Consistency of community structure in complex networks

Maria A. Riolo[1, 2] and M. E. J. Newman[1, 2, 3]

[1] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico, USA*
[2] *Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan, USA*
[3] *Department of Physics, University of Michigan, Ann Arbor, Michigan, USA*

The most widely used techniques for community detection in networks, including methods based on modularity, statistical inference, and information theoretic arguments, all work by optimizing objective functions that measure the quality of network partitions. There is a good case to be made, however, that one should not look solely at the single optimal community structure under such an objective function, but rather at a selection of high-scoring structures. If one does this one typically finds that the resulting structures show considerable variation, which could be taken as evidence that these community detection methods are unreliable, since they do not appear to give consistent answers. Here we argue that, upon closer inspection, the structures found are in fact consistent in a certain way. Specifically, we show that they can all be assembled from a set of underlying "building blocks," groups of network nodes that are usually found together in the same community. Different community structures correspond to different arrangements of blocks, but the blocks themselves are largely invariant. We propose an information theoretic method for discovering the building blocks in specific networks and demonstrate it with several example applications. We conclude that traditional community detection does in fact give a significant amount of insight into network structure.

I. INTRODUCTION

Many networks, from social and information networks to biological networks and the Internet, are found to divide into distinct groups of nodes, referred to variously as modules, clusters, or communities [1, 2]. Community detection—the process of identifying such groups in unlabeled network data—is widely used as an analytical tool for exploring the large-scale structure of complex networks. Many algorithms for community detection have been proposed, but the most widely used ones all share one feature in common: they operate by optimizing some kind of objective function that measures the quality of candidate divisions of a network into communities. Perhaps the most widely used method is modularity maximization, as embodied for instance in the spectral modularity and Louvain algorithms, which work by optimizing the heuristic objective function known as modularity [3–5]. Inference methods, such as methods based on the stochastic block model, work by optimizing the likelihood of the observed network under an appropriate network model [6–8]. The widely used InfoMap method works by maximizing the entropy of a random walk on the network [9].

However, as pointed out by a number of authors [10–12], simply reporting the single best division of a network, as defined by an objective function, misses much of the insight that is to be gained from community analysis. In many networks, perhaps most, there are multiple divisions of the nodes that achieve high objective-function scores and in principle any of these could be the "correct" division of the network. It is a crucial question whether these competing divisions are, in some sense, similar to one another or whether, conversely, they are substantially different. If all (or most) high-scoring divisions are similar, then we may hypothesize that the community analysis is revealing some genuine underlying truth about the network: even if we don't know which of several candidate divisions is the correct one, we may still be able to draw insight from them if the candidates all tell essentially the same story. On the other hand, if the high-scoring divisions are quite different from one another then it is harder to argue that they are meaningful.

As an example, it is known that even completely random networks, such as Erdős–Rényi style random graphs, have divisions with high modularity scores [10, 13], yet such random networks have no community structure by any reasonable definition. Massen and Doye [11] generated a selection of high-modularity divisions of random graphs by Monte Carlo sampling and found that competing divisions of the same graph had little common structure, suggesting that they are probably not meaningful—a reasonable conclusion in the case of a random graph. Subsequent theoretical work has bolstered this viewpoint using ideas borrowed from the physics of glassy systems. If we consider the modularity as an energy function for a thermal model, then the random graph can be shown to undergo a transition with decreasing temperature to a replica symmetry broken state where there are many competing modularity maxima that correspond to essentially unrelated divisions [13–16].

In many real-world networks, by contrast, as well as certain model networks such as the stochastic block model, it is believed that there is clear and meaningful community structure, which we would like to be able to extract and analyze with our algorithms. In these cases we would hope that, to the extent that there are competing divisions with high scores, those divisions would be largely similar to one another, at least in their gross features. Thus, the existence of true community structure would be associated with the observation that high-scoring divisions are similar and its absence with the ob-

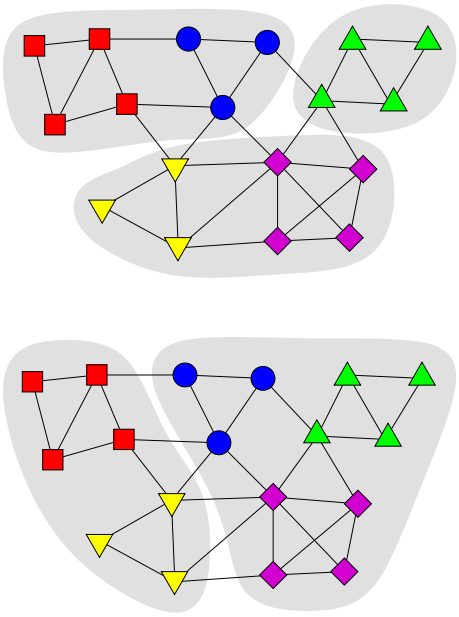

FIG. 1: Two divisions of the same set of network building blocks. The five building blocks are denoted by the shapes and colors of the nodes and the community divisions are denoted by the shaded areas. Each community division can be thought of as a different way of assembling the building blocks into communities.

servation that they are different. Equivalently, true community structure would correspond to replica symmetry and lack of it to replica symmetry breaking.

Unfortunately, though this picture seems intuitive and reasonable, it has been found not to apply in many real-world situations. For example, Good *et al.* [17] generated Monte Carlo samples of high-modularity divisions for a range of networks, including both models and real-world examples, and found in all cases that even though the networks in question were believed to possess strong community structure there were nonetheless a large number of high-scoring divisions that appeared to be quite different. This raises serious questions about whether our community detection algorithms are returning meaningful results.

