

# Are communities just bottlenecks? Trees and treelike networks have high modularity

James P. Bagrow\*

*Department of Engineering Sciences and Applied Mathematics,  
Northwestern Institute on Complex Systems,  
Northwestern University, Evanston, Illinois, USA*

(Dated: January 4, 2012)

Much effort has gone into understanding the modular nature of complex networks. Communities, also known as clusters or modules, are densely interconnected groups of nodes that are only sparsely connected to other groups in the network. Discovering high quality communities is a difficult and important problem in a number of areas. The most popular approach is the objective function known as Modularity, used to both discover communities and measure their strength. To understand the modular structure of networks it is then crucial to know how such functions evaluate different topologies, what features they account for and what implicit assumptions they may make. We show that trees and treelike networks can have unexpectedly and often arbitrarily high values of modularity. This is surprising since trees are maximally sparse connected graphs and are not typically considered to possess modular structure, yet the non-local null model used by modularity assigns low probabilities, and thus high significance, to the densities of these sparse tree communities. We further study the practical performance of popular methods on model trees and on a genealogical dataset, and find that the discovered communities also have very high modularity, often approaching its maximum value.

PACS numbers: 89.75.Hc, 89.75.Fb, 05.10.a, 89.20.Hh

## I. INTRODUCTION

Complex networks have made an enormous impact on research in a number of disciplines [1–5]. Networks have revolutionized the study of social dynamics and human contact patterns [6–8], metabolic and protein interaction in the cell [9, 10], ecological food webs [11–13], and technological systems such as the world wide web [14, 15] and airline transportation networks [16, 17]. Seminal results include the small-world [18] and scale-free nature [14] of many real world systems.

One of the most important areas of network research has been the study of community structure [19, 20]. Communities, sometimes called modules, clusters, or groups, are subsets of nodes that are densely connected among themselves while being sparsely connected to the rest of the network. Networks containing such groups are said to possess modular structure. Understanding this structure is critical for a number of applications from link prediction [21] and the flow of information [22] to a better understanding of population geography [23–25].

Much effort has been focused on finding the best possible partitioning of a network into communities. Typically this is done by optimizing an objective function that measures the community structure of a given partition. Many algorithmic approaches have been devised. Most partition the entire network while some focus on local discovery of individual groups [26–28]. Overlapping community methods, where nodes may belong to more than one group, have recently attracted much interest [29–31]. For a lengthy review of community methods see [19].

Given the reliance on objective functions, it is important to understand how the intuitive notion of communities as in-

ternally dense, externally sparse groups is encoded in the objective function. Some functions simply measure the density of links within each community, ignoring the topological features those links may display, while other functions rely upon those links forming many loops or triangles, for example. We show the importance of understanding these distinctions by revealing some surprising features of how communities are evaluated. In particular we show that the only requirement for strong communities—according to the most popular community measure—is a lack of external connections, that bottlenecks [32] leading to isolated groups can make strong communities even when those groups are internally maximally sparse.

This paper is organized as follows. In Sec. II we present several measures of community quality and discuss their different features and purposes. In Sec. III we show analytically that trees and treelike graphs can possess partitions that display very high, often arbitrarily high values of modularity. This is our primary result. In Sec. IV we apply two successful community discovery algorithms to these trees and show that the discovered communities can have even higher modularities. We also study the community structure of a treelike network derived from genealogical data. We finish with a discussion and conclusions in Sec. V.

## II. MEASURING COMMUNITIES

Given a network—represented by a graph  $G$  of  $N$  nodes whose structure is encoded in an  $N \times N$  adjacency matrix  $A$  where  $A_{ij} = 1$  if nodes  $i$  and  $j$  are connected and zero otherwise—we wish to determine to what extent  $G$  possesses modular structure. To put the notion of a community or module onto a firm foundation, objective functions have been introduced to measure or quantify how “good” or “strong” a community or a partitioning into communities is. These ob-

---

\* james.bagrow@northwestern.edu; <http://bagrow.com>

jective functions are also often the goal of an optimization algorithm, where the algorithm attempts to find the community or communities that maximizes (or minimizes) the objective function. Here we briefly discuss three objective functions: subgraph conductance, modularity, and partition density. Due to its popularity and wide use we will focus primarily on modularity.

