Identifying Community Structures from Network Data via Maximum Likelihood Methods

Jernej Čopič, Matthew O. Jackson, and Alan Kirman *

First Draft: August, 2005
This Draft: September 3, 2008

Abstract

Networks of social and economic interactions are often influenced by unobserved structures among the nodes. We develop a simple model of how hidden community structure generates networks of interactions and then axiomatize a method of identifying the latent community structures from network data. The method is based on maximum likelihood estimation.

*Čopič is the department of economics at UCLA, email: jcopic@econ.ucla.edu. Jackson is at the Department of Economics, Stanford University, Stanford, California 94305-6072, USA, email: jacksonm@stanford.edu , web site: http://www.stanford.edu/∼jacksonm/ . Kirman is at GREQAM, EHESS, 2 rue de la Charité, 13236 Marseille CEDEX 02, France, email: kirman@univmed.fr . We gratefully acknowledge financial support under NSF grant SES-0316493 and SES-0647867, as well as from the Guggenheim Foundation, the Center for Advanced Studies in the Behavioral Sciences, the Lee Center for Advanced Networking, the Institute for Advanced Study and the Social Information Sciences Laboratory at Caltech. We thank Bhaskar Dutta, Matteo Marsili, and participants of the ISS seminar at Cornell for helpful discussions; Pritha Dev for pointing out an error in a previous version, and Rik Pieters and Hans Baumgartner for making their data available.

1 Introduction

In many settings where we observe networks of interactions, there are groups of nodes that interact more intensively with each other than with nodes outside their group. If nodes are people, they may belong to the same club, be of the same ethnicity or profession. In the case of trade unions, for example, individuals with similar jobs are more likely to interact. In many cases the underlying structure that influences network interactions is of interest but is not directly observable. In such cases we can infer which nodes should be grouped together by observing their interaction patterns. In this paper, we axiomatize a technique for uncovering latent communities that underlie networks of interactions.

Just to fix ideas, let us mention a few examples. Consider a job market network where the nodes are universities and weighted links represent the relative rates at which they hire each other's PhD graduates. Are there unobserved ideological groupings that bias hirings so that departments are more likely to hire graduates of other departments with similar ideologies? Can we reconstruct an objectively "most likely" ideological partition of universities by using the observed hiring patterns to infer the ideological biases? A similar problem, for which we examine data in the last section of this paper, is to uncover groupings of economics journals based on the rates at which they cite each other. Can we reconstruct scientific communities based on the observed citation patterns? Beyond these examples there are many others, such as uncovering biases in trading patterns, uncovering hidden organizations or cartels from networks of communications, classifying types of interactions based on networks of chemical or biological interactions, and so forth.

Given the importance in many disciplines of partitioning nodes and deriving community structures based on network data, there is a rich literature proposing a variety methods for doing so. This dates to notions of structural equivalence which identified nodes that were equivalent or interchangeable in terms of their network positions (e.g., see the seminal work on block modeling and positional analysis by Lorrain and White (1971) and White, Boorman and Breiger (1976)). As more applications have arisen, and given that nodes are rarely fully structurally equivalent, a burgeoning set of algorithms for partitioning nodes into communities based on network data has emerged .[1] As one might expect with such a

[1]See Wasserman and Faust (1994), Newman (2003) and Jackson (2008) for overviews of different aspects of the literature. These include methods that examine correlations of relationships between nodes (e.g., CONCOR, see White, Boorman and Breiger (1976), Boorman and White (1976), and Schwartz (1977)); methods based on (repeated) bisection of the network (e.g., Kernighan and Lin (1970), Fiedler (1973), and Pothen, Simon and Liou (1990)); methods based on building hierarchies of similar nodes (for instance starting with all nodes separate and then matching the two most similar according to various measures, and

variety of techniques, different methods can end up producing very different partitions from the same network data. This obviously means that the methods are identifying different things. Without some systematic study of techniques, it can be difficult if not impossible to know which method to use for any given problem, or even to know exactly what the resulting community structure means. Although the algorithms are often clever and have some nice intuition behind them, identifying community structures is still more of an art than a science.

The purpose of this paper is to provide a first step in terms of providing foundations for various approaches to partitioning nodes into community structures based on network data. In particular, we examine a simple method of identifying community structures based on maximum-likelihood estimation. The basic ideas behind the likelihood approach and the model on which it is based are as follows. (We describe this for a special case where links are either present between two nodes or not, and provide the full description of weighted and/or directed networks in the body of the paper.) The model starts with a given set of nodes. There is some true underlying community structure which is a partition of the nodes into groups. The communities of nodes can be thought of as groups of nodes that have some natural affinity for each other or some basic characteristics in common. Links between nodes are formed at random, but in a way that depends on the underlying community structure.[2] The key is that two nodes that lie in the same community are more likely to interact with each other than two nodes that lie in different communities. In particular, there is some probability p_{in} that two members of the same community will be linked to each other, and another probability p_{out} that members of different communities will be linked to each other.

As observers, we do not directly observe the partition of nodes into communities nor do we observe the probabilities with which different nodes interact (are linked to each other). Instead, we simply observe the resulting network and use that to form an estimate of which underlying community structure and probability structure would be most likely to have generated the observed network. This follows standard maximum likelihood techniques. That is, for any given community structure and p_{in} and p_{out}, where $1 \geq p_{in} > p_{out} \geq 0$, one can calculate the probability each possible network being observed. Maximum likelihood estimation identifies the community structure and p_{in} and p_{out} that maximize the probability of observing the data at hand. In fact, this method not only provides a "best" or likelihood

then iterating upwards (e.g., Lance and Williams (1967)), methods based on identifying cliques (Bron and Kerbosch (1973)), and methods based on edge removal based on betweenness measures (Girvan and Newman (2002), Tyler, Wilkinson and Huberman (2003)) or loops (Radicchi et al (2003)).

[2]The randomness could also come from measurement error. For our analysis here, the distinction is inconsequential.

maximizing network structure, but also provides a complete ordering over all community structures and in fact provides a relative likelihood that any one could have generated the data. It is then also straightforward to develop statistical measures of fit, confidence, and significance, through the standard tools associated with likelihood estimation.

This approach to identifying community structures differs from the bulk of the approaches used in the literature in the following manner: Rather than starting an algorithm based on some intuition and then declaring communities to be what the algorithm identifies, this approach begins with a model of what a community structure is and how it generates networks and then uses standard statistical techniques to recover the community structure.

Beyond offering this perspective on identifying community structures, the central technical contribution of the paper is to provide an axiomatization of this maximum likelihood approach to identifying community structures. In fact, our axiomatization provides new and basic insights into maximum likelihood estimation as a general statistical technique. We show that this technique is the only one that satisfies a set of properties including a monotonicity property (increasing the observed interaction between nodes that are grouped together keeps them grouped together), an independence axiom (if one community structure is considered more likely than another, and we rearrange the network in a way that only changes links among nodes that are grouped in the same way in both community structures then the first community structure is still considered more likely than the second) a neutrality axiom (if a community structure is identified and one rearranges some of the relationships but in a way that preserves the total number of links within groups and the total number of links across groups, then the same community structure is identified), and a normalization axiom. The purpose in offering this characterization of the maximum likelihood approach is that it provides the properties that uniquely identify this technique for grouping nodes into communities. If such characterizations are eventually provided for other methods, we will have much better understandings of the contrasts between, and relative strengths and weaknesses of various approaches.

Implementing the maximum method of identifying network structures presents two major challenges, and so we also spend some time in the paper discussing how to implement the technique. First, the p_{in} and p_{out} described above are not known to the observer but have to be estimated in conjunction with the community structure. While these estimates depend on the community structure identified, we show that there is a nice consistency property that leads to accurate estimation of these parameters. Second, a challenge faced by any method for identifying community structures is that the number of possible partitions is exponential in the number of nodes, making a full search impossible for almost all applications. We show that there are nice properties that allow us to start by identifying something which is not quite a partition, but instead is an indication of which pairs of nodes are most likely to be in the same communities. We call this a "pseudo-community structure." For example, it may be that we end up saying that nodes 1 and 2 look like they belong in the same community, and nodes 2 and 3 also do, but 1 and 3 do not. This structure, which is not quite a community structure but is easy to compute, maximizes the likelihood in a well-defined way. From there, we then work with an algorithm which finds the community structure which is closest to this "pseudo-community structure" with a high probability. We also prove that nearby structures have similar likelihoods.

In addition to the analysis of the maximum likelihood approach and its characterization, and discussion of some algorithms for implementing it; we also illustrate the method with an application to a network of citations among economics journals.

As a final note, it should be obvious that the simple model that we examine here is an oversimplification of how communities might affect network formation. In particular, the only observables here are the networks of interactions and not any other characteristics of the nodes. In most applications, there are additional data that one might draw upon. Moreover, links are formed in an over-simplified way that only considers two different probabilities: one for nodes within the same community and another for nodes in separate communities. In most interactions there may be a richer probability structure underlying how different community structures interact. However, it should also be obvious that the techniques suggested here can be generalized to richer underlying models. The maximum likelihood method is quite flexible in this regard, and the axiomatization that we provide has easy extensions to variants on the basic model. Our purpose here is not to provide the definitive method of identifying community structures, but instead to begin a program of characterizing techniques and developing a richer understanding of their relative properties.

Relation to the literature

As mentioned above this fits into a rich literature on estimating community structures from network data (again, for overviews see Wasserman and Faust (1994), Snijders and Nowicki (199?), Newman (2003), and Jackson (2008)). The contribution here is to begin a program of characterizing different techniques based on their properties, and also to base a technique on an underlying model of what community structures are and how they lead to network formation.

Given the prominence of maximum likelihood as a technique of statistical estimation, it has been used in analyzing network data. In particular, Holland, Laskey, and Leinhardt (1983) (see also Holland and Leinhardt (1977) and Fienberg and Wasserman (1981)) and

Snijders and Nowicki (1997) proposed models where nodes have group identities and then the probabilities of linking depends on those identities.[3] In order to meaningfully fit such a model one has to restrict the number of parameters as otherwise one could have each node be its own group and then assign probability 1 between groups where links exist and 0 between those where links do not exist. Here we have moved to an extreme with just two probabilities that must be the same across all communities. This simplification is useful for characterizing this sort of method and identifying properties that single out maximum likelihood as a technique. Clearly generalizing this will be useful.[4]

2 Background Definitions

2.1 Networks and Sizes

Let $N = \{1, 2, ..., n\}$ be a set of *nodes* or vertices.

A *network* on a set of nodes N is matrix $g \in \mathbb{R}_+^{n \times n}$.

We allow networks to be weighted and/or directed, although our methods also apply equally well to special cases where the network is undirected and/or unweighted.

A set of *sizes* or capacities on a set of nodes N is a matrix $s \in \mathbb{R}_+^{n \times n}$.

A network g on a set of nodes N with sizes s is *feasible* if

$$g_{ij} \leq s_{ij}$$

for each ij.