In this paper we revisit this question and argue that in fact the high-scoring divisions of many networks *are* similar, but in a more subtle sense. Specifically, we show that while it is true that the communities discovered by these algorithms vary substantially between high-scoring divisions, the variation is of a limited and specific type. We show that for both real and model networks it is possible to find an elemental set of "building blocks," groups of nodes such that most high-scoring community divisions are formed by combining these blocks in one way or another, while the blocks themselves are essentially indivisible—see Fig. 1 for a sketch. Thus most high-scoring community divisions are similar in the sense of being built from the same set of building blocks.

To put this another way, if we know the blocks then it takes very little additional information to specify how they are joined together and hence specify the complete community structure. We use this observation to create an information-theoretic algorithm for determining the building blocks and demonstrate its use on a range of example networks. Though our results are numerical only and hence do not constitute a proof, we find that the algorithm gives convincing and plausible results when applied to synthetic networks with known blocks, synthetic networks known to have no blocks, and real-world networks. We conclude that community structure analyses do in fact convey consistent and believable information about the large-scale structure of networks, when interpreted in an appropriate manner.

II. SAMPLING NETWORK DIVISIONS

Like the previous studies discussed above, our investigation starts with the generation of a random sample of network divisions that score highly according to an appropriate objective function. Previous studies sampled divisions according to modularity, but this approach is arguably somewhat ad hoc: there is no rigorous principle that tells us the relative sampling weight one should give to divisions with different modularity. Massen and Doye and others [11, 16–18] have employed a Boltzmann distribution, which is convenient for numerical simulation but does not have a formal justification in this context. In our work we use an alternative approach that has become popular in recent years, that of sampling from the posterior distribution of an appropriate generative model. The model we use, which is standard in calculations of this kind, is the degree-corrected stochastic block model [8], a random graph model in which the probabilities of edges depend on the communities they belong to. Inverting the probability relation using Bayes' rule allows us to calculate the probability of a particular community division given an observed network and it is from this distribution that we sample. Specifically, the approach is as follows. (This part of the paper follows the outline of our previous presentation in [19]—see that paper for additional details.)

The degree-corrected stochastic block is a random generative model of a community-structured network. When used to generate networks (rather than for community detection), it works as follows. Initially, each of n nodes is assigned to one of k groups, then a Poisson-distributed number of edges is added between each node pair such that the expected number of edges between nodes i and j is $\theta_i\theta_j\omega_{g_ig_j}$, or a half this many when $i = j$, where θ_i and ω_{rs} are parameters that we choose and g_i is the community to which node i belongs. This leaves θ_i and ω_{rs} arbitrary to within multiplicative constants, which are fixed by normalizing the θ_i such that their mean is 1 in

each community thus:

$$\frac{1}{n_r}\sum_{i=1}^{n}\theta_i\delta_{r,g_i} = 1, \tag{1}$$

where δ_{ij} is the Kronecker delta and $n_r = \sum_i \delta_{r,g_i}$ is the number of nodes in group r.

Note that in this model the number of edges between each node pair is Poisson distributed, meaning there can in principle be multiple edges between the same nodes. In some respects this is unrealistic—most real-world networks have only single edges between nodes—but for the common case of sparse networks the probability of multiple edges is small and can usually be neglected, and the Poisson model has technical advantages that have made it the accepted formulation for community detection applications.

Now consider a specific undirected network with structure described by its adjacency matrix $\mathbf{A}$ having elements $A_{ij} = 1$ if there is an edge between nodes i and j and 0 otherwise. The probability, or likelihood, that this network is generated by the degree-corrected stochastic block model defined above is

$$\begin{aligned} P(\mathbf{A}|\theta,\omega,g,k) &= \prod_{i<j}\bigl(\theta_i\theta_j\omega_{g_ig_j}\bigr)^{A_{ij}} e^{-\theta_i\theta_j\omega_{g_ig_j}} \\ &\qquad\times\prod_i\bigl(\tfrac12\theta_i^2\omega_{g_ig_i}\bigr)^{A_{ii}/2} e^{-\theta_i^2\omega_{g_ig_i}/2} \\ &= \prod_i\theta_i^{d_i}\prod_{r<s}\omega_{rs}^{m_{rs}}e^{-n_rn_s\omega_{rs}}\prod_r\omega_{rr}^{m_{rr}}e^{-n_r^2\omega_{rr}/2}, \end{aligned} \tag{2}$$

where we have used Eq. (1) in the second equality, $d_i = \sum_j A_{ij}$ is the degree of node i, and

$$m_{rs} = \begin{cases} \sum_{ij}A_{ij}\delta_{g_i,r}\delta_{g_j,s} & \text{when } r\neq s, \\ \frac12\sum_{ij}A_{ij}\delta_{g_i,r}\delta_{g_j,r} & \text{when } r=s, \end{cases} \tag{3}$$

which is the number of edges between groups r and s. We have also neglected an overall multiplying factor in Eq. (2) which is independent of the parameters θ, ω, and g, and will therefore have no effect on our calculations.