### A. Conductance

The conductance  $\phi$  of a subgraph is a measure of how ‘isolated’ the subgraph is, in analogy with electrical conductance [33]. Subgraphs with many connections to the rest of the network will have high conductance, whereas a subgraph will have low conductance if it relies on a few links for external connectivity. For a given subgraph  $S$  such that  $|S| \leq N$ , one form of conductance is

$$\phi(S) = \frac{\sum_{i,j} A_{ij} [i \in S] [j \notin S]}{\sum_{i,j} A_{ij} [i \in S] [j \in S]} = \frac{K_S - 2m_S}{2m_S}, \quad (1)$$

where  $[P] = 1$  if proposition  $P$  is true and zero otherwise,  $K_S = \sum_{i \in S} \sum_j A_{ij}$  is the sum of the degrees (number of neighbors) of all nodes in  $S$ , and  $m_S$  is the total number of links in  $S$ . (The factor of two in the denominator is sometimes dropped.) In other words, subgraph conductance is the ratio between the number of links exiting the subgraph to the number of links within the subgraph.

While low  $\phi$  may appear to be a good indicator of community structure, we remark that it primarily measures isolation or ‘bottleneckedness,’ meaning that, e.g., a random walker moving in a subgraph with low conductance will be very few opportunities to exit the subgraph, whereas it would have many opportunities if the subgraph had high conductance. This is also true if the subgraph is a densely interconnected module. However, consider a large 2D periodic lattice of size  $L_x \times L_y$ ,  $L_x \geq L_y$ . This graph has  $N = L_x L_y$  nodes and  $M = 2N$  links and is generally considered to have no modular structure. The conductance of a subgraph created by cutting the lattice in half along the  $y$  direction is  $\phi = 2L_y / (L_x L_y) = 2/L_x$ . As the lattice grows, the conductance of this subgraph decreases, despite there being no modular structure.

### B. Modularity

A key point lacking in earlier definitions of communities such as conductance is that they fail to quantify the statistical significance of the subgraph. It may be possible for a randomized null graph to contain subgraphs exhibiting comparable conductance, for example, and conductance alone does not capture this. Modularity [34, 35] was introduced to account for this in an elegant way. It has become the most common community objective function [19, 20] and possesses a number of distinct advantages over previous approaches, such as not requiring the number of communities to be known in advance. However, it has some drawbacks as well: it is known

to possess a *resolution limit* where it prefers communities of a certain size that depends only on the global size of the network and not on the intrinsic quality of those communities [36].

Modularity  $Q$  can be written as:

$$\begin{aligned} Q &= \frac{1}{2M} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2M} \right] \delta(c_i, c_j) \\ &= \sum_c \left[ \frac{m_c}{M} - \left( \frac{K_c}{2M} \right)^2 \right], \end{aligned} \quad (2)$$

where  $M = \frac{1}{2} \sum_{i,j} A_{ij}$  is the total number of links in the network;  $k_i = \sum_j A_{ij}$  is the degree or number of neighbors of node  $i$ ;  $\delta(c_i, c_j) = 1$  if node  $i$  and node  $j$  belong to the same community ( $c_i = c_j$ ) and zero otherwise;  $m_c = \frac{1}{2} \sum_{i,j} A_{ij} \delta(c_i, c) \delta(c_j, c)$  is the total number of links inside community  $c$ ; and  $K_c = \sum_i k_i \delta(c_i, c)$  is the total degree of all nodes in community  $c$ . The first definition of  $Q$  illustrates the intuition of its form: For every node pair that shares a community we sum the difference between whether or not that pair is actually linked with the expected ‘number’ of links between those same two nodes if the system was a purely random network constrained to the same degree sequence (this null model is known as the configuration model, and the loss term is approximate). This is then normalized by the total number of links in the network. By rewriting the sum over node pairs as a sum over the communities themselves, the second definition of  $Q$  makes clear the resolution limit: global changes to the total number of links  $M$  will disproportionately affect each community’s local contribution to  $Q$ . This can potentially shift the maximal value of  $Q$  to a different partition even when the local structure of the communities remains unchanged. It has also been shown that random graphs can have large modularities due to fluctuations [37].

Equation 2 gives values between  $-1$  and  $1$ . When  $Q \approx 0$  there is strong evidence that the discovered community structure is not significant, at least according to this null model, while the communities are considered better and more significant as  $Q$  grows. In practice, it is typically assumed that a network possesses modular structure when  $Q > 0.25$  or  $0.3$  [34].