Let $G(s)$ denote the set of feasible networks on nodes N given a size matrix s.

[3] More generally, an independent paper by Handcock and Raftery (2006) uses likelihood techniques to analyze a spatial social model for generating networks. Other examples include applications of likelihood techniques to the clustering of things like Potts variables by Giada and Marsili (2005). The Potts model comes from statistical physics where Potts spins within groups are correlated but across groups are not. Giada and Marsili explore grouping objects based on correlation patterns in the observed attribute vectors, based on an underlying Gaussian model and maximum likelihood. As Giada and Marsili point out, this technique can be applied to settings beyond physics, such as to the stock market, where returns of stocks within groups are presumed correlated but across groups are not.

[4] Since the writing of this paper, independent work by Newman and Leicht (2007) has worked with a similar model to the one here with two probabilities, but their paper is interested in introducing the model whereas our focus is on providing a characterization. A new paper by Clauset, Moore and Newman (2008) provides a nice extension of the model to allow for a hierarchy of communities that dictates the relative probabilities of linking within and across communities.

Sizes represent the maximal potential interaction. In a standard 0-1 network, this would simply be 1. But in applications where one has data on the frequency or intensity of interaction, the size represents the maximal amount the two nodes could have interacted. The network g then represents the actual interaction.

As an example, consider an application where the network is composed of citations among journals during some time period. In that case each node is a journal. As any article can cite any other article at most once, the most that journal i can cite journal j is $s_i s_j$, where s_i is the number of articles published in journal i during the time period and s_j is the number of articles appearing in journal j during the time period. This is an example of a more general case where $s_{ij} = s_i s_j$, with the interpretation that node i is composed of s_i units, and that each unit in every node can have at most one directed connection with each other unit in each other node.

As another example, consider a network of coauthorships, where each node is an individual. Here one possibility is that $s_{ij} = 1$ for all ij, as there is a potential for two nodes to be co-authors. This would apply if the network simply keeps track of whether two individuals have ever written a paper together. If instead, the network keeps track of how many papers two individuals have written together over some time period, then s_{ij} would be the capacity of papers that could be written involving individuals i and j during the time period, and g_{ij} would be the number that they did write.

2.2 Community Structures

A community structure is a partition of the set of nodes.

Let $\Pi(N)$ be the set of all partitions of N.[5]

For any $\pi \in \Pi(N)$ and any $i \in N$, let $c_\pi(i)$ be the component of π containing i.

To illustrate the definition, here is an example of a community structure with six communities on a sixteen node undirected network, where sizes are all 1 and pairs of nodes are either linked or not linked.

2.3 Community Structure Rankings

We use an approach which is new to the community structure literature: that of a ranking of community structures. This is a bonus of the likelihood approach, and embodies our view of community structures. The network data are generated with some inherent randomness

[5] Thus an element $\pi \in \Pi(N)$ is a collection of subsets of N such that $\cup_{c\in\pi} c = N$ and if $c, c' \in \pi$ and $c \neq c'$ then $c \cap c' = \emptyset$.

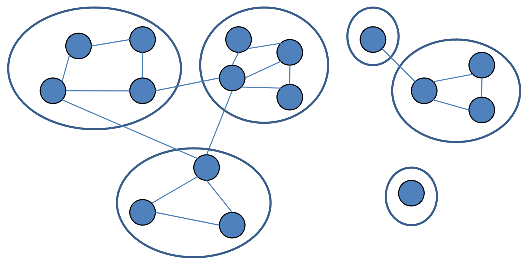

Figure 1: An Example of Community Structure in a 0-1 Undirected Network, where Ovals Capture the Communities

and/or are subject to measurement error. As such, rather than only trying to uncover a single "best" community structure, it makes sense to rank all possible community structures. The interpretation is that higher community structures in the rankings look more likely to be the "true" underlying community structure that generated the data; but lower ones in the rankings could still be the "true" one.

In fact, the (maximum) likelihood method not only allows us to rank the possible community structures, but also tells us the relative likelihood with which each one is the true community structure. This is an attractive by-product of our approach.

Let $R(N)$ denote the set of all weak orders (complete and transitive binary relations) on the set community structures on the set of nodes N, and let $\succeq$ denote a generic element of $R(N)$.

Thus, $\pi \succeq \pi'$ indicates that the community structure π is a (weakly) "better" or more likely partitioning of the nodes of N than is π' according to the criterion that is embodied in $\succeq$.

We wish to produce a ranking of community structures based on observed data. A general method for doing this should describe how the ranking will be determined as a function of the observed data (g, s). That is, a method of ranking community structures should tell us how we will determine a ranking (and then implicitly, a "best" community structure) for each possible situation we might face.

Formally, a *community structure ranking* is a function that selects a weak order in $R(N)$ over all community structures for each observation of a network and sizes, (g, s).

We denote such a ranking by $\succeq_{s,g}$, which indicates the ranking over the community structures that results when having observed sizes described by s and the network $g \in G(s)$.[6]

Let $\succ_{s,g}$ denote the associated strict relationship associated with the community structure ranking $\succeq_{s,g}$, so that $\pi \succ_{s,g} \pi'$ if and only if *not* $\pi' \succeq_{s,g} \pi$.

3 The Model and Likelihood Ranking

We now present the likelihood method for ranking community structures. It is naturally associated with a view of what communities and networks represent.

3.1 A Model and its associated Likelihood Function

Our model is that a community is a group of similar nodes in terms of their probabilities of interaction with other nodes. So, there is some true community structure $\pi \in \Pi$ and the probability of any two given nodes inside a community interacting is p_{in} and the probability of two given nodes from different communities interacting is p_{out}, where $1 \geq p_{in} > p_{out} \geq 0$.

If s_{ij} is the potential size of the interaction between i and j, then the chance of seeing exactly g_{ij} interactions between i and j is proportional to (omitting the binomial coefficient)

$$p_{in}^{g_{ij}}(1-p_{in})^{s_{ij}-g_{ij}}$$

if i and j are in the same community under π, and proportional to

$$p_{out}^{g_{ij}}(1-p_{out})^{s_{ij}-g_{ij}}$$

if i and j are in different communities under π.

From this, we can calculate what would be the probability of observing any given $g \in G(s)$ if the true community structure were $\pi \in \Pi$. This likelihood is[7,8]

$$L_{s,g}(\pi) = C \times_{i\in N} \left[\left(\times_{j\in c_\pi(i)} (p_{in})^{g_{ij}}(1-p_{in})^{s_{ij}-g_{ij}} \right) \left(\times_{j\in N\setminus c_\pi(i)} (p_{out})^{g_{ij}}(1-p_{out})^{s_{ij}-g_{ij}} \right) \right], \quad (1)$$

[6]As N is generally fixed, we omit it from the notation even though a community structure ranking is also a function of N.

[7]We adopt the convention that $0^0 = 1$.

[8]This expression is for the case of a directed network, where g_{ij} is not constrained to be equal to g_{ji}. The case of a non-directed network only needs to consider one direction. The expression here simply ends up being the square of that calculation if the network is non-directed; and so a simple adjustment to the expressions that follow handles the non-directed case.

where C is a constant consisting of binomial coefficients. C does not affect the relative likelihoods, and so we can effectively ignore it.[9]

Under the *likelihood ranking*, the community structures are ranked according to the likelihood that they generate. That is, π as being (weakly) more likely than π' given the sizes s and observed network g, if $L_{s,g}(\pi) \geq L_{s,g}(\pi')$. This provides a likelihood ranking, $\succeq_{s,g}^{L(p_{in},p_{out})} \in R(N)$ on $\Pi(N)$.[10]

The likelihood ranking presumes knowledge of p_{in} and p_{out}. These can be estimated from the data, as we describe in our section on estimation.

3.2 Alternative Representations of the Likelihood Function

We rewrite the likelihood function in a way that makes clear that the likelihood ranking has some simple and attractive properties.

Given $\pi \in \Pi$, let

$$In(\pi) = \{ij \mid i \in N, j \in c_\pi(i)\},$$

and

$$Out(\pi) = \{ij \mid i \in N, j \notin c_\pi(i)\}.$$

Thus, $In(\pi)$ is the set of all pairs of nodes that are in the the same community under π and $Out(\pi)$ is the set of all pairs of nodes that are in different components under π.

Let

$$T(g) = \sum_{ij \in N \times N} g_{ij} \text{ and } T(s) = \sum_{ij \in N \times N} s_{ij}.$$

These are the total weighted links in the network g and the total sizes from s. Let

$$T^{In(\pi)}(g) = \sum_{ij \in In(\pi)} g_{ij} \text{ and } T^{In(\pi)}(s) = \sum_{ij \in In(\pi)} s_{ij}$$

and similarly define $T^{Out(\pi)}(g)$, and $T^{Out(\pi)}(s)$.

This keeps track of the total weights that are inside communities and outside communities for a given community structure π.

[9]Note that C is dependent on the observed g and s. Nevertheless, it does not vary as we vary the community structures, and thus can be ignored in ranking community structures and in comparing their relative likelihoods.

[10]Given the relative likelihood function L, we not only have an ordering, but we also have relative likelihoods. In addition to ranking community structures, this will allow us to say how close two community structures are to each other in a very precise sense. We return to this below.

By taking logs of the likelihood function L from (**??**), we preserve the ranking over partitions, but end up with a function that is easier to work with. Thus, let $\ell_{s,g}(\pi) = log(L_{s,g}(\pi))$. We can then write

$$\ell_{s,g}(\pi) = log(L_{s,g}(\pi)) = k_1 T^{In(\pi)}(g) + k_2 T^{In(\pi)}(s) + k_3 T^{Out(\pi)}(g) + k_4 T^{Out(\pi)}(s), \quad (2)$$

where $k_1 = log(p_{in}/(1-p_{in}))$, $k_2 = log(1-p_{in})$, $k_3 = log(p_{out}/(1-p_{out}))$, and $k_4 = log(1-p_{out})$.

This representation of the log-likelihood makes it clear that the likelihood of a given community structure only depends on the total number of "inside connections" within communities relative to their total capacity, as well as the total number of "outside connections" across communities relative to their total capacity. The likelihood ranking thus does not care where the connections fall exactly, but simply how the total number of inside (within-community) connections compares to what the total capacity is, as well as how the total number of outside (cross-community) connections compares to its capacity.

Noting that $k_1 > k_3$ but $k_2 < k_4$, we see why an optimum will not always be an extreme partition where either all nodes are grouped together or all nodes are grouped apart. As more nodes are grouped together, some interactions are shifted from $T^{Out(\pi)}(g)$ to $T^{In(\pi)}(g)$, and are thus weighted more (by the factor k_1 rather than k_3), but the sizes are also shifted from $T^{Out(\pi)}(s)$ to $T^{In(\pi)}(s)$ which affects the term in the opposite direction. It is this tradeoff that determines the optimal partition.