The values of the parameters θ and ω are not of interest in the present case, so we integrate them out using maximum-entropy priors as described in [19], to get

$$\begin{aligned} P(\mathbf{A}|g,k) &= \prod_r n_r^{\kappa_r}\frac{(n_r-1)!}{(n_r+\kappa_r-1)!} \\ &\qquad\times\prod_{r<s}\frac{m_{rs}!}{(pn_rn_s+1)^{m_{rs}+1}}\prod_r\frac{m_{rr}!}{(\frac12pn_r^2+1)^{m_{rr}+1}}, \end{aligned} \tag{4}$$

where

$$\kappa_r = \sum_i d_i\delta_{r,g_i} \tag{5}$$

and we have discarded a further multiplying constant. Now we apply Bayes' rule to get

$$P(g,k|\mathbf{A}) = \frac{P(\mathbf{A}|g,k)P(g,k)}{P(\mathbf{A})}. \tag{6}$$

The denominator $P(\mathbf{A})$ is a simple normalizing constant that plays the role of a partition function and, like other constants, will not be important for our calculations. For the prior probability $P(g,k)$ we again follow our previous work, making the choice $P(g,k) = n^{-k}\prod_r n_r!$, which is derived from a simple "restaurant process" [19]. With this choice, and again neglecting overall constants, we have

$$\begin{aligned} P(g,k|\mathbf{A}) &= n^{-k}\prod_r n_r^{\kappa_r}\frac{n_r!(n_r-1)!}{(n_r+\kappa_r-1)!} \\ &\qquad\times\prod_{r<s}\frac{m_{rs}!}{(pn_rn_s+1)^{m_{rs}+1}}\prod_r\frac{m_{rr}!}{(\frac12pn_r^2+1)^{m_{rr}+1}}. \end{aligned} \tag{7}$$

We now generate community divisions (g,k) from this distribution using Metropolis–Hastings Monte Carlo sampling. Our sampling algorithm, which makes specific use of the structure of the prior on g and k to enhance sampling speed, is described in detail in [19]. The implementation, which is written in the C programming language, performs about 1 million Monte Carlo steps per second on a typical desktop computer, allowing our calculations to scale to networks of tens or hundreds of thousands of nodes with relative ease, although we will have no need of such large networks in this paper.

It is worth mentioning that the methods employed in this paper to study the building blocks of community structure are agnostic about the particular scheme we use for sampling community divisions. While we favor the algorithm described here for its principled statistical foundations, other methods such as modularity-based sampling could be employed as well. One could also make other choices of the prior $P(g,k)$ on the community divisions, such as the hierarchical prior of Peixoto [20], which allows for different probabilities for structures at different scales. The simple prior we use is representative of typical applications of community detection in the literature, making it a good choice for the examples we give, but others could easily be applied. Indeed, there is some overlap between our calculations of building blocks and Peixoto's hierarchical method, which aims to find communities at all scales simultaneously. In some cases the smaller communities found in his work may be similar to the blocks we find, although given the differing nature and goals of our approaches it seems certain there will be differences as well.

III. RESULTS

Our goal is to use the algorithm described above to generate a random sample of high-probability commu-

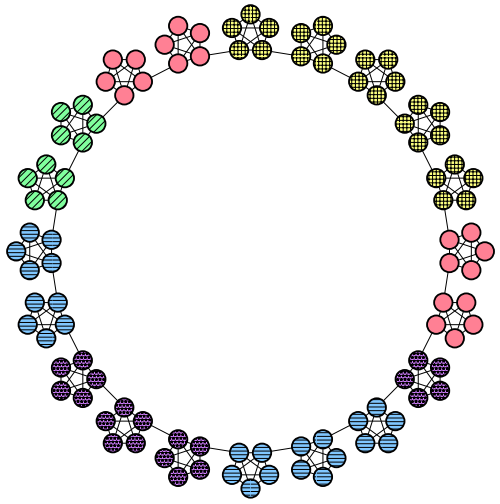

FIG. 2: The model network network of Good *et al.* [17] with nodes colored to indicate the highest probability division found over 20 runs of 10^7 steps of the Monte Carlo algorithm described in Section II.

nity divisions and then compare the structure of those divisions to try to determine what features they have in common. As described in the introduction, we find that in most cases they can be represented as the union of a collection of elemental and largely indivisible blocks of nodes that appear to represent the fundamental "atoms" of community structure in the network.

A. An example model network

To illustrate our approach we take as a first example a simple model network proposed by Good *et al.* [17] precisely as an illustration of the issues discussed in the introduction. This network, which is shown in Fig. 2 and is similar to the "connected caveman" model of Watts [21], is composed of a number of cliques (i.e., completely connected subgraphs) joined together in a ring. In the example shown in the figure there are 20 cliques of five nodes each.

We now apply our Monte Carlo sampling algorithm to this network. Figure 2 shows the highest probability division found over twenty runs of the algorithm. The division is perhaps not the one we would at first guess—it is not the division into the 20 cliques themselves. Instead, as the figure shows, the algorithm has divided the network into five groups of varying size. The cliques themselves are still intact—none of them has been split between communities—but some cliques have been joined together to make larger communities of 10, 20, or even 25 nodes.

As we will see, this result is typical. There are natural blocks of nodes in many networks that want to be in the same community—the cliques in this case. (A somewhat similar idea has been highlighted by Chakraborty *et al.* [22], who noted the existence of subsets of nodes in networks that were often assigned to the same community.) However these blocks are, in most cases, not themselves communities. The communities are assembled by putting blocks together. Moreover, it is easy to see in this case that there are many ways of putting the blocks together that are as good as the one shown in Fig. 2, or nearly so. For instance, since the network has a discrete rotational symmetry around the ring, there are trivially 20 rotational variants of the division shown that have the exact same probability but which join the blocks in different ways. The result is that if one samples many high-scoring divisions of the network one will see the same blocks repeatedly but not necessarily the same communities. Indeed the communities can change dramatically from one state of the sampling algorithm to another: large pieces can shear off and form their own community, or join another. If one were to compare different community divisions, therefore, particularly using elementary numerical measures of similarity such as the Rand index, one might conclude that there was wide variation between divisions and little consistency—and hence that the algorithm was not giving useful information about network structure. This, however, would be a mistake. Once we understand the nature of the building blocks from which the communities are assembled we see that the structures sampled by the algorithm are in a sense highly similar and consistent.

