Hierarchical Problems for Community Detection in Complex Networks

Benjamin Good[1, 2, *] and Aaron Clauset[2, †]

[1] *Physics Department, Swarthmore College, 500 College Ave., Swarthmore PA, 19081*

[2] *Santa Fe Institute, 1988 Hyde Park Rd. Santa Fe NM, 87501*

(Dated: August 15, 2008)

An objective method for extracting network community structure is an extremely useful tool for understanding the large complex networks found in the social and biological sciences. One such method, which relies on the maximization of the modularity quality function Q, has received a great deal of attention and is now widely used. We find that, for networks with a hierarchical modular structure, there exist many different community assignments with comparably high modularity Q that mix different levels and modules in the hierarchy. Since many real-world networks are thought to possess hierarchical structure, this throws doubt upon the ability of any single-layer community detection algorithm to reliably uncover the underlying group structure and calls into question the significance of individual modularity maximization results.

I. INTRODUCTION

The search for an objective method of network community detection has been fueled by the recent explosion of available data in the biological and social sciences, which often deal with large complex systems containing many interacting parts. A reliable method of community detection serves the useful purpose of allowing us to "coarse-grain" our view of the system, hopefully reducing it to a more easily understandable form. In addition, it is often hypothesized that the topological structure of a network is closely tied to the underlying function of the system [1], so a community detection algorithm might allow us to identify functional modules for further study.

A useful and popular method of community detection has been proposed by Girvan and Newman [2]. They introduce the modularity function Q, which gives a measure of how "good" a particular community assignment is. This function is then maximized over the space of all community assignments to determine the best partition of the network into communities.

It has been shown that the problem of maximizing Q is NP-Complete [3], meaning that in practice, one must turn to approximate optimization algorithms, several of which have been previously discussed in the literature [4–10]. However, many of these approximation algorithms return a single community assignment, leading us to ask whether this approximate result accurately reflects the underlying group structure. Such a property is crucial if we wish to make further inferences based on the community analysis.

A seemingly unrelated weakness in the modularity maximization method is its focus on returning a "flat" community structure, even when the network has a clear hierarchical organization. To collapse the hierarchy into this flat community structure, the modularity function must make a choice as to the "best" level of resolution of the hierarchy, leading to a loss of information about the other levels. Since many real-world networks are assumed to contain an underlying hierarchical structure [11], this weakness could have far-reaching consequences.

Various methods have been proposed to get around this issue. Sales-Pardo *et. al.* use an ensemble of local modularity maxima to construct a picture of the underlying hierarchical structure [12]. Reichardt and Bornholdt introduce a small parameter which can be tuned to uncover different levels of resolution [9]. Yet these methods all rely on the fact that the modularity function is eventually able to explore every level of the hierarchical structure. This assumption has never been demonstrated in a rigorous manner and is brought into question by recent work showing that maximizing modularity can sometimes fail to uncover small communities in large networks [13].

In the current work, we find that these two seemingly unrelated issues – the significance of individual approximation results and the problem of hierarchy – are inextricably tied together. Using several simple analytical models, we see that the presence of a hierarchical community structure leads to a large number of community assignments with high modularity that mix different levels of the hierarchy. We then test this hypothesis by exploring the modularity landscape of several real-world networks thought to contain a hierarchical structure.

II. MODULARITY AND APPROXIMATION SIGNIFICANCE

We begin by outlining a quantitative definition of "significance" for results of approximation algorithms as well as the necessary tools required to measure this significance.

The modularity of a partition of a network into m communities is given by

$$Q = \sum_{s=1}^{m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right] \qquad (1)$$

where l_s is the number of edges in community s, d_s is

[*] Electronic address: bgood1@swarthmore.edu
[†] Electronic address: aclauset@santafe.edu

the total degree of the nodes in the community, and L is the total number of edges in the network. Intuitively, one can view the modularity as the difference between the observed fraction of edges within a community and the "expected" fraction of edges for the null-model of a random graph obeying the same degree distribution.

If we assume that the partition with maximum modularity corresponds to some "true" underlying community structure of the network, then the question of the significance of individual approximation results can be rephrased as follows: are partitions with a high modularity Q essentially small variations on the optimal community assignment or do they contain large deviations from this community structure?