Since $T^{out}(\cdot) = T(\cdot) - T^{in}(\cdot)$, we can rewrite (**??**) as

$$\ell_{s,g}(\pi) = (k_1 - k_3) T^{In(\pi)}(g) + (k_2 - k_4) T^{In(\pi)}(s) + k_3 T(g) + k_4 T(s). \quad (3)$$

Thus, if we compare two partitions π and π', it follows that $\pi \succ_{s,g}^{L(p_{in},p_{out})} \pi'$ if and only if

$$(k_1 - k_3) T^{In(\pi)}(g) - (k_4 - k_2) T^{In(\pi)}(s) > (k_1 - k_3) T^{In(\pi')}(g) - (k_4 - k_2) T^{In(\pi')}(s). \quad (4)$$

4 Properties of Community Structure Rankings

Based on the simple representation of the likelihood ordering over community structures in (**??**), we can derive some basic properties that characterize the likelihood ranking.[11]

[11] Throughout the statement of properties and characterization, we treat N as given and fixed, and one can simply require that the following properties hold for each N and then the characterizations still proceed exactly as stated.

The first property is a very weak one that implies that we are maximizing something rather than minimizing it.

Let us say that (g, s) has only two non-degenerate nodes if there exist $i, j \in N$ such that $g_{kl} = s_{kl} = 0$ for all $kl \neq ij$.

Property **1** [Weak Monotonicity] *A community structure ranking $\succ_{\cdot,\cdot}$ is weakly monotonic if for each (g, s) with only two non-degenerate nodes $i, j \in N$, and any two community structures π and π' such that $ij \in In(\pi) \cap Out(\pi')$:*

- *If $0 < g_{ij} = s_{ij}$ then $\pi \succ_{s,g} \pi'$,*
- *if $0 = g_{ij} < s_{ij}$ then $\pi' \succ_{s,g} \pi$, and*
- *if $0 = g_{ij} = s_{ij}$ then $\pi \sim_{s,g} \pi'$.*

While the property has a long statement, it is very simple. The property concentrates on settings where we are essentially just comparing two-node networks, as $g_{kl} = s_{kl} = 0$ for all $kl \neq ij$. In such situations it says that if the amount of interaction is equal to the sizes, then a partition that groups the two nodes together is preferred to one that has them apart; while if the amount of interaction is 0 then it is better to have the nodes in separate communities than together. The last part simply says that if all capacities are 0, then we cannot order partitions so all of them are indifferent to each other.

This axiom should be satisfied by any reasonable community structure ranking.

The next property, independence, states that the ranking of two partitions does not depend on the links that are classified similarly across the two partitions (either both "in" or both "out").

Property **2** [Independence] *A community structure ranking $\succeq_{\cdot,\cdot}$ satisfies independence if*

$$\pi \succeq_{s,g} \pi' \iff \pi \succeq_{s',g'} \pi', \tag{5}$$

for every feasible (s, g) and (s', g') such that $g_{ij} \neq g'_{ij}$ or $s_{ij} \neq s'_{ij}$ implies that $ij \in In(\pi) \cap In(\pi')$ or $ij \in Out(\pi) \cap Out(\pi')$.

The property says that if we look at two different network situations (s, g) and (s', g'), and two community structures such that any differences between the data only occurs in parts of the community structures that are identical, then the ordering over the community structures is the same. Put differently, if we have some pair of nodes ij which either lie in the same community in both π and π', or else are both in different communities in both π and π', then making changes only the interaction or size between i and j will not change the relative ranking of the two community structures.

This property is clearly satisfied by the likelihood ranking as we saw in (**??**) that only the totals of "ins" and "outs" matter, and so interactions that enter in similar ways in both partitions are irrelevant to the ordering, we only need to keep track of which interactions change as we change partitions.

The next property, Neutrality, while tedious to write, is also straightforward. It states that if we rearrange interactions, while keeping the relative difference between the total of "ins" and "outs" the same between two community structures, then the relative ranking of those community structures should not be affected. That is, take two partitions π, π', and take $i, j, k, l \in N$. Neutrality then says that if under π, i, j are in the same community and k, l are in different communities, while under π', i, j are in different communities and k, l are in the same community, then if we increase the level of interaction between i and j by an amount $x > 0$ and also increase the level of interaction between k and l by x, the ranking between π and π' doesn't change. Similarly, the ranking is unchanged if we transfer interaction from nodes that are in different communities under both partitions to nodes that are in the same community under both partitions. The same is required of changes in sizes.

PROPERTY **3** [NEUTRALITY] *A community structure ranking $\succeq_{\cdot,\cdot}$ satisfies neutrality if the following holds. Consider any s, s', $g, g' \in G(s) \cap G(s')$, $i, j, k, l \in N$, $x > 0$, and $\pi, \pi' \in \Pi(N)$.*

1. *If $ij \in In(\pi) \cap Out(\pi')$ and $kl \in Out(\pi) \cap In(\pi')$, and $g_{ij} = g'_{ij} + x$, $g_{kl} = g'_{kl} + x$, and $g_{hm} = g'_{hm}$ for $hm \notin \{ij, kl\}$, then*
$$\pi \succeq_{s,g'} \pi' \iff \pi \succeq_{s,g} \pi'.$$
Similarly, if $s'_{ij} = s_{ij} + x$, $s'_{kl} = s_{kl} + x$, and $s_{hm} = s'_{hm}$ for $hm \notin \{ij, kl\}$, then
$$\pi \succeq_{s',g} \pi' \iff \pi \succeq_{s,g} \pi'.$$

2. *If $ij \in In(\pi) \cap Out(\pi')$ and $kl \in In(\pi) \cap Out(\pi')$, and $g_{ij} = g'_{ij} + x$, $g_{kl} = g'_{kl} - x$, and $g_{hm} = g'_{hm}$ for $hm \notin \{ij, kl\}$, then*
$$\pi \succeq_{s,g'} \pi' \iff \pi \succeq_{s,g} \pi'.$$
Similarly, if $s_{ij} = s'_{ij} + x$, $s_{kl} = s'_{kl} - x$, and $s_{hm} = s'_{hm}$ for $hm \notin \{ij, kl\}$, then
$$\pi \succeq_{s',g} \pi' \iff \pi \succeq_{s,g} \pi'.$$

This neutrality property is a key one in identifying the likelihood ranking, as it really is the essence of the fact that we do not care about precisely which nodes are involved in interactions, just how many interactions are occuring within communities and how many interactions are occurring across communities (relative to the sizes). In our model, this is a natural requirement since communities are equivalence classes of nodes and it does not matter which of the nodes within the equivalence class is involved in a specific interaction.

The last property, scaling, requires that there exist some relative rate so that if we increase interaction at a rate proportional to sizes, then the relative rankings of community structures is unaffected.

PROPERTY **4** [SCALING] *A community structure ranking* $\succeq_{\cdot,\cdot}$ *satisfies scaling if there exists* $\gamma > 0$ *such that if* g, s, g', s', ij, *and* x *are such that* $g'_{ij} = g_{ij} + x$ *and* $s'_{ij} = s_{ij} + \gamma x$, *and* $g'_{kl}, s'_{kl} = g_{kl}, s_{kl}$ *otherwise, then*

$$\pi \succeq_{s,g} \pi' \iff \pi \succeq_{s',g'} \pi'.$$

4.1 Two Characterization Theorems

THEOREM **1** *A community structure ranking* $\succeq_{\cdot,\cdot}$ *satisfies monotonicity, independence, neutrality, and scaling, if and only if there exist* $p_{in}, p_{out} \in (0,1)$ *such that* $\succeq$ *is the likelihood ranking associated with probabilities* p_{in} *and* p_{out}.

To prove Theorem **??**, it is useful to prove an auxiliary theorem, which replaces independence and neutrality with another property. This property says that rankings of partitions depend only on how they differ in terms of the total interaction they have within communities and total size they have within communities. (Note that since the total interaction summed (within and across communities) is the same in both cases, comparing the "ins" also incorporates the relevant information about the "outs".)

Let

$$D(g, \pi, \pi') = T^{In(\pi)}(g) - T^{In(\pi')}(g)$$

be the difference between the partitions in terms of the total interaction within communities, and similarly let

$$D(s, \pi, \pi') = T^{In(\pi)}(s) - T^{In(\pi')}(s).$$

Property **5** [Internal Differences] *A community structure ranking $\succeq_{\cdot,\cdot}$ satisfies internal differences if whenever (g, s), (g', s'), π, π', π'', π''' are such that $D(g, \pi, \pi') = D(g', \pi'', \pi''')$ and $D(s, \pi, \pi') = D(s', \pi'', \pi''')$, then*

$$\pi \succeq_{s,g} \pi' \iff \pi'' \succeq_{s',g'} \pi'''.$$

This incorporates both the neutrality and independence conditions.

Theorem **2** *A community structure ranking $\succeq_{\cdot,\cdot}$ satisfies monotonicity, internal differences, and scaling, if and only there exist $p_{in}, p_{out} \in (0, 1)$ such that $\succeq$ is the likelihood ranking associated with probabilities p_{in} and p_{out}.*

The proof of both theorems appears in the appendix.

It is natural to ask whether the above characterizations of the likelihood ranking are tight. It is clear that scaling and neutrality imply a form of linearity. Relaxing these will yield other rankings with non-linear functional forms. Relaxing monotonicity, we would get for example a "minimum-likelihood" ordering as well as anything in between the two. An example of a ranking that doesn't satisfy independence is one similar to a likelihood ranking, but where the comparison between each two partitions depends on the number of common components.

5 Communities and Consolidation

A natural interpretation of communities is that they are collections of "equivalent" nodes. This can be formalized by requiring that if we have a community of nodes and we combine those nodes to become one large node, then the overall structure between communities would be preserved. We formalize this as follows.

Consolidations

We call a combination of two nodes into one a consolidation. The formal definition is long, but simple and straightforward. The idea is that the single node inherits the sum of the sizes and the interactions of the nodes that it replaces.

Consider N, s, and $g \in G(N, s)$, and some $i, j \in N$. Define the *consolidation of g by combining i into j*, denoted $g^{(i\sim j)} \in G(N \setminus i, s^{(i\sim j)})$, as follows.

First, the sizes associated with the new node are the sum of the sizes of the combined nodes, while other nodes keep their sizes. Thus,

$$s_{jk}^{(i\sim j)} = s_{ik} + s_{jk}, \text{ and } s_{kj}^{(i\sim j)} = s_{ki} + s_{kj} \text{ for } k \neq i, j,$$

$$s_{jj}^{(i\sim j)} = s_{ii} + s_{jj} + s_{ji} + s_{ij}, \text{ and}$$
$$s_{kl}^{(i\sim j)} = s_{kl} \text{ if } \{k, l\} \cap \{i, j\} = \emptyset.$$

Second, the interactions of the new node are the sum of the connections the previous nodes (so as above, with g replacing s).