One way to make these observations more quantitative is illustrated in Fig. 3. In panel A of the figure we demonstrate that the individual cliques in the network are rarely split between communities. The plot shows a histogram of the probability that each pair of nodes in the network find themselves in the same community, averaged over a large number of divisions of the network sampled using the Monte Carlo algorithm. (Such probabilities have been studied in the past—see for example Ref. [23].) The histogram is colored according to the distance between cliques, where distance 0 means node pairs in the same clique, distance 1 means adjacent cliques, and so forth. As we can see, nodes in the same clique have probability close to 1 of being in the same community, but nodes at all other distances have substantially lower probability.

In Fig. 3B we show another representation of the same probability measurements, a density plot of the pairwise probabilities. This plot clearly picks out the individual building blocks of the network as the dark squares along the diagonal of the figure. If we did not already know what the blocks were for this network, we could deduce them by examining this figure.

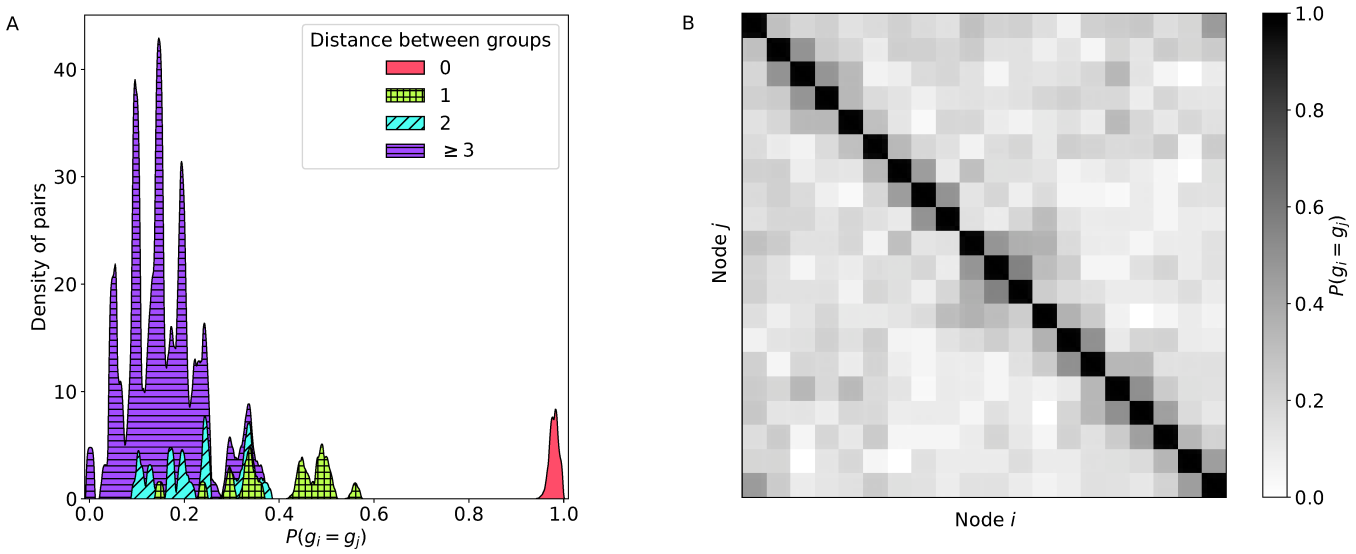

FIG. 3: (A) The distribution of the probabilities $P(g_i = g_j)$ that two nodes i and j are found to be in the same community. We show separate histograms for node pairs in the same clique ("distance 0"), adjacent cliques ("distance 1"), and so forth. (B) A density plot of the same set of probabilities. The ground-truth cliques in the network are clearly visible as the dark squares along the diagonal.

B. Social network

Let us now apply the same ideas to a more complex example. In Fig. 4 we show the results of applying the algorithm of Section II to a standard and widely studied network from the community detection literature, the social network of fictional characters in the novel *Les Misérables* by Victor Hugo [24], which provides a good illustration of the phenomena discussed above. As an initial test of the method we perform a single run of our algorithm for over a million Monte Carlo steps and then select four high-probability states from the latter portion of the run as shown in Fig. 4A. The first state has the highest probability of the four, but the others are also competitive. Panels B to E in Fig. 4 show the community divisions found in each of the four cases.

As we can see from the figure, the four divisions have much in common, but there are also substantial differences between them. From one panel to the next, groups of nodes break off from communities and join others in a manner similar to that of the previous section. To get a clearer picture of these changes we show in Fig. 5 two different comparisons of the four divisions. In Fig. 5A we show the network with nodes colored in pie-chart fashion to indicate which communities they belong to in each of the four divisions. In Fig. 5B we perform a simple reconstruction of the building blocks of the network by assigning a different color to each set of nodes that are found together in the same community in all four divisions. Each of the four divisions is composed of combinations of these elemental groups, and yet these groups are not themselves communities. Consider, for example, the group of nodes in Fig. 5B that are colored yellow with horizontal stripes. This group does not form a standalone community in any of the four divisions pictured in Fig. 4B–E, yet whatever group they fall in they are always found together.

This approach for reconstructing the building blocks is, however, somewhat ad hoc. Ideally we would prefer a more rigorous method. We describe such a method in the next section.

C. Choosing optimal building blocks

How can we make the notion of building blocks for community structure more rigorous? One approach is to think in terms of information content. A good set of building blocks is one that describes most of the structure in most commonly occurring community divisions, meaning that given the building blocks only a small amount of additional information is needed to define a division. For instance, we could describe a community division by first specifying to which community each block belongs and then specifying a (hopefully small) set of corrections to the resulting division for any nodes that we put in the wrong community.