### C. Partition density

Yet another approach to quantifying community structure is that of partition density [30]. Partition density was introduced specifically for the case of link communities, where links instead of nodes are partitioned into groups. This allows for communities to overlap, since nodes may belong to multiple groups simultaneously. We do not consider overlapping communities here, but partition density can still be calculated for non-overlapping node communities.

The partition density  $D$  is

$$D = \frac{1}{2M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 1)(n_c - 2)}. \quad (3)$$

Partition density measures, for each community, the number of links within that community minus the minimum number

of links necessary to keep a subgraph of that size connected,  $n_c - 1$ , the size of its spanning tree. This is then normalized by the maximum and minimum number of links possible for that connected subgraph,  $\binom{n_c}{2}$  and  $n_c - 1$ , respectively. The partition density is then the average of this quantity over the communities, weighted by the fraction of links within each community. For a link partition that covers an entire connected network, we have  $\sum_c m_c = M$ , but this does not necessarily hold for a node partition.

A crucial feature of the partition density is that it explicitly compares the link density of a subgraph to that of a tree of the corresponding size. This controls for the fact that the subgraph in question is connected, making the reasonable assumption that communities should be connected. The null model used by modularity on the other hand, does not make this assumption, and it may potentially assign very low probabilities to such an event. As we will show, this aspect of modularity is crucial and has been overlooked.

### III. COMMUNITIES IN TREES AND TREELIKE GRAPHS

We now study a model tree graph that is typically considered to not possess modular structure and show that these graphs possess partitions with arbitrarily high modularity values. We also study a mixed case graph containing both modular and non-modular structures.

#### A. Cayley tree

The Cayley tree is a regular graph with no loops and where every node  $i$  has the same degree  $k_i = z + 1$  (except for leaf nodes on the boundary who possess  $k = 1$ ). See Fig. 1. It can be constructed by first starting from a root node at generation 0, giving that node  $z + 1$  child nodes, and then repeatedly giving each new child  $z$  children of its own. This continues for a fixed number of generations  $g$ . These trees can grow either in “width” (via  $z$ ) or in “depth” (via  $g$ ). The number of nodes in generation  $g > 0$  is  $n(g) = (z + 1)z^{g-1}$  and the total number of nodes is  $N(g) = 1 + \sum_{g'=1}^g n(g')$ . Since this is a tree, the total number of links is  $M(g) = N(g) - 1 = (z + 1)(1 - z^g)/(1 - z)$ . Since the bulk of the graph is regular, the Cayley tree has no density fluctuations (except possibly considering the edge) and so it does not in an obvious way conform to our preconceived notions of community structure as internally dense, externally sparse groups. In the thermodynamic limit the Cayley tree is known as the Bethe lattice. We concern ourselves here primarily with finite graphs, however, such that finite size and edge effects cannot be ignored.

We now compute the modularity of a particular “community” partition of the Cayley tree, which we call the *analytic partition*. First place the root node into a community of its own. Then create a new community for each child of the root node, containing that child *and all of its descendants*. Thus there are  $z + 2$  communities in total. Apart from the singleton community containing the root node, every community is a complete  $z$ -ary tree (which is not exactly a Cayley tree) with

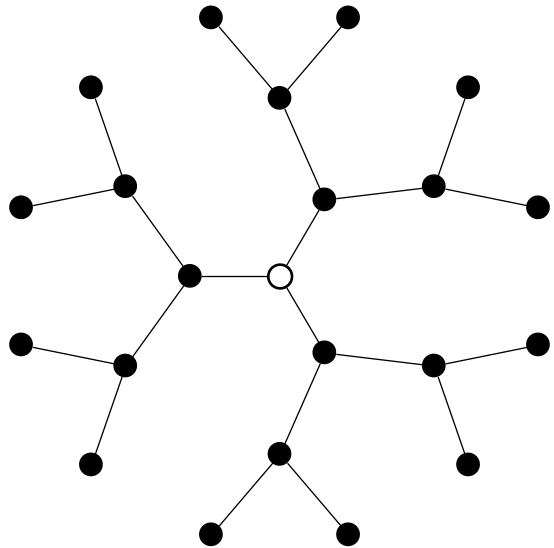


FIG. 1. Cayley tree for  $z = 2$  and  $g = 3$ . The root node is indicated in white.