In order to fully answer this question, we must first specify the nature of the space of partitions of a network. For a given partition C, we define its *neighbors* as the partitions generated by moving a single node into either one of the existing groups in the partition or to a new group on its own. This space can be difficult to analyze because the number of neighbors depends on the number of groups in the current partition. Yet we can still gain a notion of "distance" between two partitions by employing the Variation in Information (VI) metric introduced by Meila [14]. Given two partitions C and C' (with m and m' groups, respectively), the variation in information between them is given by

$$VI(C, C') = -\sum_{x=1}^{m}\sum_{y=1}^{m'} P_{xy} \log\left(\frac{P_{xy}^2}{P_x P_y}\right) \quad (2)$$

where P_x (respectively P_y) is the probability that a node is in group x (y) in partition C (C') and P_{xy} is the probability that a node is in group x in partition C *and* group y in partition C'. Two partitions of the network are the same if and only if $VI(C, C') = 0$ and the maximum possible VI is given by $\log n$ where n is the number of nodes in the network.

With our partition space and distance metric in hand, we now turn back to the question of the significance of approximation results. If a network's modularity landscape has only one competitive peak (the global optimum), then we can be confident in an approximation result with a high modularity because it is guaranteed to be located on this central peak somewhere close to the optimal partition. However, if the modularity landscape is littered with competitive local optima representing different community structures, then a high modularity partition does not necessarily have to be close to the global optimum because it could be located on one of the local optima. Thus, we see that the question of approximation significance can be rephrased as a question about the nature of the modularity landscape.

Previous work in this area used a simulation method called parallel tempering to determine the nature of the modularity landscape [15], and it was generally concluded that for real-world networks, there only exists one competitive peak in the modularity landscape [16]. However,

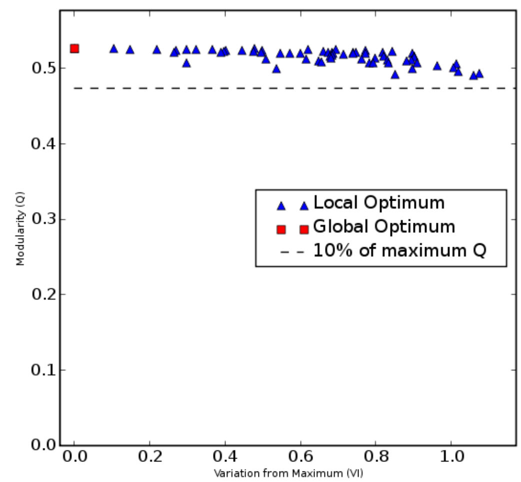

FIG. 1: Local optima sampled from the *American Political Books* network using our simulated annealing algorithm.

this conclusion suffers from the fact that it is difficult to determine whether there is truly a single peak or whether the parallel tempering algorithm becomes trapped in one of the local optima. To correct for this issue, we use a simulated annealing (SA) algorithm with a random initial partition and the COUPON COLLECTOR termination criterion to ensure that when the algorithm finishes, there is a high probability that no single move can increase the modularity of the partition. In other words, with high probability the algorithm terminates when it reaches a local optimum. These conditions help ensure that after running the SA algorithm many times, we sample a wide variety of the local optima (if they exist).

We tested our procedure on the *American Political Books* network compiled by V. Krebs. In this network, nodes represent different U.S. political books available online at amazon.com, and the nodes are connected by an edge if the two books are frequently purchased together. We ran our simulated annealing algorithm 700 times to produce 64 unique local optima, which were then plotted by their modularity Q and their variation in information from the previously calculated [17] global optimum (see FIG. 1).

As our results show, there are many local optima within 10% of the maximum modularity and most of these have non-trivial distances from the optimal partition. To get a sense the extent to which these local optima differ from the optimal partition, we can examine whether they constitute changes to the largest groups in the partition or whether they are just modifications of the smaller groups. We transform each partition by keeping the largest k groups and lumping the remaining nodes into a single group. If the variation in information is still high for these more "coarse-grained" partitions, then we have evidence that these local optima contain modifications to the largest group structures in the net-

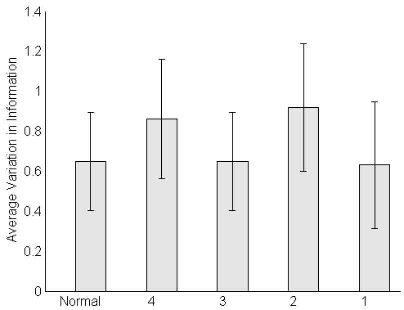

FIG. 2: The average VI for the *American Political Books* network samples when only the top k groups are retained.

work. We conducted these measurements for the political books network using values of k between 1 and 4 (the optimal partition contained only 5 groups) and the results are shown in FIG. 2. Since the average VI does not decrease with decreasing k, we conclude that most of the local optima contain significant structural deviations from the optimal community assignment.