Information theory tells us that in general the amount of information needed to specify the community structure given the blocks is equal to the conditional entropy of the latter given the former and we may consider a particular set of building blocks to be "good" if the con-

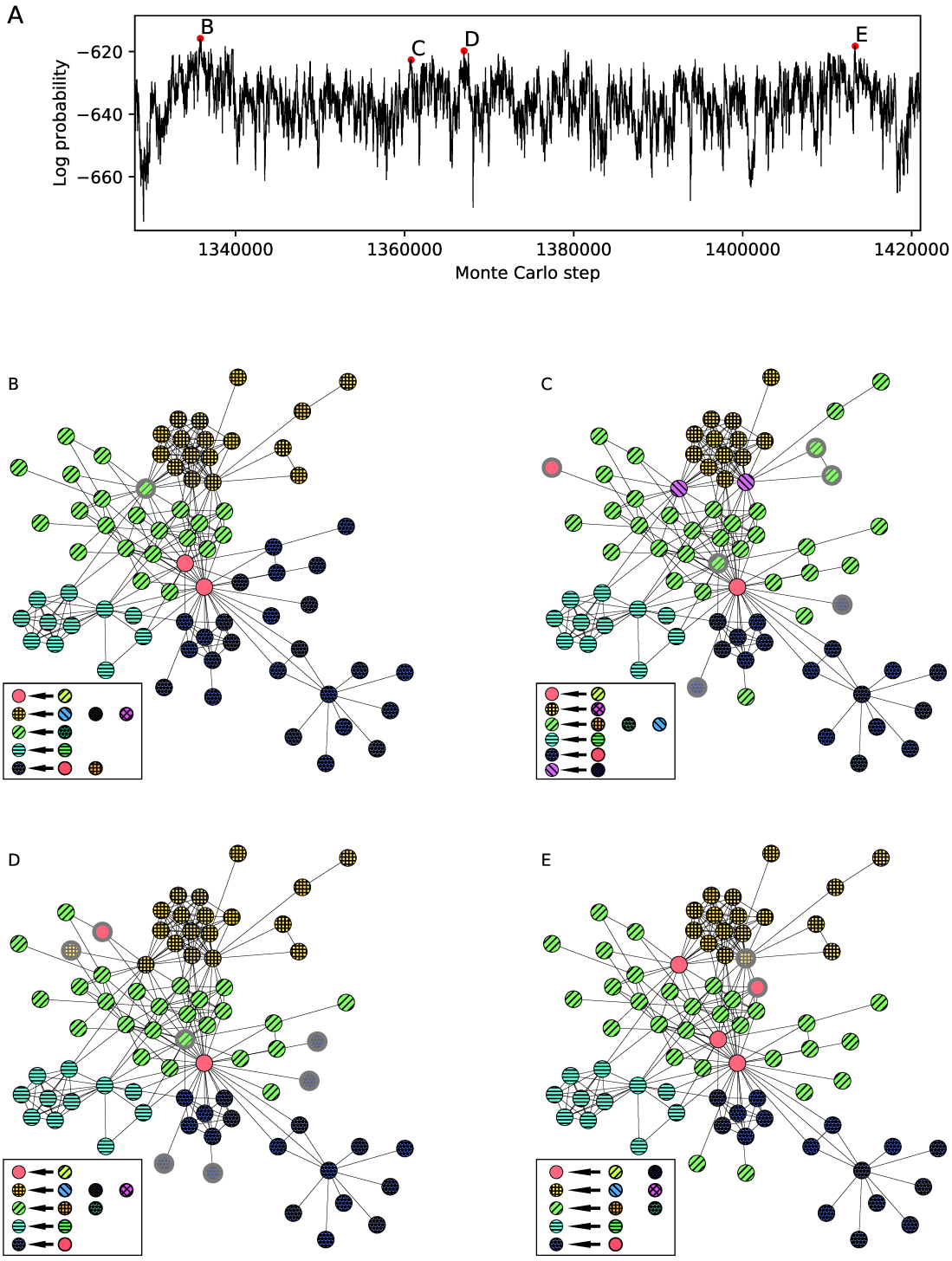

FIG. 4: Results from a single run of our Monte Carlo community detection algorithm on the social network of fictional characters from the novel *Les Misérables* [24]. In panel A, we show the log-likelihood of states visited as a function of time for a portion of the run. Selected peaks in the likelihood are labeled B to E and the community assignments at these peaks are shown in the lower four panels. Inset in each of these four panels is a legend showing how the communities discovered by the algorithm correspond to the building blocks shown in Fig. 6. These mappings reproduce the community structures well but not perfectly—a few nodes are placed in the wrong community and we have highlighted these nodes in each panel.