$g - 1$  generations. Partitioning the tree in this way requires cutting only  $z + 1$  links. There are zero links inside the singleton community and

$$m = \frac{N(g) - 1}{z + 1} - 1 = z \frac{1 - z^{g-1}}{1 - z} \quad (4)$$

links inside the  $z + 1$  other communities. To compute the total degree of nodes within the community, we note that all  $(N(g) - 1)/(z + 1)$  nodes have degree  $z + 1$  except the  $n(g)/(z + 1)$  boundary nodes that have degree 1. Thus the total degree is

$$\begin{aligned} K &= (z + 1) \left( \frac{N(g) - 1}{z + 1} - \frac{n(g)}{z + 1} \right) + \frac{n(g)}{z + 1} \\ &= \frac{1 + z - 2z^g}{1 - z}. \end{aligned} \quad (5)$$

The final modularity is then given by substituting these expressions for  $m$ ,  $K$ , and  $M$  into:

$$Q_{\text{cayley}} = (z + 1) \left[ \frac{m}{M} - \left( \frac{K}{2M} \right)^2 \right] - \left( \frac{z + 1}{2M} \right)^2, \quad (6)$$

where the functional dependence on  $z$  and  $g$  has been suppressed. For  $z = 10$  and  $g = 4$ , for example,  $Q_{\text{cayley}} \approx 0.91$ , an extremely high modularity. Even for  $z = 3$  and  $g = 3$  we have a high modularity of  $Q_{\text{cayley}} \approx 0.7$ . In general, the limiting value of  $Q_{\text{cayley}}$  for a given  $z$  is

$$\lim_{g \rightarrow \infty} Q_{\text{cayley}}(z, g) = \frac{z}{z + 1}. \quad (7)$$

Even for a finite  $g > 1$ ,  $Q_{\text{cayley}} \rightarrow 1$  as  $z \rightarrow \infty$ . Thus the Cayley tree is able to achieve **arbitrarily high** modularity partitions.

Meanwhile, the  $z + 1$  branch communities of the Cayley tree’s analytic partition each have conductance

$$\phi_{\text{cayley}} = \frac{1}{m} = \frac{1 - z}{z - z^g}. \quad (8)$$

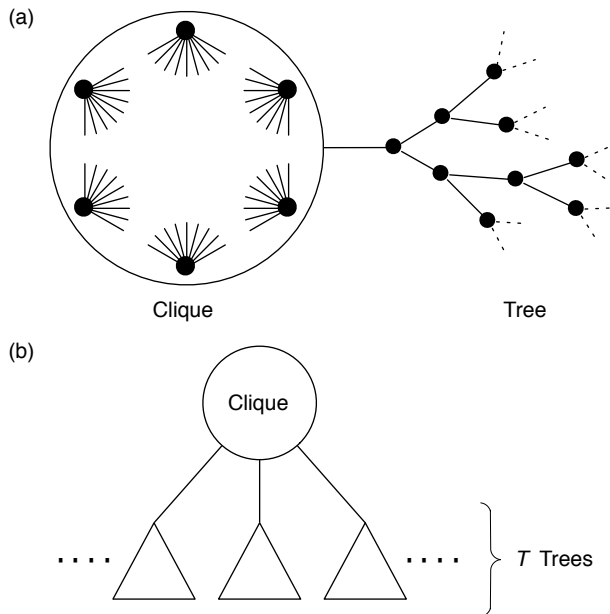


FIG. 2. (left) A mixed test case consisting of a single clique (complete subgraph) of  $n_{\text{clique}}$  nodes connected by a single link to a  $z$ -ary tree. This is partitioned into communities by cutting the single bridging link. (right) A generalization where now  $T$  trees are connected to the single clique ( $T < n_{\text{clique}}$ ).

For  $z = 4$  and  $g = 10$ , for example,  $\phi_{\text{cayley}} \approx 2.86 \times 10^{-6}$ , a very small value. This makes sense since only a single link separates that entire branch from the rest of the graph. This also emphasizes that conductance is primarily a measure of bottlenecks and isolation and should be approached with caution when applied to community structure.

Finally, we remark that the partition density of the Cayley tree is zero since  $m_c = n_c - 1$ . This is true not just for the analytic partition but for all partitions of the Cayley tree where each community is connected.

### B. A clique and a tree

In practice, one may deal with networks with wide fluctuations in local density, meaning there may exist localized subgraphs of low and of high density at the same time. We analyze a simple example consisting of a single completely dense graph known as a clique connected by one link to the root of a  $z$ -ary tree of  $g$  generations. See Fig. 2.