This result represents a significant departure from the view commonly held in the literature that partitions with high modularity Q must be small variations on the optimal partition. However, even with this information, the exact nature of the group structure of these local optima is not immediately clear from the graph above, and short of visually examining every sampled partition, there does not seem to be a clear way to characterize the nature of these local optima. In addition, there is the question of why there are so many local optima in the first place? In what other kinds of networks can we expect to find the same behavior?

III. ANALYTICAL MODELS

To answer these questions, we consider several simple analytical networks whose modularity landscapes are much more amenable to direct calculation. We try to demonstrate that it is the hierarchical nature of the group structure that generates this explosion in local optima.

A. The Ring Network

We begin by examining the ring network, in which n cliques each containing m nodes are connected by single edges to form a ring structure (see FIG. 3). This model was chosen for its extreme analytical simplicity as well as the fact that it has a clear, intuitive group structure. Initially, one might expect that the optimal partition corresponds to putting each clique into its own group. The modularity of such a partition can be easily calculated to be

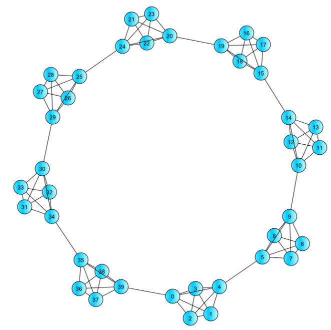

FIG. 3: An example ring network with $n = 8$ and $m = 5$.

$$Q_1 = 1 - \frac{1}{\binom{m}{2} + 1} - \frac{1}{n} \tag{3}$$

The partition generated by placing adjacent *pairs* of cliques in the same group is intuitively non-optimal — a certain aspect of the network's community structure is missed when viewed at this "lower resolution." Nevertheless, the modularity of this partition can be calculated to be

$$Q_2 = 1 - \frac{1}{2\left(\binom{m}{2} + 1\right)} - \frac{2}{n} \tag{4}$$

and we actually have $Q_2 > Q_1$ whenever n and m satisfy the inequality

$$n > 2\binom{m}{2} + 2 \tag{5}$$

Fortunato and Barthélemy first termed this phenomenon the *resolution limit* due to the inability of the modularity maximization to uncover group structure below a certain level [13].

This argument can easily be generalized to groups containing k cliques and again we find a similar resolution limit effect. By analogy with the above derivation, we can show that $Q_{k+1} > Q_k$ when n and m satisfy

$$n > k(k+1)\left[\binom{m}{2} + 1\right] \tag{6}$$

In this way, the resolution limit induces a hierarchical structure on the ring network, although since there are many different ways to group k adjacent cliques together, the resolution limit actually induces many different hierarchical structures.

Returning to the $k = 2$ case, we see that a partition that pairs non-adjacent cliques together is clearly non-optimal (see FIG. 4). In fact, for a partition containing β non-adjacent pairings, the modularity can be calculated to be

$$Q_2' \;=\; Q_2 - \frac{\beta}{n}\frac{1}{\binom{m}{2} + 1} \tag{7}$$

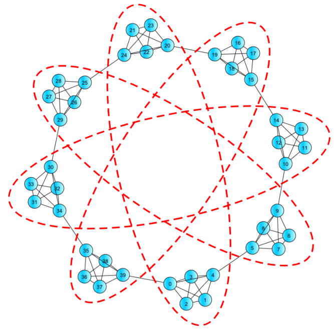

FIG. 4: A partition of the ring network that groups non-adjacent cliques together.

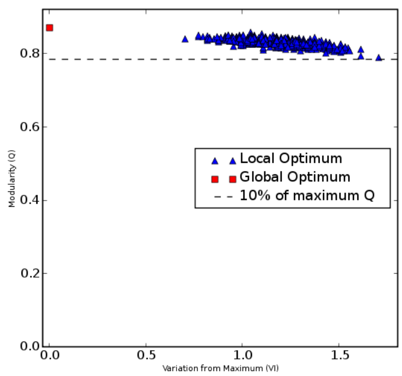

FIG. 5: Local optima sampled from the ring network with $n = 24$ and $m = 5$.