If a consolidation involves two nodes that are in the same community according to some community structure, then we have a corresponding collapsing of the community structure. Given $\pi \in \Pi(N)$ such that $j \in c_\pi(i)$, let $\pi^{(i\sim j)} \in \Pi(N \setminus i)$ be the associated consolidated community structure. That is, $c_{\pi^{(i\sim j)}}(j) = c_\pi(j) \setminus i$ and $c_{\pi^{(i\sim j)}}(k) = c_\pi(k)$ for $k \notin c_\pi(j)$.

The next property requires that if we consolidate two nodes that are in the same community under two different partitions, then the relative ranking of the corresponding consolidated community structures is the same as the relative ranking of the original community structures.

Property **6** [Consolidation] *A community structure ranking $\succeq_{\cdot,\cdot,\cdot}$ satisfies consolidation if for every (N, s), $g \in G(N, s)$ and consolidation of g combining i into j, $g^{(i\sim j)} \in G(N \setminus i, s^{(i\sim j)})$,*

$$\pi \succeq_{N,s,g} \pi' \iff \pi^{(i\sim j)} \succeq_{N\setminus i, s^{(i\sim j)}, g^{(i\sim j)}} (\pi')^{(i\sim j)}$$

for every $\pi \in \Pi(N)$ and $\pi' \in \Pi(N)$ such that $i \in c_\pi(j)$ and $i \in c_{\pi'}(j)$.

Proposition **1** *Given p_{in} and $p_{out} \in [0, 1]$, with $p_{in} \geq p_{out}$, the corresponding likelihood ranking satisfies consolidation.*

The proof follows directly from the expression for the log-likelihood given in (**??**).

6 Implementation and Estimation

In this section we present an application of the maximum-likelihood ranking. In order to apply the ranking, we have to resolve several estimation issues.

6.1 Estimation of Probabilities

The probabilities p_{in} and p_{out} are often unknown and have to be estimated. In general, given a partition π there are associated estimates $\hat{p}_{in}(\pi), \hat{p}_{out}(\pi)$ given by[12]

$$\hat{p}_{in}(\pi) = \frac{T^{In(\pi)}(g)}{T^{In(\pi)}(s)},$$

$$\hat{p}_{out}(\pi) = \frac{T^{Out(\pi)}(g)}{T^{Out(\pi)}(s)}.$$

Given that the estimates of p_{in} and p_{out} depend on the partition, and the optimal partition (in terms of being highest ranked) depends on p_{in} and p_{out}, some interesting issues arise. One obvious requirement is that the partition π be optimal (of maximum likelihood) under $\hat{p}_{in}(\pi), \hat{p}_{out}(\pi)$. We call this property *Consistent Optimality.* Hypothetically, there could exist more than one combinations of a partition and estimated probabilities satisfying this requirement. We show that as the sizes become large, then the true community structure satisfies consistent optimality and only that community structure. Proving that the true partition satisfies consistent optimality is a fairly direct consequence of the law of large numbers. Proving that no other partition satisfies consistent optimality is slightly more subtle, and follows from the next proposition.

PROPOSITION **2** *Let $\bar{\pi} \in \Pi(N)$ be nondegenerate and be the unique optimal partition under the likelihood model given probabilities $p_{in}, p_{out} \in (0,1), p_{in} > p_{out}$. If g, s are such that such that $\frac{g_{ij}}{s_{ij}} = p_{in}, \forall ij \in In(\pi)$, and $\frac{g_{ij}}{s_{ij}} = p_{out}, \forall ij \in Out(\pi)$, then $\pi \in \Pi(N)$ satisfies consistent optimality if and only if $\pi = \bar{\pi}$.*

The law of large numbers implies the following corollary.

COROLLARY **1** *Let s^t be a sequence of sizes such that $\lim_{t\to\infty} s^t_{ij} = \infty, \forall ij \in N \times N$. For each $t = 1, ...,$ let a network $g^t \in G(s^t)$ be generated by the likelihood model with associated probabilities $p_{in} > p_{out}$ and some true nondegenerate community structure $\pi \in \Pi(N)$. With probability 1, as $t \to \infty$, there exists a unique $\hat{\pi}$, which is optimal according to the ordering $\succ_{L(\hat{p}^t_{in}(\hat{\pi}), \hat{p}^t_{out}(\hat{\pi}))}$, and $\hat{\pi} = \pi$ and $(\hat{p}^t_{in}(\hat{\pi}), \hat{p}^t_{out}(\hat{\pi})) \to (p_{in}, p_{out})$.*

Thus, as sizes become larger, the only partition that will end up satisfying consistent optimality is the true partion.

[12] We take 0/0=1. In order to have well-defined estimates in all cases, when faced with a degenerate partition set $\hat{p}_{in}(\pi_d) = \max_{ij} \frac{g_{ij}}{s_{ij}}$ and $\hat{p}_{out}(\pi_t) = \min_{ij} \frac{g_{ij}}{s_{ij}}$

6.2 Nearby Partitions have Similar Likelihoods

We are still faced with the problem of an exponential number of partitions to compare in finding an optimal one. It is useful to show that partitions that are "close" to each other have likelihoods that are close to each other. In particular, we now show the following result: If two partitions are close to each other according to a simple metric that counts how many objects are in different groups across the two partitions, then the likelihoods of the two partitions are also close to each other. This is useful, since then if we can examine a grid of partitions and identify the likelihood of each element of the grid, we can bound the likelihood of any partition that we do not directly examine by knowing that is close to something on the grid. By being sure to search a fine enough grid, we are sure to find a community structure close to the maximum likelihood structure.

Recall that

$$\ell_{s,g}(\pi) = log(L_{s,g})(\pi) = k_1 T^{In(\pi)}(g) + k_2 T^{In(\pi)}(s) + k_3 T^{Out(\pi)}(g) + k_4 T^{Out(\pi)}(s), \qquad (6)$$

Let us consider a situation where $1 > p_{in} > p_{out} > 0$, and all sizes s_{ij} are bounded above by s_1, and the true partition is given by π_0.

As the likelihood L is a number strictly between 0 and 1, ℓ is a negative number. Let us define a loose upper bound on the difference between $\ell_{s,g}(\pi)$ and $\ell_{s,g}(\pi')$, for two arbitrary partitions.

Let

$$Dist(\pi, \pi') = |In(\pi) \Delta In(\pi')|$$

where Δ is the symmetric difference. Note that $|In(\pi)\Delta In(\pi')| = |Out(\pi)\Delta Out(\pi')|$.

Thus, from (**??**) we easily obtain an upper bound

$$|\ell_{s,g}(\pi) - \ell_{s,g}(\pi')| \leq Dist(\pi, \pi') \left[|k_1 - k_3| + |k_2 - k_4|\right] s_1. \qquad (7)$$

Thus, the difference in log-likelihoods is bounded by a linear factor of $Dist(\pi, \pi')$. Thus, we have deduced the following proposition.

Proposition **3** *Consider a p_{in}, p_{out}, $1 > p_{in} > p_{out} > 0$, in a situation where s_{ij}'s are bounded above. Then there exists a constant K, independent of N, such that*

$$|\ell(\pi) - \ell(\pi')| < K Dist(\pi, \pi').$$

It is important that the constant K be independent of N. Otherwise, fixing any N there are only finitely many partitions, and so obtaining such a relationship would be trivial. The fact that this works for any N, means that there is truly a relationship between the distance between partitions and the difference in log-likelihoods.

Bounding the absolute difference in log-likelihoods may not be so informative in some problems, as we might not know how much variation in log-likelihoods there is to begin with. Thus, it is also useful to bound the *relative* difference. To do this, we begin by defining a relative measure of distance between partitions.

Let $dist(\pi, \pi')$ be the normalized distance between two partitions. That is,

$$dist(\pi, \pi') = \frac{Dist(\pi, \pi')}{n(n-1)/2},$$

where $n(n-1)/2$ is the maximal possible distance between any two partitions (i.e., the distance between the discrete and degenerate partitions which differ in the grouping of all $n(n-1)/2$ pairs of nodes) and n is the number of nodes.

Next, we need some measure of how much variation in log-likelihoods we should expect. We do this by computing the distance between the optimal partition and the worst possible grouping of nodes, which is essentially the opposite of the optimal partition (all ij's in $In(\pi)$ are switched to *Out* and vice versa). This is not necessarily a partition, but it gives us a measure of the order of magnitude of how much the log-likelihood varies as we change from the optimal partition to the worst possible grouping of nodes. Let $M(n)$ denote this magnitude, which can vary with n. In fact, in the appendix, we bound this below by a factor that is proportional to $n(n-1)/2$. From this and Proposition **??** we then deduce the following.

PROPOSITION **4** *Consider any p_{in}, p_{out}, $1 > p_{in} > p_{out} > 0$, and suppose that each $s_{ij} = s_1 > 0$ for some s_1. There exists a constant k, such that for any n and partitions π and π'*

$$\frac{|\ell(\pi) - \ell(\pi')|}{M(n)} < k dist(\pi, \pi'),$$

Proposition **??** helps in developing algorithms. By searching over a grid of partitions that comes close enough to any given partition, then we are sure to get within some distance of the optimal log-likelihood. Thus we can simplify the problem of approximating the optimal log-likelihood into a problem of approximating partitions.

6.3 Pseudo Community Structures

Before proceeding, we make another observation about simplifying the problem. If we loosen the problem, so that instead of searching over partitions, we instead search over groupings of pairs of nodes into the In and Out categories, then it is very easy to derive an optimal log-likelihood. Essentially, this is a problem where we can classify each pair of nodes into being "in the same community" or "in different communities", without worrying about whether there is consistency across nodes. For example, in doing this we might end up classifying i in the same community as j, and j in the same community as k, but i and k in different communities. Thus, this is not a community structure as it is not a partition over nodes. We call such a structure a *pseudo community structure*.

More formally, a *pseudo community structure* $\widehat{\pi}$ is a subset of $N \times N$ with the interpretation that if $\{i, j\} \in \widehat{\pi}$ then i and j are in the same community. If a pseudo community structure is transitive, so that $\{i, k\}$ is in $\widehat{\pi}$ whenever $\{i, j\}$ and $\{j, k\}$ are both in $\widehat{\pi}$ for some j.

The method for finding the Pseudo community structure is as follows. Given any estimates for p_{in} and p_{out}, we have estimates for the parameters $k_1, \ldots, k_4$ in (**??**). Then we set ij in the In category, if and only if $k_1 g_{ij} + k_2 s_{ij} > k_3 g_{ij} + k_4 s_{ij}$, or

$$\frac{g_{ij}}{s_{ij}} > \frac{k_4 - k_2}{k_1 - k_3}.$$

This procedure only requires n^2 steps. Constructing a pseudo community structure in this manner we obtain the absolute highest possible log-likelihood score. This is also a computationally easy task. If the pseudo community structure that we generate turns out to be a partition, then it is the community structure that maximizes the likelihood. The resulting pseudo community structure will not always be a community structure; but then if we can find a community structure that is close to it, then by the above propositions,[13] we obtain an approximately optimal community structure.