We wish to compute the modularity of a community partition containing the entire clique in one community and the entire tree in the other, where only the link from the root of the tree to the clique was cut. We assume that there are  $n_{\text{clique}}$  nodes in the clique and  $n_{\text{tree}} = (1 - z^{g+1}) / (1 - z)$  nodes in the tree. The numbers of links in each subgraph are  $m_{\text{clique}} = \binom{n_{\text{clique}}}{2}$  and  $m_{\text{tree}} = z(1 - z^g) / (1 - z)$ , respectively. The total number of links is  $M = m_{\text{clique}} + m_{\text{tree}} + 1$  and the total degrees are  $K_{\text{clique}} = n_{\text{clique}} + (n_{\text{clique}} - 1)^2$  and  $K_{\text{tree}} = z^g + (z + 1)(1 - z^g) / (1 - z)$ . The final modularity of

the partition is then given by substituting these expressions into

$$Q_{\text{clique-tree}} = \frac{m_{\text{clique}} + m_{\text{tree}}}{M} - \frac{K_{\text{clique}}^2 + K_{\text{tree}}^2}{4M^2}. \quad (9)$$

We plot Eq. (9) as a function of  $g$  in Fig. 3a for  $n_{\text{clique}} = 100$  and several values of  $z$ . We see that  $Q$  attains a maximum of  $1/2$ , a non-random value but not as high as the pure Cayley tree previously analyzed despite the addition of a “perfect” community. We also see that, as  $z$  increases,  $Q$  becomes more sharply peaked as a function of  $g$ . This is due to the resolution limit: the larger  $z$  is, the more quickly the tree will grow from one generation to the next and thus the tree community more quickly passes beyond the size preferred by modularity. This leads to a  $Q$  that grows more rapidly and then decays more rapidly as  $g$  increases.

We also study a generalization of Fig. 2a from one tree to  $T$  trees (Fig. 2b), where each tree is its own community. For this model  $m_{\text{tree}}$  and  $K_{\text{tree}}$  are unchanged for each tree, while now  $K_{\text{clique}} = T + n_{\text{clique}}(n_{\text{clique}} - 1)$ ,  $M = m_{\text{clique}} + Tm_{\text{tree}} + T$  and

$$Q_{\text{clique-trees}} = \frac{m_{\text{clique}} + Tm_{\text{tree}}}{M} - \frac{K_{\text{clique}}^2 + TK_{\text{tree}}^2}{4M^2}. \quad (10)$$

We plot Eq. (10) in Fig. 3b as a function of  $n_{\text{clique}}$  for several values of  $T$ . We see that increasing  $T$  raises the overall modularity of the partition, giving quite high values of  $Q_{\text{clique-trees}} > 0.8$ . We also see that, as  $T$  increases, the curve becomes more flat, meaning that good quality partitions, according to modularity, exist for a wide range of clique sizes.

## IV. REAL-WORLD EXAMPLES

The above derivations show that trees may possess arbitrarily high values of modularity. However, these calculations did not consider the resolution limit of modularity. In fact, real world optimization of modularity will result in partitions that give *even higher* values of  $Q$  than those of the analytic partitions discussed in Sec. III.

To see this we apply two of the most popular and successful community discovery methods. The first is known as **Fast Unfolding** (sometimes referred to as the Louvain method) and can efficiently find very high modularity partitions [38]. The second method is called **Infomap** [39]. Infomap does not optimize modularity, instead exploiting information-theoretic arguments, but the partitions it does find are often high in modularity, especially for undirected networks.

We apply these algorithms to the Cayley tree. In Fig. 4 we plot the modularities discovered by each algorithm and the modularities  $Q_{\text{cayley}}$  of the analytic partitions (Eq. 6). We see that the methods find communities that appear as strong as the analytic method or stronger. Fast Unfolding typically exceeds  $Q_{\text{cayley}}$  as the trees grow, and even approaches  $Q = 1$ . Infomap tends to stay closer to  $Q_{\text{cayley}}$  but it too can exceed these bounds, especially for trees with  $z = 2$ . If these methods were applied blindly to a network, such high values of modularity would suggest that these communities are extremely high quality and that the network was extremely modular.