Yet it is also easy to see that this partition is a local optimum: moving any single node will break up one of the cliques and thus reduce the overall modularity. Furthermore, for a ring network with $n = 24$ and $m = 5$, there are more than 10^{10} such partitions within 10% of Q_2. With numbers like these, any algorithm that searches for the global optimum by sampling the local optima (like the SA algorithm we use) has a very poor chance of success.

Indeed, after running our SA algorithm on the ring network and plotting the results in the same manner as before, we see that our suspicions are confirmed (see FIG. 5). The algorithm samples many local optima with comparably high modularity and non-trivial VI and fails to sample the global optimum even once. By visually and numerically analyzing several of these local optima, we do not find a single case where these local optima ever break up a clique. However, they also very rarely group adjacent cliques together. In addition to the non-adjacent clique groupings we discussed above, these local optima also contain groups of 3 (and sometimes 4) cliques, thus mixing groups from different levels of the induced hierarchy.

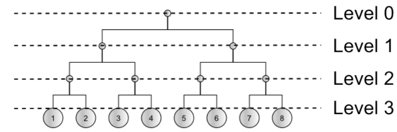

FIG. 6: An example of our specialized HRG with 8 nodes and 4 levels.

B. Hierarchical Random Graphs

A slightly more realistic model of a hierarchical network is the Hierarchical Random Graph (HRG) introduced by Clauset, Newman, and Moore [18]. In particular, we consider a special case of HRG in which there are 2^n nodes organized into a hierarchical group structure consisting of a balanced binary tree with $n+1$ levels (see FIG. 6) To implement this hierarchical structure, the nodes are connected in the following manner: nodes which share a common ancestor at level i of the tree are connected to each other with probability p_i. In general, the p_i can be any decreasing function of i, but for mathematical convenience we take

$$p_i = 2^{-(n-i-1)} \tag{8}$$

In order to prevent unusual edge assignments of individual HRG instances from affecting our results, we create an ensemble of HRGs and consider a partition's *average modularity* $\langle Q \rangle$, defined by

$$\langle Q \rangle = \left\langle \sum_{s=1}^{m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right] \right\rangle \tag{9}$$

We now make a mean field approximation (later confirmed by experiment) and take

$$\langle Q \rangle \approx \sum_{s=1}^{m} \left[\frac{\langle l_s \rangle}{\langle L \rangle} - \left(\frac{\langle d_s \rangle}{2\langle L \rangle} \right)^2 \right] \tag{10}$$

Because of the symmetry of the binary tree, the maximum modularity partition must consist of groups of the same size. Thus, to maximize the modularity, we must merely find a level k which maximizes the average modularity

$$\langle Q \rangle = 1 - \frac{k}{n} - 2^{-k} \tag{11}$$

If we treat k as a continuous variable, we see that $\langle Q \rangle$ is maximized when

$$k^* = \log_2(n \ln 2) \tag{12}$$

so that the optimal partition is the one with groups taken from the level of the hierarchy that is closest to k^*.

Again, we see that the HRG suffers from a resolution limit similar to the ring network. The optimal level grows

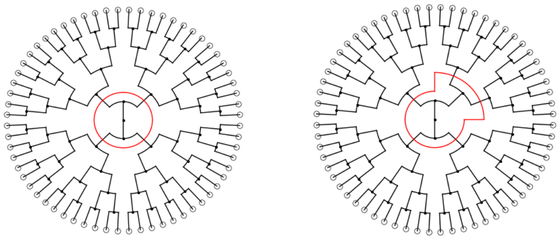

FIG. 7: The global optimum (left) and an example local optimum (right) for an HRG with 7 levels and 64 nodes.

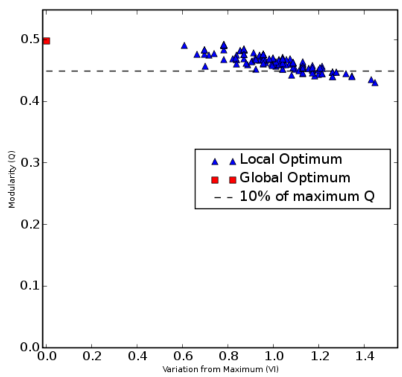

FIG. 8: Local optima sampled from the HRG network with 9 levels and 256 nodes.

with, but not as fast as n, so that the distance between the bottom level and the optimum level grows with the size of the network. Thus, for large enough networks, previously significant group structures become mere building blocks for communities rather than communities themselves. By analogy to the cliques of the ring network, we might expect that the local optima for the HRG mix groups from various levels of the hierarchy and/or join branches from different parts of the tree together (see FIG. 7). In most of these cases, moving a single node will break up one of the tightly bound groups at the bottom of the hierarchy and decrease the modularity.