The following proposition shows that as sizes grow the pseudo community structure constructed as above will be a community structure with a probability converging to 1.

PROPOSITION **5** *Let s^t be a sequence of sizes such that $\lim_{t\to\infty} s^t_{ij} = \infty, \forall ij \in N \times N$. For each $t = 1, ...,$ let a network $g^t \in G(s^t)$ be generated by the likelihood model with some associated probabilities $p_{in} > p_{out}$ and some true nondegenerate community structure $\pi \in \Pi(N)$. Let $\widehat{\pi}^t$ be the pseudo community structure that maximizes the likelihood on $g^t \in G(s^t)$. Then, with probability 1, as $t \to \infty$, $\widehat{\pi}^t$ is a community structure and optimal according to the ranking $\succ^{L(\widehat{p}^t_{in}(\widehat{\pi}^t), \widehat{p}^t_{out}(\widehat{\pi}^t))}$, and $\widehat{\pi}^t \to \pi$, and $(\widehat{p}^t_{in}(\widehat{\pi}^t), \widehat{p}^t_{out}(\widehat{\pi}^t)) \to (p_{in}, p_{out})$.*

The proposition follows from the fact that as $s^t \to \infty$, the probability that any two nodes are misidentified as being in the same group or separate groups goes to 0 by the strong law

[13]The propositions extend directly to pseudo community structures.

of large numbers. As the set of nodes is fixed, the proposition follows directly and so the proof is omitted.

6.4 An algorithm

Our practical algorithm for finding the maximum-likelihood partition is as follows. We set up a grid of p_{in}'s and p_{out}'s. For each point on that grid we do the following. First we construct the pseudo community structure that maximizes the overall likelihood of observing the given data (g, s). Second, we search for the community structure that is closest to the pseudo community structure. From the results above, this will be close to an overall maximizer. These two steps are described as follows.

In the first step, a pseudo community structure is obtained by examining each pair of nodes i and j and checking whether the likelihood is higher when the pair is put *In* category (with the interpretation that they should end up in the same community) or the *Out* category (that they should end up in different groups). In particular, working from (**??**), if

$$g_{ij} \log\left(\frac{p_{in}(1-p_{out})}{p_{out}(1-p_{in})}\right) > s_{ij} \log\left(\frac{1-p_{out}}{1-p_{in}}\right)$$

then i and j are considered to be in the *In* category and not otherwise. This can be done independently across pairs of nodes in any order since we are constructing a pseudo community structure and the status of one pair of nodes has no impact on other pairs.

In a second step, given that the pseudo-community structure is not a community structure, we search for the community structure that is closest to this pseudo community structure. In most applications, this can only be done approximately since the number of community structures close to a pseudo community structure grows exponentially in the number of nodes. The approach we follow in the application below is described as follows. Let $\widehat{\pi}$ be the pseudo community structure obtained from the first step. Given $\widehat{\pi}$ and a subset of nodes $a \subset N$, let $m_{\widehat{\pi}}(a)$ be the number of pairs of nodes that have to be added to or deleted from $\widehat{\pi}$ in order to make a a completely connected component (so that all pairs of nodes in a are included and no pairs of nodes that include one element in a and one out of a are included). Let $m^*(\widehat{\pi})$ be the minimum number of pairs of nodes that have to be added to or deleted from $\widehat{\pi}$ to turn it into a community structure. We can then write a Bellman equation,

$$m^*(\widehat{\pi}) = \min_{a \subset N} \left(m_{\widehat{\pi}}(a) + m^*(\widehat{\pi}\mid_a)\right).$$

Even though this suggests an iterative procedure for finding the closest community structure, the procedure can be impossible to implement because of all the subsets of nodes that need to be checked. The algorithm that we implement is a variation of this. Randomly pick one of the nodes x that in the most pairs in $\widehat{\pi}$. We then examine random subsets of x's (extended) neighborhoods in $\widehat{\pi}$ to see which one requires the minimal number of added or deleted links in order to become a community. Specifically, starting with a parameter δ fixed in [0,1]:

1 Start with a pseudo community structure $\widehat{\pi}$, and let $MD(\widehat{\pi})$ be the nodes with maximal degree (involved in the most pairs) in $\widehat{\pi}$

2 Uniformly at random pick one node $x \in MD(\widehat{\pi})$.

3 Randomly pick a subset S of x's neighbors in $\widehat{\pi}$.

4 with probability δ add a uniformly randomly chosen neighbor of a uniformly randomly chosen node in S (if it is not already in S), then begin step [4] again with the new set S. With probability $1 - \delta$ (or if there are no more nodes that are neighbors of any nodes in S that are not already in S), then proceed to step [5].

5 Make S into a community by adding all pairs of nodes in S to $\widehat{\pi}$ and deleting any pairs of nodes that involve only one node in S.

6 Iterate on steps 3 to 5 a number of times equal to the number of nodes. Select the choice of S that requires the fewest number of added and deleted pairs in step 5.

7 Discard the resulting S from [6] as a community, and repeat steps [1] to [6] on $\widehat{\pi}$ restricted to the remaining nodes.

8 Stop when the remaining $\widehat{\pi}$ in step 7 is a community structure (it may be empty).

Our algorithm for finding the closest community structure to the pseudo community structure can be biased toward communities that are too small, depending on how S is chosen in step [3] above (for instance, by picking it by flipping a fair coin to decide if a given neighbor is in or out). Thus, we add a third step where we consolidate two nodes that are in the same community obtained from the process described above. We then repeat the process on the consolidated network that now forces these two nodes to be in the same community. We keep iterating until either the resulting community structure consists of singletons, or the community structure of the consolidated problem has a lower likelihood that the previous community structure.

6.5 An Application

We now illustrate the likelihood method. This is not meant to be an empirical analysis of economics journals, but rather to illustrate the method and to show how it compares a standard method.

The data are of cross-citations of 42 economic journals for the period of 1995-1997, from Pieters and Baumgartner (2002). The nodes are journals and the number of articles, and the entries g_{ij} are the number of citations by articles published in journal i during the time period of 1995-1997 of articles in journal j (of any previous time). We take a short cut in order to estimate the sizes s_{ij} as follows. The exact count should be the sum across articles published in i during 1995-1997 of the total number of articles that were ever published in j prior to the respective article in i. We simply estimate the size s_{ij} to be a factor proportional to the relative sizes of journals i and j in terms of articles published per year.[14] We estimate these based on the respective number of articles published in 2003, as reported on the ISI web of science journal performance metrics.

Table 1 presents a summary of the sample. Tables 2-4 present a few estimated partitions. Table 2 presents the partition obtained by our algorithm. This partition approximately satisfies consistent optimality. For comparison, we present in Table 3 the partition obtained by Pieters and Baumgartner (2002), which has a lower likelihood than either of the other partitions.[15] Table 4 presents the partition with the highest likelihood which we found by a heuristic search. This partition was found via a combination of exhaustive searches over subsets of nodes of sufficiently small sizes (15 nodes) with some subjective decisions over sets of nodes to examine. The community structure in Table 4 has a higher likelihood than that found via the algorithm, which is not surprising, but suggests that finding better approximation algorithms is worthy of study.

The partition that we found which has the highest likelihood (see Table 4) is represented in Figure **??** where the nodes of the same color are in the same community and the thickness of the lines represents the weights of the citations.

[14] So, this approximation does not account for the fact that some journals have been around for more years than others or that the number of articles per year may have varied differently across journals over time. Noting that there is a strong bias towards more recent articles in citations, this estimate should not be far off.

[15] This data set was studied by Peters and Baumgartner [2002], and they derived community structures using a hierarchical clustering algorithm and prespecifying that there should be seven communities. The partition that they obtain is different than the partition that we obtain. Most notably, our estimated maximum-likelihood partition has nineteen communities, while theirs has a pre-specified seven. As one should expect, their partition has a lower likelihood score according to our measure.

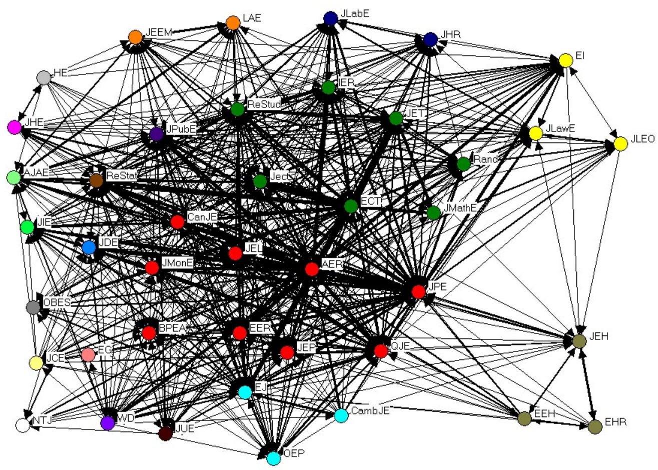

Figure 2: Community Structure in Economic Journal Citations (drawn using UCINET) - Same Colored Nodes are in the Same Community

It can be difficult to identify community structures without a methodology. The graph above makes it fairly clear how the communities are structured, and from looking at that graph one might get the false impression that one can intuite the communities simply from examining the network directly. To see why this is not the case, consider the following picture of the raw data, where line thickness is again an indication of relative citations, but where we have not colored or grouped nodes according to communities. From such a graph, it is essentially impossible to derive the community structure without a careful method.

6.6 Hypothesis Testing

The log likelihoods in the above community structures are very small, just as they should be, since there are a huge number of possible community structures. Thus, in order to make sense out of the results, it helps to do a careful statistical test to see if one community structure is significantly (in a statistical sense) more likely than another.

Here, we can do a likelihood ratio test, which has a firm foundation in the statistics literature on likelihood estimation techniques. The form for a likelihood ratio test in our setting is as follows.

Consider a null hypothesis H_0, which can be thought of as stated as some subset of all the

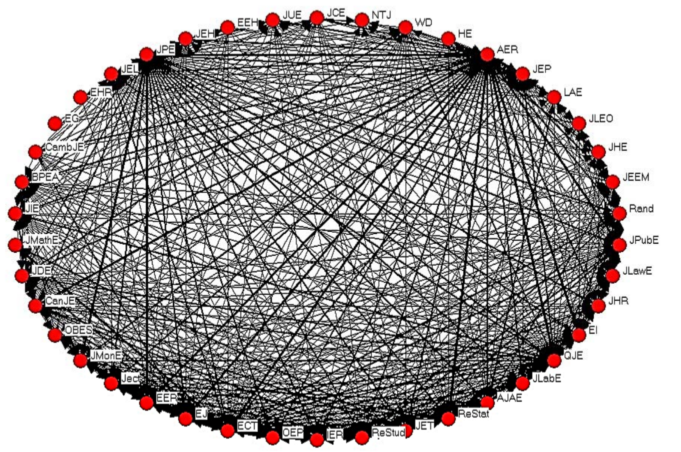

Figure 3: Raw Data of Economic Journal Citations (drawn using UCINET) - Line Thickness indicates Interaction Level

possible (pseudo) community structures together with some subset of possible (p_{in}, p_{out})'s. For instance, it might be that the null hypothesis is that the true community structure is the trivial one, where all nodes are grouped together[16] with some potential restrictions on p_{in} (and p_{out} is then irrelevant).