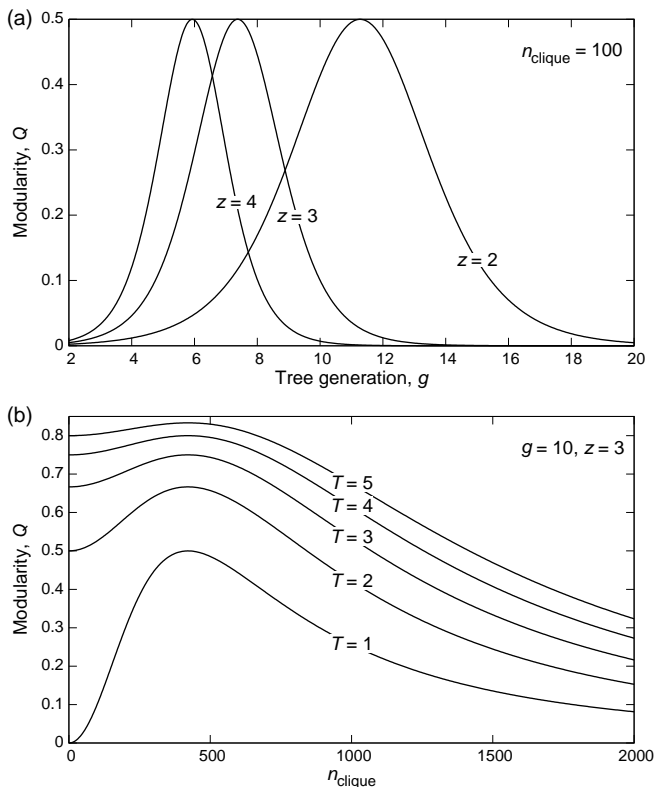


FIG. 3. (a) Modularity for the example network illustrated in Fig. 2a. As we increase the size of the tree for a fixed clique size, the modularity grows to a maximum value and then decays away. This is due to the resolution limit: there exists a specific tree size that maximizes  $Q$  for each clique size. (b) For the generalization of one clique and  $T$  trees, shown in Fig. 2b, we see that the analytic partition again attains high modularity, especially as more trees are added. Likewise, as  $T$  increases we see that the peak of  $Q$  flattens out and that the partition has high modularity for a range of clique sizes. This means that much of the resolution limit can be compensated for if the network is sufficiently treelike.

As a final practical example, we also apply both methods to a treelike network derived from a genealogical dataset capturing the advisor-advisee relationships between mathematicians and their students [40, 41]. (This genealogy is not exactly a tree since some students have multiple advisors.) We only consider the giant connected component of the network, capturing approximately 90% of the dataset. In total the network has  $N = 133319$  nodes and  $M = 148247$  links. The modularities of the partitions found by Fast Unfolding and Infomap are  $Q_{FU} = 0.951083$  and  $Q_{IM} = 0.877146$ , respectively. These high values would again imply that the network is strongly modular.

Another interesting aspect of a community partition is the distribution of community sizes (numbers of nodes per community). Since any discovered modular network structure depends intrinsically on the definition at the heart of the algorithm used to find that structure, it is not known for certain what the true distribution is. Nevertheless, there has been empirical evidence showing that the size distribution may exhibit

