = \max$).
(2) The coefficient of cohesion is non-negative for all communities ($c_{ss} \geq 0$).
(3) The coefficient of adhesion between any two communities is non-positive ($a_{rs} \leq 0$).

As a community, we understand a group of nodes that has the above three properties. The presented formalism also allows for the detection of overlapping communities and community structures in possibly degenerate ground states. Sampling local minima of the energy landscape defined by the graph under study will also lead to valid community assignments. They can be regarded as sub-optimal community structures and the study of their overlap among each other and with the ground state yields valuable information about how many alternative, but sensible groupings exist for a particular network [18,16].

The ground state depends on the value γ chosen. Recall the definitions of adhesion and cohesion in Eq. (5), where γ measures the relative influence of the number of internal and external links over the respective expectation values. Further, compare the ground state properties (1) to (3). For $\gamma = 1$, a set of nodes has non-negative cohesion, if it has at least as many internal links as expected from a random assignment and for $\gamma = 2$, a set of nodes needs at least twice as many internal links than expected in order to be considered a community. Therefore, larger values of γ will lead to denser communities which are also smaller. The opposite is true for smaller values of γ. Hence, γ determines a threshold for the link density contrast in the community structure. The highest sensible value is $\gamma \approx 1/p$, since then only complete subgraphs (cliques) will have a non-negative cohesion. The lowest sensible value of γ will be the largest value which leads to a ferromagnetic ground state. Of course, an initial investigation should be performed at $\gamma = 1$, which corresponds to the natural partition of the graph. If one then wants to find additional structures, for most practical applications a sensible parameter range is $0.01 \leq \gamma \leq 100$, which should be explored starting around $\gamma = 1$.

The value of γ at which the community structure was obtained should always be quoted. Changing the value of γ allows us to detect hierarchies in the assignment of nodes into communities [18,16].

The performance of this approach to community detection was benchmarked on computer generated test networks [18, 4] and the results compared to those obtained by Girvan and Newman's betweenness algorithm [11]. The networks are Erdös–Rényi (ER) graphs [19] with an average degree of $\langle k \rangle = 16$ and 128 nodes. The nodes were divided into 4 groups of 32 nodes each. Keeping the average degree fixed, the links per node were distributed into an average of $\langle k_{\text{in}} \rangle$ to members of the same group and an average of $\langle k_{\text{out}} \rangle$ to members of the 3 remaining groups in the network such that $\langle k_{\text{out}} \rangle + \langle k_{\text{in}} \rangle = \langle k \rangle$. Obviously, increasing $\langle k_{\text{out}} \rangle$ at the expense of $\langle k_{\text{in}} \rangle$ makes the recovery of the designed community structure more difficult. At $\langle k_{\text{in}} \rangle = 4$ the network should be completely random and any trace of the built-in community structure is lost since at this

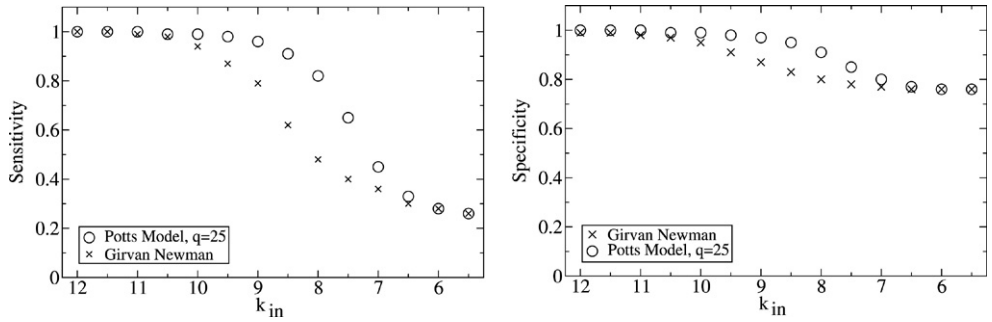

Fig. 1. Benchmarks of a community detection algorithm based on finding the ground state of a spin glass and comparison with Girvan–Newman's algorithm [11]. Tests were run on computer generated test networks with known community structure. "Sensitivity" denotes the fraction of all pairs of nodes that are classified correctly in the same community, while "specificity" denotes the fraction of all pairs of nodes classified correctly in different communities.

point the probability to link to a member of a different node equals the probability to link to a member of the same group $p_{\text{in}} = p_{\text{out}} = p$.