We can then test whatever null hypothesis we specify as follows. We estimate the likelihood of the best (pseudo) community structure and (p_{in}, p_{out}) under the null hypothesis. We then also estimate the likelihood of the best (pseudo) community structure and (p_{in}, p_{out}) without any restrictions. Let R be the ratio of these two likelihoods. Then, by standard likelihood ratio testing methods, (in the limit as the sizes grow) $-2\log(R)$ is distributed according to a χ^2 distribution with a number of degrees of freedom with is equal to the difference in the number of degrees of freedom of the full estimation problem and the number of degrees of freedom under the null hypothesis.

For example, if we find some "optimal" pseudo-community structure and estimated p_{in} and p_{out} on n nodes, then there are $2 + \frac{n(n-1)}{2}$ dimensions or degrees of freedom. The pseudo-community structure can be written as an $n \times n$ matrix of 0's and 1's, but with a

[16]Note that either extreme community structure, grouping all nodes together or all apart, are effectively the same. They all interact with each other in the same manner. Even though they are opposite extremes of the model, they are equivalent.

restriction that the diagonal be 1's and that the matrix be symmetric[17] (so this accounts for the $n(n-1)/2$), and then the additional 2 is for the p_{in} and p_{out}. Under a null hypothesis of the trivial partition, we estimate only p_{in}, and so there is just one degree of freedom. The difference is then $\frac{n(n-1)}{2}+1$, which is then the number of degrees of freedom for the χ^2 distribution. If we restrict our attention to only community structures, and not pseudo-community structures, then there are additional restrictions and so the number of degrees of freedom is even smaller.

Note that the mean and variance of a χ^2 distribution with d degrees of freedom are d and 2d, respectively; and so here the mean is $\frac{n(n-1)}{2}+1$ and variance here is n^2-n+2, and the standard deviation is then roughly n.

In the application to the economics journals, the degrees of freedom are 884-1=883. The log likelihood of the trivial partition is -546661 and of the optimal partition from the algorithm (Table 2) is -409725. Thus, -2 log(R) is 273872. This is thousands of standard deviations away from the mean (9816 standard deviations from the mean), and so the identified community structure is significantly more likely (in a statistical sense, with a p-value of essentially 0) than the trivial community structure; and we reject the null hypothesis.

If we test the optimal partition against the null hypothesis of the partition in table 3, we find over seven hundred standard deviations of difference. We can also test the optimal partition from table 4 against that in table 2 and find over two hundred standard deviations of difference.

We can also examine the pseudo-community numbers. The log likelihood of the optimal pseudo-community structure at its associate $p_in = 0.0053$ and $p_out = 0.00071$ is -394737.[18] As we know, this is a higher likelihood than each of the community structures since it is optimal among all pseudo community structures. This is statistically significant when tested against any of the community structures.

Thus, we see statistically significant differences between each of the community structures found, under the presumption of the model.

The distance between the optimal pseudo-community structure and the heuristic optimal community structure (table 4) is 102 "pairs", and its distance from the and the algorithm's best approximation is 88 "pairs" (and recall that there are 883 pairs in total). So the algorithm did find a community structure that was closer to the pseudo-community structure

[17] One can also work with pseudo-community structures where the matrix is not symmetric, so that i is put with j, but not vice versa. This allows any $n \times n$ matrix of 0's and 1's, with the diagonals being all ones, or $n^2 - n$ dimensions.

[18] At the $p_in = 0.0064$, $p_out = 0.00085$ of the approximate partition, the log likelihood of pseudo community is -396474.

than the heuristic best, but nevertheless the heuristic best is more likely. Thus, proximity is not directly related to likelihood, as we should expect.

7 Concluding Remarks

7.1 Summary

We have analyzed ranking community structures from network data based on (maximum) likelihood methods. The contribution here was in

(i) providing a model of how network data arises on which to base such methods,

(ii) demonstrating properties that characterize the likelihood method,

(iii) providing large sample properties of the model and likelihood method,

(iv) suggesting approaches to implementing it algorithmically, and

(iv) illustrating it in an application.

There are some interesting issues for further development.

7.2 Other Methods

Given the plethora of different methods for analyzing community structures, it is important to begin to systematically study their properties. Providing characterizations of different methods, as we have done for the likelihood ranking, would be very helpful in being able to compare methods.[19] In the past, comparisons across methods have usually been done simply by applying them and seeing which one seems to give a more subjectively sensible partition.

Along with this, it makes sense to see if one can rationalize some of the more popular previously proposed algorithms through some model. That is, what models of how network data is generated would justify previously studied methods, so that they are finding the true or most likely to be true community structure under some well-defined view of the world?

[19] One can also examine a Bayesian version of what we have done here. Likelihood estimation is a classical statistical approach when one does not have (or is unwilling to adopt) a prior distribution. It is easy to see how our approach can be adopted to do a Bayesian analysis, just replacing likelihood expressions with posterior probabilities.

7.3 Further Development of Algorithms for Implementing the Likelihood Ranking

As we have started from a model and derived conditions characterizing the optimal community structures, rather than starting from an algorithm, we face the task of finding algorithms that can search for approximately optimal structures. We used a specific approach based on pseudo community structures for the application above, and have presented some results that suggest approaches for finding algorithms. But there is still much more that can be done in terms of deriving algorithms that will find approximately optimal community structures quickly.

7.4 Heterogeneous Communities and Hierarchies

There are some obvious extensions that can be made to the model we have suggested here. The model we specified has a single p_{in} that indicates the probability of linkings between nodes in the same community, and a single p_{out} for the probability of linkings between nodes of different communities. Many applications may have more heterogeneity than this, where p_{in}'s and p_{out}'s vary across communities of nodes. Most notably, things might tend to be asymmetric. For instance, in the journal citation data, it is clear that some communities have much higher rates of being cited by other communities, than they do of citing other communities.

This suggests that a hierarchical model might be worth exploring, where there are different levels of communities, and the chance of lower level communities attaching to higher level communities would be higher than the reverse. Such a model is an easy extension of what we have done here, and then there is a direct analog in terms of the likelihood expressions. There is essentially no conceptual difference, only some extra parameters (more p_{in}'s and p_{out}'s) to estimate. For example, here is one variation. In addition a community structure, there is also a hierarchical structure among communities. There might be several communities in any stratum. Communities in lower strata interact with communities in higher strata at a higher rate than the reverse. A simple version would be one where the p_{out} depends on whether one is interacting with a higher strata, the same strata, or a lower strata. But one could also envision having the p_{out}'s depend on how different the strata are.

A general specification that includes the hierarchical model as a special case is one where each community has a different p_{out} for each other community that it interacts with. In order to keep the model well-specified, one needs some restrictions on the p_{in}'s and p_{out}'s, so that the world does not become one of each node being its own community. A natural restriction would be to have p_{in} be the same for all nodes interacting within any given community, and to be higher than each of the p_{out}'s that community has for any other community. But there are many others worth exploring.[20]

With any specification of such a model, our basic approach is still valid and our results on large samples, and the ideas behind using pseudo community structures in implementing the method also still work. Clearly, characterizing such variations on the method will require some modification of the properties, and the specifics of the implementation will require modifications of the implementation algorithm.

References

[1] Boorman, S.A. and H.C. White (1976) "Social Structure from Multiple Networks II: Role Structures," *American Journal of Sociology*, 81: 1384- 1446.

[2] Bron, C. and J. Kerbosch (1973) "Finding all cliques of an undirected graph," *Comm. ACM*, 16: 575-577.

[3] Clauset, A., C. Moore and M.E.J. Newman (2008) "Hierarchical Structure and the Prediction of Missing Links in Networks," *Nature*, 453:1, 98-101.

[4] Fiedler, M. (1973) "Algebraic connectivity of graphs" *Czech. Math. J.* 23: 298-305.

[5] Fienberg, S.E. and S.S. Wasserman (1981) "Categorical Data Analysis of Single Sociometric Relations," *Sociological Methodology*, 12, 156-192.

[6] Giada, L. and M. Marsili (2005) "Data clustering and noise undressing of correlation matrices," revision of arXiv:cond-mat/0101237 v1 16 Jan 2001.

[7] Girvan, M. and M. E. J. Newman, (2002) "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences* 99: 7821-7826.

[8] Handcock, M.S. and A.E. Raftery (2006) "Model-Based Clustering for Social Networks," mimeo: University of Washington.

[9] Holland, P.W. and S. Leinhardt (1977) "A Dynamic Model for Social Networks," *Journal of Mathematical Sociology*, 5, 5-20.

[20] Clauset, Moore and Newman (2008) present an interesting way in which a hierarchy might generate probabilities across communities.

[10] Holland, P.W., K.B. Laskey, and S. Leinhardt (1983) "Stochastic Blockmodels: First Steps," *Social Networks*, 5, 109–137.

[11] Kernighan B.W. and S. Lin (1970), "An efficient heuristic procedure for partitioning graphs," *Bell System Technical Journal* 49: 291-307.

[12] Lance, G.N. and W.T. Williams (1967) "A General Theory of Classificatory Sorting Strategies," *Computer Journal* 9:373-380.

[13] Lorrain F. and H.C. White (1971) "Structural Equivalence of Individuals in Social Networks," *Journal of Mathematical Sociology*, 1: 49-80.

[14] Newman, M.E.J. (2004) "Detecting Community Structure in Networks," *Physical Review E* 69, 066133.

[15] Newman, M.E.J. and E.A. Leicht (2007) "Models and Exploratory Analysis in Networks," *Proceedings of the National Academy of Sciences*, 104: 9564-9569.

[16] Pieters, F.G.M. and H. Baumgartner (2002) "Who talks to whom? Intra- and interdisciplinary communication of economics journals ," *Journal of Economic Literature*, 40(2):483-509.

[17] Pothen, A., H. Simon, and K.-P. Liou, (1990) "Partitioning sparse matrices with eigenvectors of graphs," *SIAM J. Matrix Anal. Appl.*, 11: 430-452.

[18] Radicchi, F., C. Castellano, F. Cecconi, V. Loreto, and D. Parisi (2003) "Defining and identifying communities in net- works," cond-mat/0309488.

[19] Schwartz, J.E. (1977) "An Examination of CONCOR and related methods for blocking sociometric data," in *Sociological Methodology 1977*, edited by D.R. Heise, Jossey-Basse: San Francisco, 255-282.

[20] Snijders, T.A.B. and K. Nowicki (1997) "Estimation and Prediction for Stochastic Block Models for Graphs with Latent Block Structure," *Journal of Classification* , 14, 75-100.