This hypothesis was confirmed after running our SA algorithm on an HRG with 9 levels and 256 nodes (see FIG. 8). As with the previous two networks, we see a large number of local optima within 10% of the maximum modularity. A visual and numerical analysis confirms our suspicions: the local optima correspond to mixing different levels of the hierarchy. We found no cases where the branches were broken except when they were split into sub-branches, and even in these cases no branch of size 8 or smaller was ever split. Thus, we see that even without the peculiarities of the ring network, a hierarchical structure by itself can generate this explosion in local optima. In addition, for these kinds of networks we now have a better idea of the nature of these local optima, which simply correspond to a mixture of different levels of the hierarchy.

C. Non-Hierarchical Graphs

To determine whether a hierarchical structure is a *necessary* condition for the appearance of the local optima, we now turn to a simple non-hierarchical graph which contains 4 groups of 32 nodes each. Nodes within a group are connected with probability P_{in} and nodes spanning two groups are connected with probability P_{out}, where $P_{in} > P_{out}$. We choose P_{in} and P_{out} so that there are on average 768 links within groups and 256 links between groups. Hence by construction, there is no explicit hierarchical community structure and because of the small number of groups the resolution limit does not come into effect, so there is no induced hierarchy either. This network is of interest not only for its simple non-hierarchical structure, but also because it is a commonly used test case for approximate modularity maximization methods.

Using this network, we ran our SA algorithm 300 times and found the global maximum in every single instance. This result is important for several reasons. First, it makes more plausible our hypothesis that a hierarchical structure is the cause of the local optima. Second, it provides evidence that there is nothing peculiar about our algorithm that causes it to find these local optima. But finally, it serves as a lesson that this traditional test of modularity maximization methods might be too simple. Although it contains roughly the same number of nodes as the *American Political Books* network, its structure is obviously much more simple because it fails to generate the complex modularity landscape that we see for the latter.

IV. BACK TO THE REAL WORLD

With this possible cause of the local optima in mind, we return to the results for the *American Political Books* network shown in FIG. 1. Although we might naturally think these books divide into traditional political categories like left, right, and center, it is also fairly plausible that there exists a more hierarchical community structure, with groups like far-left, center-right, etc. In FIG. 9 and FIG. 10 we show a visual depiction of the global optimum and two sampled local optima from the experiment. We can see that for the most part, these local optima represent modifications to the global optimum in the sense that they merge and split groups in the optimal partition. This seems to agree with our earlier observations that the local optima mix of different levels of the hierarchical structure.

V. CONCLUSION

These analytical and numerical results show that when coupled with the resolution limit, a hierarchical structure creates a modularity landscape littered with competitive local optima, as significant community structures

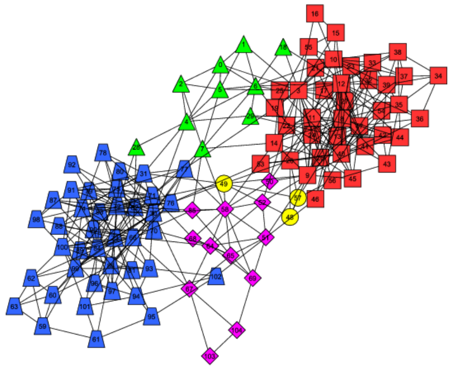

FIG. 9: The global optimum from the *American Political Books* network.

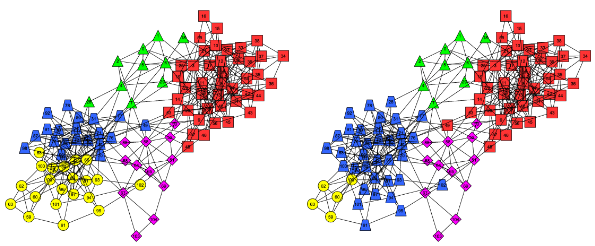

FIG. 10: Two particular local optima from the *American Political Books* network.

at one level become community building blocks at another level. This makes it very difficult, if not impossible to deduce the "natural" community structure using modularity maximization methods. One would not want to make further inferences based on a partition that mixes different levels of the hierarchy (possibly even mixing subgroups from different branches) because it yields only indirect information about the underlying hierarchical structure. A local optimum like the one for the ring network pictured in FIG. 4 would be even worse, because it gives very misleading information about which cliques are related. Furthermore, one would expect that it becomes even harder to visually discern hierarchical mixtures from particular levels of the hierarchy when the hierarchical structure becomes more asymmetric and more complex – something we might expect to occur in real world networks. Because of these difficulties, the significance of individual approximate modularity maximizations is called into question.