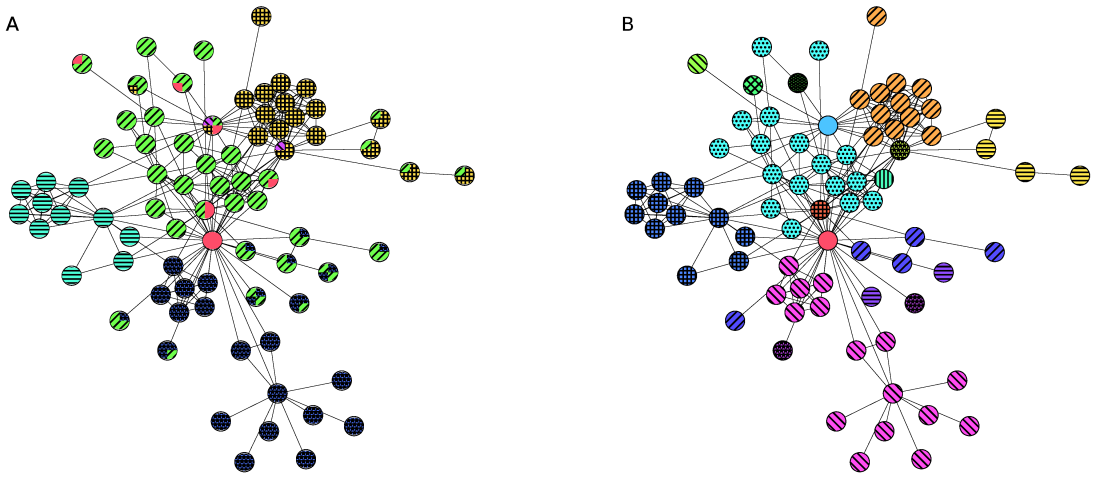

FIG. 5: Comparison of the four divisions of the *Les Misérables* social network pictured in Fig. 4. In (A) the nodes are colored in pie-chart fashion to indicate which community they belong to in each division. In (B) we assign new group labels so that only nodes that belonged to the same group in all four assignments are grouped together. This gives us a simple estimate of the identity of the building blocks that make up the community divisions.

ditional entropy is small when averaged over the distribution of community divisions, Eq. (7), or (more practically) a suitable set of divisions sampled using Monte Carlo. As we will see, it is indeed possible to find such building blocks.

In practice it is conventional to use not the conditional entropy for comparing network divisions but the mutual information, which is a simple linear transform of the conditional entropy that inverts the information scale so that divisions are maximally similar for maximum mutual information. In our calculations we make use of the "reduced mutual information" of Ref. [25], which includes a correction term that allows for accurate computation of the amount of information even in cases where, as here, the number of blocks may be very different from the number of communities. (Other information measures, such as the "variation of information" [26], might also work in this context, but we favor the reduced mutual information since it is tailored directly to answering the specific question at hand here.) The details are as follows.

Consider a network of n nodes and a specific community division of that network with k communities, and suppose that we have q building blocks. Let g_i represent the community to which node i belongs as previously, and let h_i represent the building block. We define a *contingency table*, which is the matrix of elements c_{gh} equal to the number of nodes that belong to community g and block h. Then the *reduced mutual information* is given by

$$M = \frac{1}{n} \log \frac{n! \prod_{gh} c_{gh}!}{\prod_g a_g! \prod_h b_h!} - \frac{1}{n} \log \Omega, \tag{8}$$

where $a_g = \sum_h c_{gh}$ and $b_h = \sum_g c_{gh}$ are the row and column sums of the contingency table and Ω is the number of distinct possible contingency tables that have these row and column sums. Roughly speaking, the second term—the term in $\log \Omega$—represents the amount of information needed to determine which community each block belongs to and the first term represents the amount of information needed to subsequently specify the full community structure once the blocks have been assigned. The optimal choice of building blocks is the one that maximizes (8) when averaged over community structures, in other words the one that contains maximal information about those structures on average.

In the limit where we have only a single building block, we have $b_h = n$ and $c_{gh} = a_g$, so the first term in (8) is equal to $\log 1$, and vanishes. At the same time there is only one possible contingency table, so $\Omega = 1$ and the second term vanishes also. Thus we have $M = 0$ in this limit. At the other extreme, where every node is its own building block, we have $c_{gh} = 0$ or 1 for all g, h, and $b_h = 1$. Hence the first term in (8) is equal to $\log(n!/\prod_g a_g!)$. At the same time, the number of contingency tables is equal the number of ways one can assign the n one-node blocks to communities of the given sizes a_g, which is given by the multinomial coefficient $\Omega = n!/\prod_g a_g!$. Hence in this case the two terms in (8) are exactly equal to one another and again we have $M = 0$. For all other choices of blocks we expect M to take a value larger than zero, so somewhere in between the two limits must lie the optimal choice of building blocks.

Our goal is to maximize the Monte Carlo average of Eq. (8) over possible choices of the blocks. Exhaustive maximization is impractical in most cases because the number of choices is exponentially large in the size of the network, so instead we use an approximate greedy algorithm, which in practice seems to work well. The algorithm starts with every node in a block on its own,

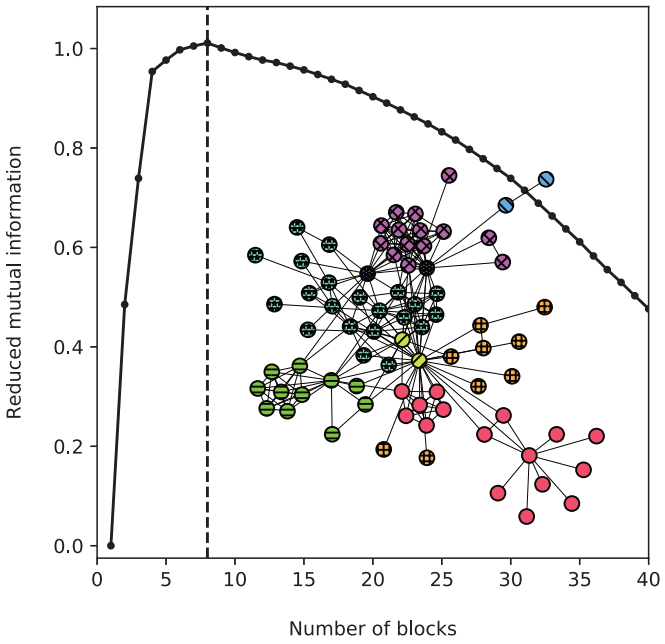

FIG. 6: Main figure: The reduced mutual information, which describes the average information about communities that is contained in the building blocks, for various different numbers of blocks, averaged over 10 000 sampled community structures for the *Les Misérables* network. As expected, the value is small for the case of very many blocks or very few, but there is an intermediate maximum, in this case at eight blocks, where the blocks contain the most complete description of the average community structure. Inset: The structure of the eight blocks at the mutual information peak.

giving $M = 0$, then joins together the two blocks that most increase (or least decrease) the value of M. We repeat this process, joining blocks in pairs until all blocks have been joined into one and the value of M is once again zero. The intermediate state that we pass through with the largest value of M is then taken to be our block division for the network.