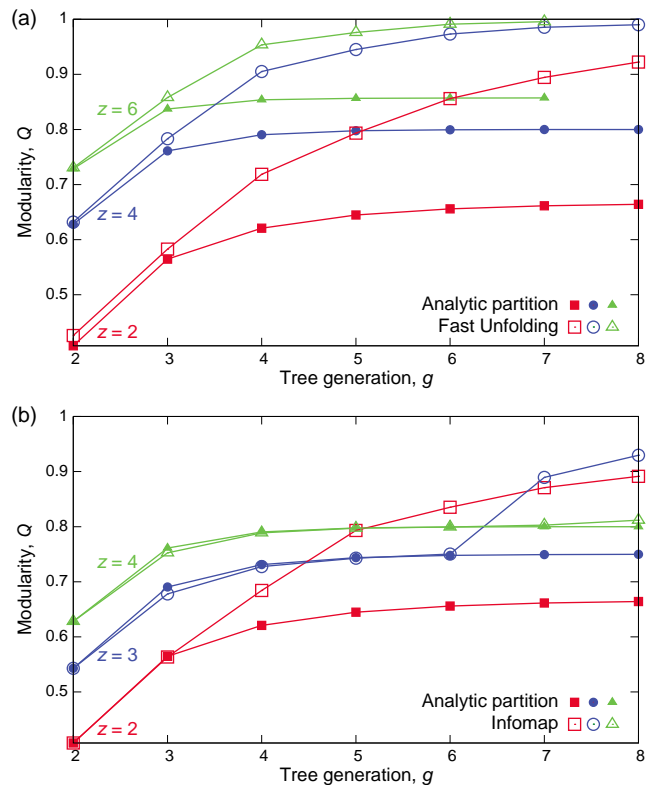


FIG. 4. (Color online) Community discovery methods find even higher values of modularity than the analytical partition of the Cayley tree. We apply two methods: (a) Fast Unfolding [38] and (b) Infomap [39] for several values of  $z$ . Closed symbols correspond to Eq. (6) while open symbols correspond to the modularities found by the methods. Fast Unfolding finds consistently higher modularities than the analytic partition, due to the resolution limit. These partitions even approach  $Q = 1$ . Infomap, which does not optimize modularity, tends to find partitions comparable to the analytic partition, although it too finds higher value partitions for  $z = 2$  and some values of  $g$  for  $z = 3$ . Note that the vertical axes do not begin at  $Q = 0$ .

a power law  $Pr(s) \sim s^{-\alpha}$ , for  $\alpha \geq 1$  [29, 30].

Yet the distributions of community sizes found in the genealogical network, shown in Fig. 5, are not heavy-tailed. Instead both methods find approximately exponential distributions, with a small number of larger communities that would be underrepresented by an exponential distribution. The lack of very large communities may be expected in graphs without hubs, but the degree distribution for this network is heavy-tailed (Fig. 5b, inset). Further study of the size distribution of community structure may prove fruitful in understanding the modular nature of complex systems.

## V. DISCUSSION AND CONCLUSIONS

We have shown that trees appear very modular. Yet connected trees are maximally sparse, going against the tenant that communities are unusually dense subgraphs. Even when

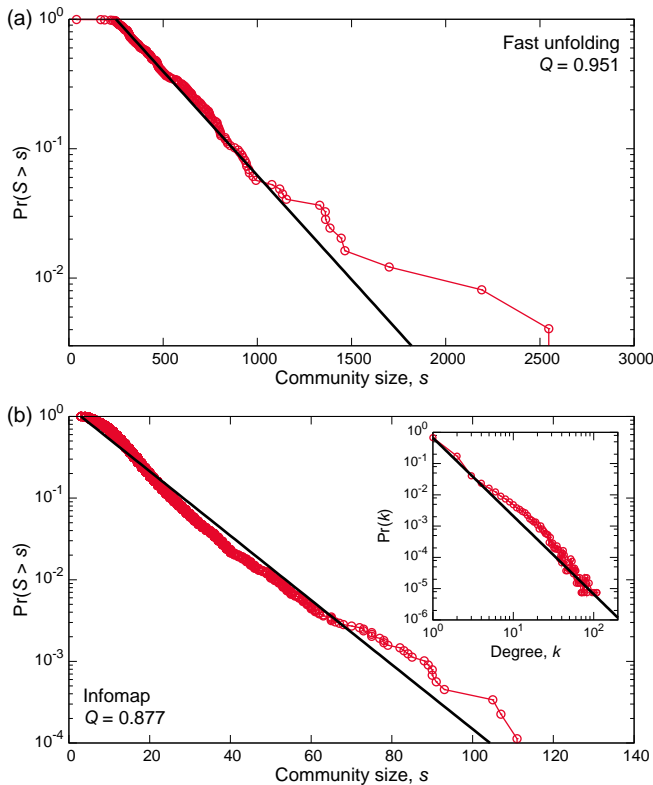


FIG. 5. The distribution of community sizes found in the genealogical network for (a) Fast Unfolding and (b) Infomap. We see that neither distribution is heavy-tailed, being instead approximately exponential (straight lines) except for a small number of the largest communities that would be underestimated by an exponential distribution. (inset) Unlike the community size distribution, the degree distribution of the network is heavy-tailed. The straight line shows a pure power law,  $\text{Pr}(k) \sim k^{-2.5}$ , for comparison.

combined with non-tree structures, the presence of trees can boost modularity. Since trees are the limiting structure as networks become sparse, sampled and missing data [42] may boost modularity, at least in some regions of networks. Incomplete data remains an issue in high-throughput biological assays, for example [43].