[21] Tyler, J.R., D.M. Wilkinson, and B.A. Huberman (2003) "Email as spectroscopy: Automated discovery of com- munity structure within organizations," In M. Huysman, E. Wenger, and V. Wulf (eds.), *Proceedings of the First International Conference on Communities and Technolo- gies*, Kluwer: Dordrecht.

[22] White, H.C., S.A. Boorman, and R.L. Breiger (1976) "Social Structure from Multiple Networks I: Blockmodels of roles and positions," *American Journal of Sociology*, 81: 730-779.

8 Appendix

8.1 Proof of Theorems ?? and ??.

First, it is easy to check that the properties hold for $\succeq_{s,g}^{L(p_{in},p_{out})}$, by direct inspection of **??**. Thus, in each case, we simply prove the converse; namely that any ranking satisfying the given properties must be a likelihood ranking for some p_{in} and p_{out}.

We start with the proof of Theorem **??**, as this is then used to prove Theorem **??**

Proof of Theorem ?? : The likelihood ranking $\succeq^L$ satisfies the properties as argued above.

For the converse, note that by internal differences, for any given ranking $\succeq_{\cdot,\cdot}$ satisfying the properties, there exists $H : I\!R^2 \to I\!R$ such that $\pi \succeq_{s,g} \pi'$ if and only if $H(D(g,\pi,\pi'), D(s,\pi,\pi')) \geq 0$.

Note that weak monotonicity implies that $H(w,0) > 0$ and $H(0,w) < 0$ whenever $w > 0$.

Let γ be defined by scaling. We now show that $H(z,y) > 0$ whenever $z - \frac{1}{\gamma}y > 0$, $H(z,y) < 0$ whenever $z - \frac{1}{\gamma}y < 0$, and $H(z,y) = 0$ whenever $z - \frac{1}{\gamma}y = 0$. Consider each of the three cases in turn.

Case 1. $z - \frac{1}{\gamma}y > 0$

By scaling, $H(z,y) > 0$ if and only if $H(z - \frac{1}{\gamma}y, 0) > 0$. Thus, since $H(w,0) > 0$ whenever $w > 0$ it follows that $H(z,y) > 0$ whenever $z - \frac{1}{\gamma}y \geq 0$.

Case 2. $z - \frac{1}{\gamma}y < 0$

By scaling, $H(z,y) > 0$ if and only if $H(0, y - \gamma z) > 0$. Thus, since $H(0,w) < 0$ whenever $w > 0$ it follows that $H(z,y) < 0$ whenever $z - \frac{1}{\gamma}y < 0$.

Case 3. $z - \frac{1}{\gamma}y = 0$

By scaling, $H(z,y) = 0$ if $H(0, y - \gamma z) = H(0,0) = 0$. Thus, since $H(0,0) = 0$ it follows that $H(z,y) = 0$ whenever $z - \frac{1}{\gamma}y = 0$.

Thus, we have shown that $\pi \succeq_{s,g} \pi'$ if and only if $\gamma D(g,\pi,\pi') \geq D(s,\pi,\pi')$. From (**??**) we know that this corresponds to a likelihood ranking. ∎

Proof of Theorem ?? : Again, a likelihood ranking $\succeq^L$ (p_{in}, p_{out}) satisfies the properties as argued above.

Given Theorem **??**, we need only show that a ranking $\succeq_{\cdot,\cdot}$ satisfying independence, neutrality, and scaling satisfies internal differences.

So, consider (g,s), (g',s'), π,π',π'',π''' such that

$$D(g,\pi,\pi') = D(g',\pi'',\pi''')$$

and

$$D(s,\pi,\pi') = D(s',\pi'',\pi''').$$

We need to show that $\pi \succeq_{s,g} \pi'$ if and only if $\pi'' \succeq_{s',g'} \pi'''$.

First, note that by independence, if $\pi \succeq_{s,g} \pi'$ the same is true regardless of the values of g_{ij} and s_{ij} such that either $ij \in In(\pi)\cap In(\pi')$ or $ij \in Out(\pi)\cap Out(\pi')$. So, without loss of generality suppose that $g_{ij}=0$ and $s_{ij}=0$ whenever either $ij \in In(\pi)\cap In(\pi')$ or $ij \in Out(\pi)\cap Out(\pi')$.

Next, by the second part of Neutrality if there is more than one link ij that is in $In(\pi)\cap Out(\pi')$, then we can equivalently consider a network g^1 such that $g^1_{kl}=0$ and $s^1_{kl}=0$, for all $kl \in In(\pi)\cap Out(\pi')$ such that $kl \neq ij$ and $g^1_{ij} = \sum_{kl\in In(\pi)\cap Out(\pi')} g_{kl}$ and $s^1_{ij} = \sum_{kl\in In(\pi)\cap Out(\pi')} s_{kl}$. Similarly, either $Out(\pi)\cap In(\pi')=\emptyset$ or we can find a link $hm \in Out(\pi)\cap In(\pi')$ and then set $g^1_{kl}=0$ and $s^1_{kl}=0$, for all $kl \in Out(\pi)\cap In(\pi')$ such that $kl \neq hm$ and $g^1_{hm} = \sum_{kl\in Out(\pi)\cap In(\pi')} g_{kl}$ and $s^1_{hm} = \sum_{kl\in Out(\pi)\cap In(\pi')} s_{kl}$.

So, $g^1_{kl}=s^1_{kl}=0$ for all kl except at most two links, $ij \in In(\pi)\cap Out(\pi')$ and $hm \in Out(\pi)\cap In(\pi')$.

Consider the case where $g^1_{ij} \geq g^1_{hm}$, as the other is analogous.

By the first part of neutrality, this is equivalent to a network g^2 where $g^2_{ij} = \sum_{kl\in In(\pi)\cap Out(\pi')} g_{kl} - \sum_{kl\in Out(\pi)\cap In(\pi')} g_{kl}$ and $g^2_{hm}=0$. Since $g_{ij}=0$ whenever either $ij \in In(\pi)\cap In(\pi')$ or $ij \in Out(\pi)\cap Out(\pi')$, it follows that

$$g^2_{ij} = D(g^2,\pi,\pi'), \tag{8}$$

and $g^2_{kl}=0$ for all $kl \neq ij$.

Case 1: $s^1_{ij} > \gamma g^2_{ij}$.

By scaling, we can consider s^3 and g^3 such that $s^3_{ij} = s^1_{ij} - \gamma g^2_{ij}$ and $g^3_{ij}=0$ and other entries are as before.

Case 1a: $s^1_{ij} - \gamma g^2_{ij} > s^1_{hm}$

By neutrality and weak monotonicity, it follows that $\pi \succ \pi'$. Note that the condition of 1a is equivalent to $D(s,\pi,\pi') > \gamma D(g,\pi,\pi')$.

Case 1b: $s^1_{ij} - \gamma g^2_{ij} < s^1_{hm}$

By neutrality and weak monotonicity, it follows that $\pi' \succ \pi$. Note that the condition of 1b is equivalent to $D(s,\pi,\pi') < \gamma D(g,\pi,\pi')$.

Case 2: $s^1_{ij} \leq \gamma g^2_{ij}$.

Again by the first part of neutrality, we can equivalently consider s^2 such that $s^2_{ij} = \gamma g^2_{ij}$ and $s^2_{hm} = s^1_{hm} + \gamma g^2_{ij} - s^1_{ij}$. Then by scaling, we can consider s^3 where $s^3_{ij} = g^3_{ij} = 0$. By weak monotonicity if follows that $\pi' \succ \pi$ provided $s^2_{hm} > 0$. Note that in this case, $s^2_{hm} = -D(s,\pi,\pi') + \gamma D(g,\pi,\pi')$. So we are in the case where $D(s,\pi,\pi') < \gamma D(g,\pi,\pi')$.

We are only left with a case where $s^2_{hm} = 0$, which corresponds to a case where $D(s,\pi,\pi') = \gamma D(g,\pi,\pi')$. Here, weak monotonicity implies indifference.

If we perform the same analysis for (g', s') we also find that $\pi' \succ \pi$ whenever $D(s',\pi,\pi') < \gamma D(g',\pi,\pi')$ $\pi \succ \pi'$ whenever $D(s',\pi,\pi') > \gamma D(g',\pi,\pi')$ and $\pi \sim \pi'$ when $D(s',\pi,\pi') = \gamma D(g',\pi,\pi')$.

Thus, if the D's are the same across (g,s) and (g',s'), then the ranking of π and π prime are the same. ∎

Proof of Proposition ??: Let $\bar{\pi} \in \Pi(N)$ be nondegenerate and be the unique optimal partition under the likelihood model given probabilities $p_{in}, p_{out} \in (0,1), p_{in} > p_{out}$. Let g, s are such that such that $\frac{g_{ij}}{s_{ij}} = p_{in}, \forall ij \in In(\pi)$, and $\frac{g_{ij}}{s_{ij}} = p_{out}, \forall ij \in Out(\pi)$.

To see that $\bar{\pi}$ satisfies consistent optimality is straightforward.

We have to prove that $\bar{\pi}$ is unique partition satisfying consistent optimality. We prove this by showing that for every $\pi \in \Pi(N), \pi \neq \bar{\pi}$, $\bar{\pi} \succ^{L(\hat{p}_{in}(\pi),\hat{p}_{out}(\pi))}_{g,s} \pi$, where $\hat{p}_{in}(\pi), \hat{p}_{out}(\pi)$ are the estimates based on (g,s) and from the partiation π.

So take $\pi \in \Pi(N)$, and let $p = \hat{p}_{in}(\pi), q = \hat{p}_{out}(\pi)$. Clearly, $p \leq p_{in}$, and $p = p_{in}$ if and only if $In(\pi) \subset In(\bar{\pi})$. Similarly, $q \geq p_{out}$, and $q = p_{out}$ if and only if $Out(\pi) \subset Out(\bar{\pi})$. At least one of these two must hold with inequality given if $\pi \neq \bar{\pi}$. Let $\bar{L}[.] = L(p_{in}, p_{out})[.]$ by $L[.] = L(p,q)[.]$, and $l[.] = \log(L[.])$, $\bar{l}[.] = \log(\bar{L}[.])$.

Consider first the case when $Out(\pi) \subset Out(\bar{\pi})$. Let $K = Out(\bar{\pi}) \cap In(\pi)$. Then,

$$
\begin{aligned}
l[\pi] - l[\bar{\pi}] = &\left(\sum_{ij\in K} g_{ij}\right)\log p + \left(\sum_{ij\in K} s_{ij} - \sum_{ij\in K} g_{ij}\right)\log(1-p) \\
&- \left(\sum_{ij\in K} g_{ij}\right)\log q - \left(\sum_{ij\in K} s_{ij} - \sum_{ij\in K} g_{ij}\right)\log(1-q).
\end{aligned}
$$

Given that $g_{ij} = q s_{ij}, \forall ij \in Out(\bar{\pi})$, we rewrite this expression as

$$
l[\pi] - l[\bar{\pi}] = \left(\sum_{ij\in K} s_{ij}\right)\left[q\log\left(\frac{p}{q}\right) + (1-q)\log\left(\frac{1-p}{1-q}\right)\right]
$$

Note that this expression is 0 when $p = q$ and is decreasing in p. Given that $p > q$, the proof follows. The argument for the the case when $In(\pi) \subset In(\bar{\pi})$ is similar.