Figure 6 shows the results of this approach applied to our social network example. The main plot shows the value of the reduced mutual information as a function of the number of blocks over the course of the calculation. The plot has the expected form, with the value increasing to a maximum then falling off again. The maximum value occurs for the case of eight blocks and the corresponding block structure is shown inset.

This choice of blocks does appear to be a good one. Referring back to Fig. 4, we show in the insets of panels B to E a key that gives the mapping from blocks to communities for each of the structures depicted. In each case it is possible to describe the entire community structure by saying to which community each block belongs, except for a small number of nodes, at most seven in any case (shown in bold), that do not fit the pattern. Thus, while our community detection algorithm does indeed return a range of different divisions for this network, it is at the same time correct to say that those divisions all reflect essentially the same underlying structure in the network, since it is possible to express them as combinations of the same set of basic building blocks.

We note in passing an interesting feature of the blocks shown in Fig. 6B: some of them are not connected, meaning they consist of two or more parts with no edges between parts. This arises because our community detection algorithm is capable of finding disassortative structure in the network as well as assortative structure. That is, it finds not only groups with a higher-than-expected number of edges, but also groups with a lower-than-expected number. Some of the groups found in this case fall into the latter category and this is then reflected in the building blocks too.

D. Random networks

One possible objection to the approach taken here is that the method we describe might give a similar division into blocks for *any* network, whether the communities found are significant or not. How can we know that the algorithm is returning meaningful blocks? To shed light on this question we can apply the same analysis approach to randomly generated networks, for which we expect there to be no meaningful blocks, despite the fact that (as discussed in the introduction) there are typically many high modularity partitions of the network [10, 13]. Because these partitions are uncorrelated [11, 13] we would expect there to be no informational advantage to describing them as a collection of blocks.

To test this hypothesis, we apply our method to a large set of randomly generated networks, consisting of Erdős-Rényi style random graphs with mean degree 10 and random 3-regular graphs, with $n = 2000$ in each case. We sample partitions using the Monte Carlo method of Refs. [11, 16–18] then compute the optimal blocks for each network by maximizing the reduced mutual information as before. In each case, we find that the reduced mutual information decreases monotonically as the number of blocks increases, and hence that, in effect, there are no blocks in these networks: the optimal choice of blocks is always to put all nodes in one block together. This lends credence to the idea that when we find a nontrivial block structure, the blocks—and hence the community divisions from which they are inferred—are in fact meaningful and contain significant information about the network.

E. Further examples

Let us return to our first example, the "ring of cliques" network shown in Fig. 2. Our claim in Section III A was that the building blocks of this network were the cliques themselves, even though the communities found by the community detection algorithm are mostly larger than a

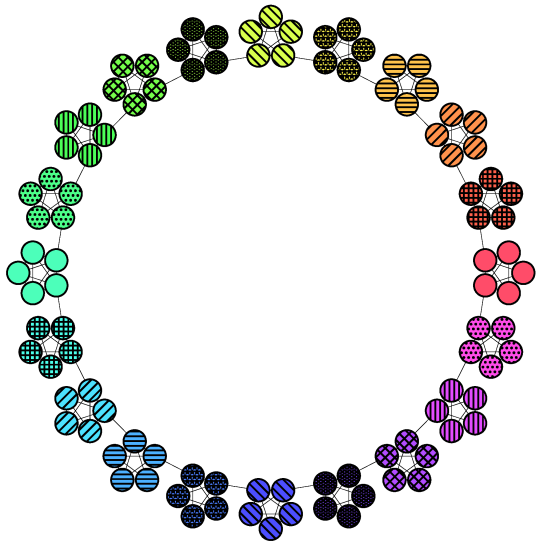

FIG. 7: Building blocks for the network of Fig. 2, found by maximizing the reduced mutual information as described in Section III C. In this case the building blocks correspond exactly to the cliques within the network, as we would expect.

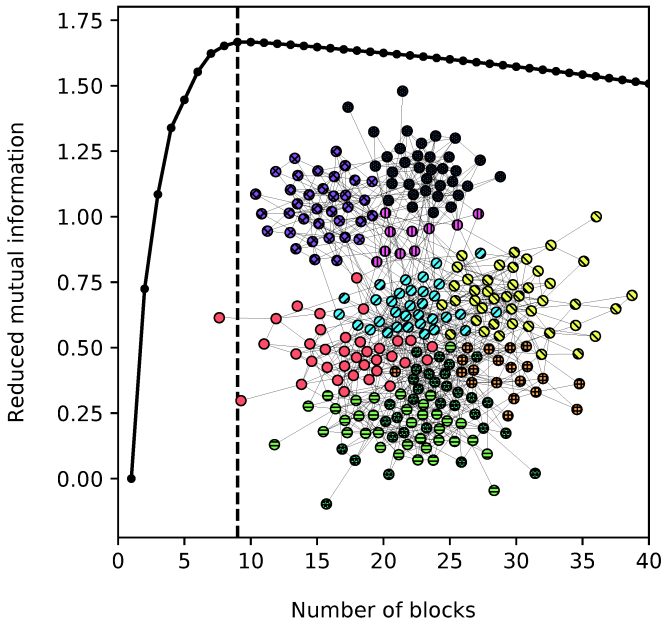

FIG. 8: Main figure: The reduced mutual information of Eq. (8) for a social network of friendships among high-school students, averaged over 10 000 sampled community structures. The mutual information has its maximum value when there are nine building blocks. Inset: The structure of the blocks at the mutual information peak.

single clique. If this were true, and if the method of the previous section is indeed able to find the building blocks of a network, then when that method is applied to this network it should find the cliques. And indeed it does. Figure 7 shows the optimal choice of building blocks for this network constructed using the greedy algorithm of the previous section and, as the figure shows, they correspond exactly to the 20 cliques in the network. (We find equivalent results for networks with other numbers of cliques as well.)