So is our definition of modular structure correct? Equation 2 depends so strongly on its null model that we must judiciously understand all facets of it. Both the resolution limit and the results shown here are due to the global nature of modularity's null model. Better approaches may be needed yet defining appropriate null models is difficult and remains an open question.

Partition density does not classify trees as modular because it *explicitly* requires groups to be denser than their corresponding spanning trees. It may prove fruitful to consider null models that incorporate or control for the fact that the communities will most likely form connected subgraphs. A local version of the configuration model, conditioned on the existence of a spanning tree in the group, may be a viable option. This approach may prove computationally intractable in practice, however, as analytic expressions for such a graph ensemble may not be possible, limiting us to monte carlo methods.

## ACKNOWLEDGMENTS

We thank Gino Biondini for providing the genealogical data.

- 
- [1] M. E. J. Newman, *Networks: an introduction* (Oxford University Press, 2010).
- [2] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47. 54 p (2001).
- [3] M. Newman, *SIAM Rev.* **45**, 167 (2003).
- [4] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical processes on complex networks* (Cambridge University Press, 2008).
- [5] A. Vespignani, *Nature Physics* **8**, 32 (2011).
- [6] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. d. Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, *New J. Phys.* **9**, 179 (2007).
- [7] M. C. González, C. A. Hidalgo, and A.-L. Barabási, *Nature* **453**, 779 (2008).
- [8] J. P. Bagrow, D. Wang, and A.-L. Barabási, *PLoS ONE* **6**, e17680 (2011).
- [9] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. Barabási, *Nature* **407**, 651 (2000).
- [10] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [11] J. Dunne, R. Williams, and N. Martinez, *Ecology Letters* **5**, 558 (2002).
- [12] A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, *Nature* **426**, 282 (2003).
- [13] L. S. Schulman, J. P. Bagrow, and B. Gaveau, *Advs. Compl. Syst.* **14**, 829 (2011).
- [14] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [15] J. Kleinberg, *Nature* **406**, 845 (2000).
- [16] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2015 (2006).
- [17] D. Brockmann, L. Hufnagel, and T. Geisel, *Nature* **439**, 462 (2006).
- [18] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [19] S. Fortunato, *Physics Reports* **486**, 75 (2010).
- [20] M. E. J. Newman, *Nature Physics* **8**, 25 (2011).
- [21] A. Clauset, C. Moore, and M. E. J. Newman, *Nature* **453**, 98 (2008).
- [22] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332 (2007).
- [23] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz, *PLoS ONE* **5**,

- e14248 (2010).
- [24] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, Proc. Natl. Acad. Sci. U.S.A. **108**, 7663 (2011).
- [25] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Brockmann, PLoS ONE **5**, e15422 (2010).
- [26] J. P. Bagrow and E. M. Bollt, Phys. Rev. E **72**, 046108 (2005).
- [27] A. Clauset, Phys. Rev. E **72**, 026132 (2005).
- [28] J. P. Bagrow, J. Stat. Mech. **05**, 001 (2008).
- [29] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature **435**, 814 (2005).
- [30] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, Nature **466**, 761 (2010).
- [31] B. Ball, B. Karrer, and M. E. J. Newman, Phys. Rev. E **84**, 36103 (2011).
- [32] S. Sreenivasan, R. Cohen, E. López, Z. Toroczkai, and H. Stanley, Phys. Rev. E **75**, 036105 (2007).
- [33] B. Bollobás, *Modern graph theory* (Springer Verlag, 1998).
- [34] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, (2004).
- [35] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).
- [36] S. Fortunato and M. Barthelemy, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).
- [37] R. Guimera, M. Sales-Pardo, and L. Amaral, Phys. Rev. E **70**, 025101 (2004).
- [38] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech. **2008**, P10008 (2008).
- [39] M. Rosvall and C. T. Bergstrom, Proc. Natl. Acad. Sci. U.S.A. **105**, 1118 (2008).
- [40] North Dakota State University, “The mathematics genealogy project,” <http://genealogy.math.ndsu.nodak.edu>.
- [41] R. D. Malmgren, J. M. Ottino, and L. A. N. Amaral, Nature **465**, 622 (2010).
- [42] J. P. Bagrow, S. Lehmann, and Y.-Y. Ahn, Arxiv preprint arXiv:1102.5085 (2011).
- [43] H. Yu, P. Braun, M. Yildrm, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, and N. Simonis, Science **322**, 104 (2008).