Fig. 1 shows the results of the benchmarks. We measured the success of the two methods in terms of sensitivity and specificity. Sensitivity measures the fraction of pairs of nodes that are correctly classified as being in the same cluster, while specificity measures the fraction of nodes correctly classified as belonging to different clusters. In other words, the two measures indicate how good the algorithms are in grouping together what belongs together and in keeping apart what does not belong together. From Fig. 1 we see that both algorithms are rather conservative in terms of grouping things together as indicated by the high levels of specificity. The change in sensitivity is much more drastic and we find that the Potts model approach outperforms the algorithm of Girvan and Newman [11]. The critical value of $\langle k_{\text{in}} \rangle$, at which the ability to recover the built-in community structure vanishes, seems to be $\langle k_{\text{in}} \rangle_c \approx 8$.

4. Communities and modularities in random networks

In our introductory paragraphs, we have already raised the question of when one may call a network truly modular. Obviously, running a clustering algorithm over a set of randomly generated data points will always produce clusters which, however, have little meaning. Similarly, minimizing the modularity Hamiltonian on a random graph results in a community structure which has all the desired properties. This does, of course, not mean that the graph we studied was in fact modular. A differentiation between graphs which are truly modular and those which are not can hence only be made if we gain an understanding of the intrinsic modularity of random graphs. By comparing the modularity of random graphs with that of real world graphs, we can assess whether the graphs under study are truly modular.

Such a comparison can of course always be made by randomizing the network under study keeping the degree distribution invariant. Such algorithms then remove all correlations and community structures possibly present in the data. Comparing the results of clustering the empirical data and a randomized version of it can always give a clue to what extent the data shows modularity above that of a random network with the same degree distribution. Nevertheless, such analysis is biased by the algorithm used to detect the community structure. Much more desirable would be a measure of modularity that can be used to compare with any algorithm.

In mapping the problem of finding a community structure onto finding the ground state of an infinite range spin glass, we have defined a coupling matrix J_{ij} with the following distribution of couplings:

$$q(J_{ij}) = p_{ij}\delta(J_{ij} - (1 - p_{ij})) + (1 - p_{ij})\delta(J_{ij} + p_{ij}), \quad (6)$$

where we have set $\gamma = 1$ and assumed we are dealing with a random network in which the links are distributed with the same p_{ij} we use for defining the weights a_{ij} and b_{ij} of the contributions of existing and missing links in the clustering. It is easy to see that this distribution has zero mean. Since the mean of this distribution couples only to the magnetization, we find a zero magnetization in the ground state [20]. This corresponds to an equi-partition of the network. The community structure of a random network consists of all equal sized communities. If we conceive community detection as looking for the "natural partition" of a network, then the natural partition of a random graph is the equi-partition.

For the number of edges to cut when equi-partitioning a random graph, a number of results exist since the 1980's, beginning with the paper by Fu and Anderson [20] about bi-partitioning a random graph. Kanter and Sompolinsky [21] have given an expression for the minimum total number of inter community edges $\mathcal{C}$, also called cut-size, when partitioning a random graph into q equal sized parts. From this, we can immediately write an expectation value for the modularity of random graphs [16]:

$$Q = -\frac{1}{M}\mathcal{H}_{GS} = \frac{N^{3/2}}{M}\sqrt{p(1-p)}\frac{U(q)}{q}. \quad (7)$$

For the $U(q)$, the ground state energy of a q-state Potts glass with Gaussian couplings of zero mean and variance $J^2 = 1$, some values for small q are given in Table 1 obtained by using the exact formula for calculating $U(q)$ from [21]. For large q, we can approximate $U(q) = \sqrt{q \ln q}$ [21].

We see that maximum modularity is obtained at $q = 5$, though the value of $U(q)/q$ for $q = 4$ is not much different from it. This qualitative behavior of dense random graphs tending to cluster into only a few large communities

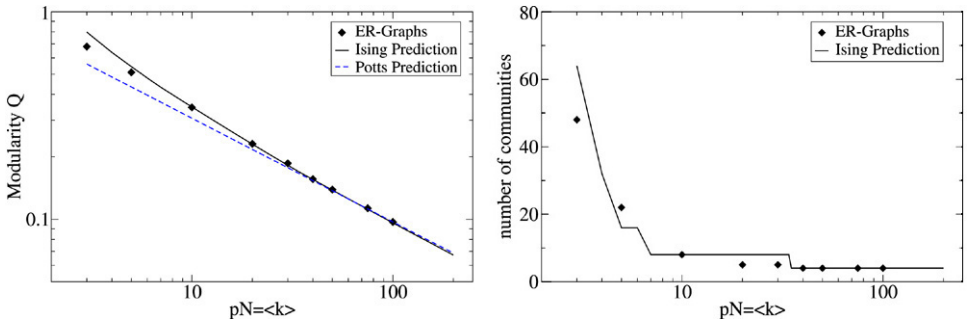

Fig. 2. Modularity and the number of communities in ER graphs. Shown are the values determined from clustering random graphs with $N = 10{,}000$ nodes and the expectation values calculated from using a Potts model (8) or an Ising model (10) recursively.