A more complex and realistic example is shown in Fig. 8. This network comes from the National Longitudinal Study of Adolescent to Adult Health (the "Add Health study"), a nationwide US social network study of friendship and dating behavior among middle- and high-school students (approximately ages 12 to 18 years). The network pictured is a friendship network of self-identified friendships between students in one school out of the many that participated in the study. (The school was picked primarily for its smaller size, which makes it easier to visualize the results.) The main panel of Fig. 8 shows again the reduced mutual information as a function of number of blocks and for this network we see that the maximum value is reached for nine blocks. Inset in the figure we show the corresponding set of blocks in the network, which in this case are all relatively compact sets of nodes.

IV. CONCLUSIONS

In this paper we have examined community structure in complex networks using a Monte Carlo algorithm that samples high-likelihood structures. We find, as a number of previous authors have also, that the typical network possesses good community divisions with a wide range of different structures. We also observe, however, that the competing structures are all related to one another in a relatively simple manner, namely they are all built from a small set of "building blocks," groups of nodes that typically appear together in the same community. The building blocks are not themselves communities in most cases, but complete community structures are formed by joining blocks together in various combinations.

This suggests that the amount of information needed to specify the community structure should be small once the building blocks are known, and we use this fact to create an algorithm that can determine the blocks for any given network. Starting from a large sample of plausible community structures generated by our Monte Carlo algorithm, we compute (a variant of) the mutual information between a proposed set of blocks and the community structure, averaged over all samples. The optimal choice of blocks is the one that maximizes this average. We find the maximum using an approximate greedy algorithm and we are able in this way to accurately recover the known blocks in a previously proposed class of test networks. We also give example applications to real-world social networks.

The lesson behind these findings is that, while the existence of large sets of competitive and apparently disparate community structures in real and model networks

appears at first to be a bad sign for community detection algorithms, the situation is actually more promising than it seems. The observed structures are, in essence, all variants of the same basic template, and the complete set of community divisions in fact provides significant information about the large-scale structure of the network.

Acknowledgments

The authors thank George Cantwell, Aaron Clauset, Alec Kirkley, and Jean-Gabriel Young for useful conversations. This work was funded in part by the James S. McDonnell Foundation (MAR) and by the US National Science Foundation under grant DMS–1710848 (MEJN).

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

[2] S. Fortunato, Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).

[3] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).

[4] M. E. J. Newman, Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).

[5] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).

[6] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps. *Social Networks* **5**, 109–137 (1983).

[7] K. Nowicki and T. A. B. Snijders, Estimation and prediction for stochastic blockstructures. *J. Amer. Stat. Assoc.* **96**, 1077–1087 (2001).

[8] B. Karrer and M. E. J. Newman, Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).

[9] M. Rosvall and C. T. Bergstrom, An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. USA* **104**, 7327–7331 (2007).

[10] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**, 025101 (2004).

[11] C. P. Massen and J. P. K. Doye, Thermodynamics of community structure. Preprint cond-mat/0610077 (2006).

[12] A. Clauset, C. Moore, and M. E. J. Newman, Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).

[13] J. Reichardt and S. Bornholdt, When are networks truly modular? *Physica D* **244**, 20–26 (2006).

[14] J. Reichardt and M. Leone, (Un)detectable cluster structure in sparse networks. *Phys. Rev. Lett.* **101**, 078701 (2008).

[15] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84**, 066106 (2011).

[16] P. Zhang and C. Moore, Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. *Proc. Natl. Acad. Sci. USA* **111**, 18144–18149 (2014).

[17] B. H. Good, Y.-A. de Montjoye, and A. Clauset, Performance of modularity maximization in practical contexts. *Phys. Rev. E* **81**, 046106 (2010).

[18] J. Reichardt and S. Bornholdt, Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).

[19] M. A. Riolo, G. T. Cantwell, G. Reinert, and M. E. J. Newman, Efficient method for estimating the number of communities in a network. *Phys. Rev. E* **96**, 032310 (2017).

[20] T. P. Peixoto, Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X* **4**, 011047 (2014).

[21] D. J. Watts, *Small Worlds*. Princeton University Press, Princeton (1999).

[22] T. Chakraborty, S. Srinivasan, N. Ganguly, S. Bhowmick, and A. Mukherjee, Constant communities in complex networks. *Scientific Reports* **3**, 1825 (2013).

[23] H. Kim and S. H. Lee, Relational flexibility of network elements based on inconsistent community detection. *Phys. Rev. E* **100**, 022311 (2019).

[24] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*. Addison-Wesley, Reading, MA (1993).

[25] M. E. J. Newman, G. T. Cantwell, and J. G. Young, Improved mutual information measure for classification and community detection. Preprint arxiv:1907.12581 (2019).

[26] M. Meilă, Comparing clusterings: An information based distance, *Journal of Multivariate Analysis* **98**, 873–895 (2007).