If neither $Out(\pi) \subset Out(\bar{\pi})$ nor $In(\pi) \subset In(\bar{\pi})$, then we consider the following. Let $\tilde{\pi}$ be the coarsest partition that is finer than both π and $\bar{\pi}$. This can be thought of as splitting the communities of π into subcommunities based on which nodes in a given community are in a common community under $\bar{\pi}$. We then show that $l[\pi] < l[\tilde{\pi}] < l[\bar{\pi}]$. This follows from a similar argument as that above. ∎

Proof of Proposition ??: Given Proposition **??**, it is enough to show that there exists k such that $M(n) > kn(n-1)/2$.

First, it is straightforward to check that $k_1 p_{in} + k_2 > k_3 p_{in} + k_4$ and that $k_1 p_{out} + k_2 < k_3 p_{out} + k_4$.

Let x denote the minimum of $(k_1 p_{in} + k_2) - (k_3 p_{in} + k_4)$ and $(k_3 p_{out} + k_4) - (k_1 p_{out} + k_2)$.

The log-likelihood of the optimal partition $\widehat{\pi}$ under the expected g is

$$s\frac{n(n-1)}{2}\left[\left(k_1 p_{in} + k_2\right)\phi + \left(k_3 p_{out} + k_4\right)(1-\phi)\right] \tag{9}$$

where ϕ is the fraction of links in $In(\widehat{\pi})$.

The worst possible grouping of nodes would lead to a log-likelihood under the expected g of

$$s\frac{n(n-1)}{2}\left[\left(k_1 p_{out} + k_2\right)(1-\phi) + \left(k_3 p_{in} + k_4\right)\phi\right]. \tag{10}$$

The difference between (**??**) and (**??**) is at least

$$s\frac{n(n-1)}{2}x,$$

and thus $M(n) \geq s\frac{n(n-1)}{2}x$, as claimed. ∎

		Cites	Cited	Self-Cites	Size 2003
1	American Economic Review (AER)	2368	6817	694	176
2	American Journal of Agricultural Economics (AJAE)	1136	479	950	124
3	Brooking Papers of Economic Activity (BPEA)	374	694	89	5
4	Cambridge Journal of Economics (CJE)	717	36	162	44
5	Canadian Journal of Economics (CJE)	2059	187	163	45
6	Economic Geography (EG)	45	21	91	19
7	Econometrica (E)	1003	6537	713	61
8	Economic History Review (EHR)	257	154	354	9
9	Economic Inquiry (EI)	1433	210	122	50
10	Economic Journal (EJ)	2089	1604	318	75
11	European Economic Review (EER)	2018	624	185	52
12	Exploration of Economic History (EEH)	454	165	102	14
13	Health Economics (HE)	278	16	81	82
14	International Economic Review (IER)	1580	858	115	54
15	Journal of Comparitive Economics (JCE)	490	30	135	37
16	Journal of Development Economics (JDE)	1454	331	220	73
17	Journal of Econometrics (JE)	1735	1052	648	79
18	Journal of Economic History (JEH)	477	367	315	29
19	Journal of Economic Literature (JEL)	1209	889	93	21
20	Journal of Economic Perspectives (JEP)	1152	670	91	37
21	Journal of Economic Theory (JET)	2009	1677	907	105
22	Journal of Environmental Economics and Management (JEEM)	789	353	344	66
23	Journal of Health Economics (JHE)	408	210	139	53
24	Journal of Human Resources (JHR)	742	625	162	44
25	Journal of International Economics (JIE)	903	706	237	58
26	Journal of Labor Economics (JLE)	1009	350	100	32
27	Journal of Law and Economics (JLE2)	363	556	124	24
28	Journal of Law, Economics and Organization (JLEO)	397	108	115	20
29	Journal of Mathematical Economics (JME)	665	246	252	41
30	Journal of Monetary Economics (JME)	1188	1310	281	64
31	Journal of Political Economy (JPE)	1158	5605	359	42
32	Journal of Public Economics (JPE2)	1938	981	491	116
33	Journal of Urban Economics (JUE)	603	126	394	58
34	Land Economics (LAE)	743	177	210	39
35	National Tax Journal (NTJ)	318	134	228	47
36	Oxford Bulletin of Economics and Statistics (OBES)	705	246	56	40
37	Oxford Economic Papers (OEP)	1043	276	79	32
38	Quarterly Journal of Economics (QJE)	1005	3474	234	40
39	Rand (Bell) Journal of Economics (RJE)	1129	709	274	30
40	Review of Economic Studies (RES)	946	2448	158	37
41	Review of Economics and Statistics (RES2)	1949	1341	217	92
42	World Development (WD)	1182	121	749	120

Table 1. Summary of the data. The number of articles ("size 2003") are as reported by ISI web of science for 2003. The citation information is from Pieters and Baumgartner (2002).

American Economic Review (AER)
European Economic Review (EER)
Journal of Economic Literature (JEL)
Journal of Political Economy (JPE)
Review of Economic Studies (RES)
Economic Inquiry (EI)
Journal of Law and Economics (JLE2)
Journal of Law, Economics and Organization (JLEO)
Rand (Bell) Journal of Economics (RJE)
Economic History Review (EHR)
Exploration of Economic History (EEH)
Journal of Economic History (JEH)
Econometrica (E)
Economic Journal (EJ)
Journal of Econometrics (JE)
Canadian Journal of Economics (CJE)
Journal of International Economics (JIE)
Quarterly Journal of Economics (QJE)
Brooking Papers of Economic Activity (BPEA)
Cambridge Journal of Economics (CJE)
Journal of Economic Perspectives (JEP)
Journal of Human Resources (JHR)
Journal of Labor Economics (JLE)
Journal of Environmental Economics and Management (JEEM)
Land Economics (LAE)
Journal of Economic Theory (JET)
Journal of Mathematical Economics (JME)
World Development (WD)
Review of Economics and Statistics (RES2)
Oxford Economic Papers (OEP)
Oxford Bulletin of Economics and Statistics (OBES)
National Tax Journal (NTJ)
Journal of Urban Economics (JUE)
Journal of Public Economics (JPE2)
Journal of Monetary Economics (JME)
Journal of Health Economics (JHE)
Journal of Development Economics (JDE)
Journal of Comparitive Economics (JCE)
International Economic Review (IER)
Health Economics (HE)
Economic Geography (EG)
American Journal of Agricultural Economics (AJAE)

Table 2. The best partition found by approx. alg. with a grid search, for values of p_{in} ranging between 0.004 and 0.009 with the increments of 0.0002, and values of p_{out} ranging between 0.0005 and 0.001 with the increments of 0.0005. This partition was found at $p_{in} = 0.0064$, $p_{out} = 0.00085$, the estimates are $\hat{p}_{in} = 0.00634$, $\hat{p}_{out} = 0.00084$, and the log likelihood at these parameter values is -409725. Note that this partition approximately satisfies CO.

Canadian Journal of Economics (CJE)
Economic Journal (EJ)
European Economic Review (EER)
Journal of Comparitive Economics (JCE)
Journal of Development Economics (JDE)
Journal of International Economics (JIE)
Journal of Monetary Economics (JME)
Oxford Economic Papers (OEP)
Econometrica (E)
International Economic Review (IER)
Journal of Econometrics (JE)
Journal of Economic Theory (JET)
Journal of Mathematical Economics (JME)
Oxford Bulletin of Economics and Statistics (OBES)
Review of Economic Studies (RES)
American Economic Review (AER)
Brooking Papers of Economic Activity (BPEA)
Journal of Economic Literature (JEL)
Journal of Economic Perspectives (JEP)
Journal of Political Economy (JPE)
Quarterly Journal of Economics (QJE)
Review of Economics and Statistics (RES2)
Journal of Law and Economics (JLE2)
Journal of Law, Economics and Organization (JLEO)
Journal of Public Economics (JPE2)
Journal of Urban Economics (JUE)
National Tax Journal (NTJ)
Rand (Bell) Journal of Economics (RJE)
Cambridge Journal of Economics (CJE)
Economic Geography (EG)
Economic History Review (EHR)
Exploration of Economic History (EEH)
Journal of Economic History (JEH)
World Development (WD)
American Journal of Agricultural Economics (AJAE)
Health Economics (HE)
Journal of Environmental Economics and Management (JEEM)
Journal of Health Economics (JHE)
Land Economics (LAE)
Economic Inquiry (EI)
Journal of Human Resources (JHR)
Journal of Labor Economics (JLE)

Table 3. The partition identified by Pieters and Baumgartner. The estimated parameter values are $\hat{p}_{in} = 0.0034, \hat{p}_{out} = 0.00075$, and the log likelihood at these parameter values is -414253. For comparison, the log likelihood of this partition at $p_{in} = 0.0064, p_{out} = 0.00085$ is -420737.

American Economic Review (AER)
Brooking Papers of Economic Activity (BPEA)
Canadian Journal of Economics (CJE)
European Economic Review (EER)
Journal of Economic Literature (JEL)
Journal of Economic Perspectives (JEP)
Journal of Monetary Economics (JME)
Journal of Political Economy (JPE)
Quarterly Journal of Economics (QJE)
Econometrica (E)
International Economic Review (IER)
Journal of Econometrics (JE)
Journal of Economic Theory (JET)
Journal of Mathematical Economics (JME)
Rand (Bell) Journal of Economics (RJE)
Review of Economic Studies (RES)
Economic Inquiry (EI)
Journal of Law and Economics (JLE2)
Journal of Law, Economics and Organization (JLEO)
Economic History Review (EHR)
Exploration of Economic History (EEH)
Journal of Economic History (JEH)
Cambridge Journal of Economics (CJE)
Economic Journal (EJ)
Oxford Economic Papers (OEP)
Journal of Human Resources (JHR)
Journal of Labor Economics (JLE)
Journal of Environmental Economics and Management (JEEM)
Land Economics (LAE)
World Development (WD)
Review of Economics and Statistics (RES2)
Oxford Bulletin of Economics and Statistics (OBES)
National Tax Journal (NTJ)
Journal of Urban Economics (JUE)
Journal of Public Economics (JPE2)
Journal of International Economics (JIE)
Journal of Health Economics (JHE)
Journal of Development Economics (JDE)
Journal of Comparitive Economics (JCE)
Health Economics (HE)
Economic Geography (EG)
American Journal of Agricultural Economics (AJAE)

Table 4. The best partition found through heuristic search. This partition satisfies CO. The estimated values of p_{in} and p_{out} are $\hat{p}_{in} = 0.0053$, $\hat{p}_{out} = 0.00071$, and the log likelihood (-constant) at these parameter values is -405184.