is confirmed by our numerical experiments. Using the largest value of Table 1, we finally arrive at an expression for the expected modularity in any ER random graph with average degree $\langle k\rangle = pN$:

$$Q = 0.97\sqrt{\frac{1-p}{pN}}. \tag{8}$$

Fig. 2 shows the comparison of Eq. (8) and experiments in which we have numerically maximized the modularity in random graphs with $N = 10{,}000$ nodes and varying connectivity $\langle k\rangle$ using a simulated annealing approach as described in Ref. [16].

The above approximation using a Potts spin glass, however, cannot explain the number of communities found experimentally in random graphs of varying connectivity since it always assumes 5 communities. Therefore, we try to approximate the ground state of a q-state Potts model by recursively bi-partitioning the network and continuing as long as the modularity increases. For every bi-partition we use the expression of the cut-size as a function of the number of N nodes and average degree $\langle k\rangle = pN$ given by Fu and Anderson [20]. After every partition, the number of links connecting to nodes in the same part and to nodes in the rest of the network is given by:

$$\langle k_{\text{in}}\rangle = \frac{pN + c\sqrt{pN(1-p)}}{2} \quad \text{and}$$
$$\langle k_{\text{out}}\rangle = \frac{pN - c\sqrt{pN(1-p)}}{2}. \tag{9}$$

The constant c corresponds to $U(2)$ and is given by $c = 1.5266 \pm 0.0002$ [20]. After b successive recursive partitions, we arrive at a modularity of

$$Q(b) = \frac{2^b - 1}{2^b} - \frac{1}{\langle k\rangle}\sum_{t=1}^{b}\langle k_{\text{out},t}\rangle \tag{10}$$

where $\langle k\rangle$ is the average degree in the total network and $\langle k_{\text{out},t}\rangle$ is the average number of external links a node gains after partition number t calculated from (9).

Though Eq. (10) only allows numbers of communities that are powers of 2, the agreement of Q with the experimental data is surprisingly good as Fig. 2 shows. Also, the number of communities is predicted almost perfectly by (10) as shown in Fig. 2.

Table 1
Values of $U(q)/q$ for various values of q obtained from [21], which can be used to approximate the expected modularity with Eq. (7)

q	2	3	4	5	6	7	8	9	10
$U(q)/q$	0.384	0.464	0.484	0.485	0.479	0.471	0.461	0.452	0.442

With the expressions (8) and (10), we are adequately able to calculate expectation values of Q for random graphs which can be used in the assessment of the statistical significance of the modularity in real world networks. Note that our analytical results improve those empirically found in Ref. [22], which reports a scaling of modularity in ER graphs as $Q \propto \langle k\rangle^{-2/3}$. We have shown that random graphs may exhibit considerable values of modularity even without any built-in group structure. Recall that the modularity Q has an upper bound of $Q < 1$. Significant community structure can hence only be attributed to graphs with values of modularity higher than those calculated for equivalent random null models. The sparser a graph, the higher the expected modularity. It is therefore particularly difficult to detect true modularity in sparse graphs. Also, the sparser a graph, the more modules it will show, while dense random graphs tend to cluster into only a handful of communities.

5. Theoretical limits of community detection

With the results of the last section we are now prepared to explain Fig. 1 and to give a limit to what extent a designed community structure in a network can be recovered. As we have seen, for any random network we can find an assignment of spins in communities that leads to a modularity $Q > 0$. For our computer-generated test networks with $\langle k\rangle = 16$ we have a value of $p = \langle k\rangle/(N-1) = 0.126$ and expect a value of $Q = 0.227$ according to (8) and $Q = 0.262$ according to (10). The modularity of the community structure built-in by design is given by:

$$Q(\langle k_{\text{in}}\rangle) = \frac{\langle k_{\text{in}}\rangle}{\langle k\rangle} - \frac{1}{4}. \tag{11}$$