All hope is not lost, however. Clauset *et. al.* have introduced a hierarchy detection method that uses HRGs and maximum likelihood methods to infer the hierarchical structure without using modularity [18]. More traditional modularity approximation methods may still be useful if they can be shown to approximate the global maximum in all cases, yet it must be stressed that it is no longer sufficient to show that they merely return high modularity partitions most of the time. Judging by the good results achieved using approximations such as the spectral method [10], it is our suspicion that some approximations *are* guaranteed to approximate the correct peak.

In addition, one could consider using techniques that use information from a large sample of local optima, such as the one developed by Sales-Pardo, *et. al.* to uncover the hierarchical structure. Sales-Pardo et al. showed experimentally that such a technique can uncover hierarchical group structure when no single result will do, but it is only now that we are able to understand why such a technique works in the first place. Each local optimum cuts through the hierarchy in a certain way, conveying only incomplete information about the hierarchical structure. Yet when considered in aggregate, enough local optima should yield a much more complete picture of the underlying hierarchical structure. However, it must be stressed that even this technique has its weaknesses. From preliminary testing using our HRGs, we find that there is both an upper and lower limit to the levels of hierarchy that can be uncovered using this method. The limits of this method have not yet been rigorously tested using a large, known hierarchical structure, and this could be an area for future work.

Yet perhaps the most important message to be gleaned from the current work is its warning against using community finding algorithms designed to find a "flat" community structure even when the network is hierarchically organized. By doing so, we are fitting the wrong model to the data and can expect poor results from doing so. This makes it likely that any quality function similar to modularity would generate similarly complicated landscapes for a hierarchical network unless it was specifically designed to find a particular level in the hierarchy, which would be quite hard to do in general. Since hierarchy is a common property for systems in the real world, we stress the need for community detection methods that incorporate hierarchy as a central assumption rather than as a useful afterthought.

Acknowledgments

This work was funded in part by the National Science Foundation's Research Experience for Undergraduates Program.

[1] Need a citation here!
[2] Newman MEJ, Girvan M (2004) *Phys Rev E* 69:026113.
[3] U. Brandes, D. Delling, M. Gaertler, R. G orke, M. Hoefer, Z. Nikoloski, and D. Wagner, in *Proceedings of the 33rd International Workshop on Graph-Theoretic Concepts in Computer Science*, Vol. 4769 of *Lecture Notes in*

Computer Science, edited by A. Brandst adt, D. Kratsch, and H. M uller (Springer, Berlin, 2007).
[4] Newman MEJ (2004) *Phys Rev E* 69:066133.
[5] Clauset A, Newman MEJ, Moore C (2004) *Phys Rev E* 70:066111.
[6] Duch J, Arenas A (2005) *Phys Rev E* 72:027104.
[7] Guimerà R, Amaral LAN (2005) *Nature* 433:895.
[8] Medus A, Acunña G, Dorso CO (2005) *Physica A* 358:593.
[9] Reichardt J, Bornholdt (2006) *Phys Rev E* 74:016110.
[10] Newman MEJ (2006) *Proc Nat Acad Sci USA* 103:8677.
[11] Need a citation here!
[12] Sales-Pardo M, Guimerà R, Moreira AA, Amaral LAN (2007) *Proc Nat Acad Sci USA* 104:15224-15229.
[13] Fortunato S, Barthelemy M (2007) *Proc Nat Acad Sci USA* 104:36-41.
[14] M. Meila, in *ICML '05: Proceedings of the 22nd International Conference on Machine Learning* (ACM Press, New York, 2005), pp. 577-584.
[15] C.P. Massen and J.P.K. Doye, e-print arXiv:cond-mat/0610077v1.
[16] Karrer B, Levina E, Newman MEJ (2008) *Phys Rev E* 77:046119.
[17] G. Agarwal, D. Kempe, e-print arXiv:0710.2533v1
[18] A. Clauset, M.E.J. Newman, and C. Moore, in *Proceedings of the 23rd International Conference on Machine Learning* (Association of Computing Machinery, New York, 2006.