Hence, below $\langle k_{\text{in}}\rangle \approx 8$, we have a designed modularity that is smaller than what can be expected from a random network

of the same connectivity! This means that the minimum in the energy landscape corresponding to the community structure that we design is less deep than those that one can find in the energy landscape defined by any network. It must be understood that in the search for the built-in community structure, we are competing with those community structures that arise from the fact that we are optimizing for a particular quantity in a very large search space. In other words, any network possesses a community structure that exhibits a modularity at least as large as that of a completely random network. If a community structure is to be recovered reliably, it must be sufficiently pronounced in order to win the comparison with the structures arising in random networks. In the case of the test networks employed here, there must be more than ≈8 intra-community links per node. Fig. 3 again exemplifies this. We see that random networks with $\langle k\rangle = 16$ are expected to show a ratio of internal and external links $k_{\text{in}}/k_{\text{out}} \approx 1$. Networks which are considerably sparser have a higher ratio while denser networks have a much smaller ratio. This means that in dense networks, we can recover designed community structure down to relatively smaller $\langle k_{\text{in}}\rangle$. Consider for example large test networks with $\langle k\rangle = 100$ with 4 built-in communities. For such networks, we expect a modularity of $Q \approx 0.1$ and hence the critical value of intra-community links to which the community structure could reliably be estimated would be $\langle k_{\text{in}}\rangle_c = 35$ which is much smaller in relative comparison to the average degree in the network.

This also means that the point at which we cannot distinguish between a random and a modular network is not defined by $p_{\text{in}} = p_{\text{out}} = p$ for the internal and external link densities as one may have intuitively expected. Rather, it is determined by the ratio of $\langle k_{\text{in}}\rangle/(\langle k\rangle - \langle k_{\text{in}}\rangle)$ in the ground state of a random network and depends on the connectivity of the network $\langle k\rangle$. This result is important whenever benchmark networks are constructed and the results of various studies on possibly different networks are to be compared. It also shows that the Potts model clustering technique is already close to optimal in the sense that it approaches the theoretical limit. Therefore, research emphasis should be laid upon finding efficient minimization routines for the Potts model Hamiltonian suggested.

From Fig. 3 we observe that sparse random graphs all show communities in the strong sense of Radicchi et al. [12]. Further, it is very difficult to find communities in the strong sense in dense graphs, even though they may exhibit a modularity well above that of a random graph.

Finally, let us examine the most widely studied network for community detection, the Zachary Karate Club [23], which was used by Girvan and Newman in their seminal paper [11] as an example. The network depicts the friendships of the members of a Karate club which eventually broke apart into two almost equally sized groups due to a dispute between the manager and the instructor. It has been used widely as an example of community detection, generally with the goal to reproduce the actual split of the club. The network contains 34 nodes and 77 links, which leads to $\langle k\rangle = 4.53$ and $p = 0.137$. According to Eq. (8), this yields an expectation value of $Q = 0.423$. This

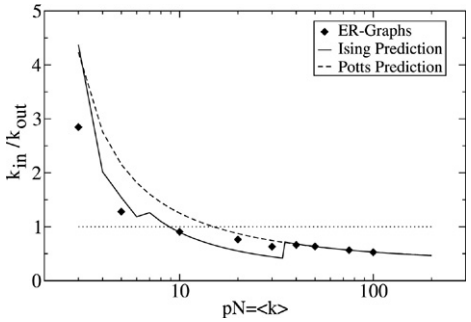

Fig. 3. Ratio of internal links to external links $k_{\text{in}}/k_{\text{out}}$ in the ground state of the Hamiltonian. Shown are the experimental values from clustering random graphs with $N = 10{,}000$ nodes and the expectation values calculated from using a Potts model (8) or an Ising model (10) recursively. The dotted line represents the Radicchi et al. definition of community in "strong sense" [12]. Note that sparse graphs will, on average, always exhibit such communities, while dense graphs will not, even though their modularity may be well above the expectation value for an equivalent random graph.

is surprisingly close to the highest value of 0.419 found in this network as reported by Duch and Arenas [24]. Alternatively, Eq. (9) yields an estimate of 12.8 for the number of links between two equal sized groups of 17 members each. A recent refinement of this estimate from Ref. [25] gives an estimated cut-size of 12.5 links. The actually observed split by Zachary into a group of 18 and one of 16 had a cut-size of 10. The split found by Girvan and Newman was a partition into a group of 19 and 15 with the same cut-size of 10. It must be understood that the analytical results obtained in the thermodynamic limit are not exactly applicable to such a small network and that a rewiring technique would be more appropriate for such a small system. This said, it still seems that the Karate club network is not truly modular in the sense that one cannot find partitions that considerably exceed the expectation values of Q for equivalent random null models. There exist only few alternative assignments of comparable modularity which overlap only in a small number of nodes. This means the community structure of maximum modularity is very stable. For such a small system, this is, however, expected. Also, one cannot say that the cut-size of the observed split differs drastically from the expected in an equivalent random network. This leads to the following conclusion: from a pure network perspective, we cannot say that there were different groups of people in the club that formed communities into which the club eventually broke apart. Rather, it was the central position of the conflicting individuals, their being highly connected and their specific positions as manager and instructor which lead to the breakup of the club. This breakup then happened along a minimal cut which is only natural. A dispute between any other pair of members would most likely not have led to a splitting of the group.

6. Conclusion

In this article, we have examined the problem of assessing statistically significant modular structure in networks. We exploited a mapping of the problem of community detection onto finding the ground state of an infinite range spin glass. The

quality function of the clustering, or modularity, is identified as the ground state energy of this spin glass. Benchmarks show the good performance of algorithms based on this mapping. Expectation values for the modularity of Erdös–Rényi random graphs were given and the dependence of these expectation values on the link density in the network was discussed. The theoretical limits of community detection were addressed. We found that only those community structures can be recovered reliably that lead to modularities larger than the expectation values of random graphs. This allowed us to quantitatively explain the observed deterioration of benchmark results for computer generated test networks and shed some light on the interpretation of widely used real world example networks. Our findings are universally applicable and independent of any algorithm used to find the community structure.

Acknowledgements

The authors would like to thank Stefan Braunewell, Michele Leone, Ionas Erb and Andreas Engel for many helpful hints and discussions.

References

[1] R. Albert, A.-L. Barabàsi, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[2] M. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[3] M. Newman, Detecting community structure in networks, Eur. Phys. J. B 38 (2004) 321.
[4] L. Danon, J. Duch, A. Arenas, A. Diaz-Guilera, Comparing community structure identification, J. Stat. Mech. (2005) P09008.
[5] R. Guimera, L.A.N. Amaral, Functional cartography of complex metabolic networks, Nature 433 (2005) 895–900.
[6] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (2005) 814.
[7] J. Reichardt, S. Bornholdt, Ebay users form stable groups of common interest, preprint physics/0503138.
[8] A. Engel, C.V. den Broeck, Statistical Mechanics of Learning, Cambridge University Press, 2001.
[9] C. Bonferroni, Il calcolo delle assicuranzioni su gruppi die teste, Studi in Onore del Professore Salvatore Ortu Carboni, 1935, pp. 13–60.
[10] S. Wasserman, K. Faust, Social Network Analysis, Cambridge University Press, 1994.
[11] M. Newman, M. Girvan, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99 (12) (2003) 7821–7826.
[12] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Defining and identifying communities in networks, Proc. Natl. Acad. Sci. USA 101 (2004) 2658.
[13] I. Derényi, G. Palla, T. Vicsek, Clique percolation in random networks, Phys. Rev. Lett. 94 (2005) 160202.
[14] M. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.
[15] M. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (2004) 066133.
[16] J. Reichardt, S. Bornholdt, Statistical mechanics of community detection, Phys. Rev. E 74 (2006) 016110.
[17] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004) 066111.
[18] J. Reichardt, S. Bornholdt, Detecting fuzzy community structures in complex networks with a Potts model, Phys. Rev. Lett. 93 (2004) 218701.
[19] P. Erdös, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17–61.
[20] Y. Fu, P.W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation, J. Phys. A: Math. Gen. 19 (1986) 1605–1620.
[21] I. Kanter, H. Sompolinsky, Graph optimisation problems and the Potts glass, J. Phys. A: Math. Gen. 20 (1987) 636–679.
[22] R. Guimera, M. Sales-Pardo, L.N. Amaral, Modularity from fluctuations in random graphs and complex networks, Phys. Rev. E 70 (2004) 025101R.
[23] W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. 33 (1977) 452–473.
[24] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E 72 (2005) 027104.
[25] J. Reichardt, S. Bornholdt, Partitioning and modularity in graphs with arbitrary degree distribution, preprint: arxiv/cond-mat/0606295 (submitted for publication).