

Network Structure and the Speed of Learning

Measuring Homophily Based on its Consequences*

Benjamin Golub[†]

Matthew O. Jackson[‡]

Draft: March 11, 2011

Forthcoming in the *Annals of Economics and Statistics*

Abstract

Homophily is the tendency of people to associate relatively more with those who are similar to them than with those who are not. In Golub and Jackson (2011), we introduced *degree-weighted homophily* (DWH), a new measure of this phenomenon, and showed that it gives a lower bound on the time it takes for a certain natural best-reply or learning process operating in a social network to converge. Here we show that, in important settings, the DWH convergence bound does substantially better than previous bounds based on the Cheeger inequality. We also develop a new complementary upper bound on convergence time, tightening the relationship between DWH and updating processes on networks. In doing so, we suggest that DWH is a natural homophily measure because it tightly tracks a key *consequence* of homophily – namely, slowdowns in updating processes.

Keywords: networks, learning, diffusion, homophily, friendships, social networks, random graphs, convergence, speed of learning, convergence rate

JEL Classification Numbers: D83, D85, I21, J15, Z13

*Jackson gratefully acknowledges financial support from the NSF under grants SES-0647867 and SES-0961481. Golub gratefully acknowledges financial support from an NSF Graduate Research Fellowship, as well as the Clum, Holt, and Jaedicke fellowships at the Stanford Graduate School of Business. We thank Christoph Kuzmics, Carlos Lever, and Amin Saberi for comments on presentations. This article includes material that originally appeared in the working paper “How Homophily Affects Diffusion and Learning in Networks” (arXiv:0811.4013v2).

[†]Graduate School of Business, Stanford University. Email: bgolub@stanford.edu, <http://www.stanford.edu/~bgolub/>.

[‡]Department of Economics, Stanford University and external faculty of the Santa Fe Institute. Email: jacksonm@stanford.edu, <http://www.stanford.edu/~jacksonm/>

1 Introduction

Homophily is the robust tendency of people to associate more with those who are similar to them on some dimensions than with those who are not. This tendency is seen when similarity is measured based on a variety of different characteristics, including ethnicity, age, religion, and occupation — as documented in a large literature in sociology (see McPherson, Smith-Lovin and Cook (2001) for a survey). Homophily produces a signature in the structure of a network: a pattern of thick concentrations of links within groups of individuals who have shared characteristics and sparser connections between such groups.

Homophily is more than a statistical curiosity about networks. Intuition suggests that the structural features caused in networks by homophily have important consequences for processes of real economic significance, such as the travel of news in a society or access to information about jobs. The segregation patterns in a network would mean that there would tend to be relatively more communication within groups of similar individuals than across groups. Although there is some work that documents the impact of homophily on behavior and outcomes (e.g., see Jackson (2007) and Golub and Jackson (2011)), we are far from a general understanding of homophily’s consequences, and furthering that understanding is the goal of this paper.

There is a related conceptual question. While it is intuitively clear what homophily basically is, it is not obvious how best to measure it. Social networks are inherently complex and multidimensional objects, and so capturing homophily concisely is a challenge, especially in settings with a great deal of heterogeneity in the sizes and compositions of various groups, and also in the traits that distinguish them. There are several measures that have been proposed in the literature, all seeking to capture different aspects of homophily (e.g., Coleman (1985), Echenique and Fryer (2007)). While these approaches offer definitions that capture certain intuitively appealing aspects of homophily (Echenique and Fryer do this in a precise axiomatic framework), it is not entirely obvious whether those are substantively the important dimensions.

Our approach seeks to provide foundations for measuring homophily based on its *consequences* for a process of interest. More specifically, we seek to identify the measure of inter-versus intra-group connectivity that affects the speed of a learning process. We are not aware of work previous to Golub and Jackson (2011) taking a process-based approach to measuring homophily. In this article, we continue the study begun in that paper by elaborating on and strengthening some of its results about a new homophily measure.

As mentioned above, we analyze how homophily — through the segregation patterns it induces — influences a particular process of learning taking place on a network. The learning model ties down exactly how network structure impacts outcomes, and confirms the intuition that learning takes longer in networks with homophily. Moreover, the formulas that describe

that relationship reveal how one should measure the structural consequences of homophily if one is interested in how it affects learning in that context.

The learning process we focus on is the classic DeGroot model of learning by repeated averaging of neighbors' signals of French (1956), Harary (1959), and DeGroot (1974). This model of learning has been used in sociology, statistics, and economics, and has been motivated as boundedly rational aggregation of dispersed information (DeMarzo, Vayanos, and Zwiebel (2003)). This process, under mild conditions, leads a society to converge to a consensus opinion after a sufficiently long period of communication.

It turns out that homophily dramatically slows down this type of learning. Moreover, the extent to which it does so can be quantified explicitly in terms of simple formulas that involve the network's structure. To do this, in Golub and Jackson (2011), we introduced a measure of network segregation called *degree-weighted homophily* (DWH) and showed that it provided a lower bound on the time to consensus. The DWH of a network can be written as a simple and interpretable formula in terms of network structure – in particular, the relative quantities of within-group and across-group links, appropriately weighted by the degrees of the agents participating in those links. This explicit approach contrasts with well-known results, which we discuss below, that pin down the convergence of this learning process in terms of certain eigenvalues of an interaction matrix. Eigenvalue measures of segregation are defined implicitly as solutions of a certain equation have some intuition behind them, but it is hard to know exactly what features of the network determine the relevant eigenvalue. With degree-weighted homophily, on the other hand, it is very clear what goes into the computation. For example, an agent's degree (number of connection) matters for his or her contributions to homophily. Perhaps surprisingly, the homophily of agents with *low* degree matters more than that of agents with *high* degree.

This explicit approach to characterizing the convergence of the learning process in terms of network structure relates to a literature in mathematics on how segregation affects random walks on networks (Diaconis and Stroock, 1991; Chung et al., 2004; Montenegro and Tetali, 2006). In this article, we focus on the quality of the DWH bound and its relationship to the previous state of the art. First, we show that the bound is asymptotically tight — that is, in certain cases it provides an exact characterization of convergence time rather than just one bound. Second, it is better than existing bounds based on the Cheeger inequality in an important class of random networks. Finally, we give a new result toward a complementary *upper* bound on convergence time in terms of DWH, with the goal of relating the learning process to homophily as closely as possible. We find that DWH is more closely related to the convergence speed of the linear updating process than previously known. The fact that DWH tightly tracks a key *consequence* of homophily – namely, slowdowns in updating processes – motivates its relevance as a useful homophily measure.

2 Model, Definitions, and Basic Results

2.1 Networks

We begin by reviewing basic network definitions.¹ Given a set of n nodes $N = \{1, \dots, n\}$, a network is represented via its adjacency matrix: a symmetric n -by- n matrix \mathbf{A} with entries in $\{0, 1\}$. The interpretation is that $A_{ij} = A_{ji} = 1$ indicates that nodes i and j are linked, and we restrict attention to undirected networks. Let $d_i(\mathbf{A}) = \sum_{j=1}^n A_{ij}$ denote the *degree* of node i , which is the number of agents i is linked to. Let $D(\mathbf{A}) = \sum_{i=1}^n d_i$ be the sum of degrees in the network. A *walk* in \mathbf{A} is a sequence of nodes i_1, i_2, \dots, i_K , not necessarily distinct, such that $A_{i_k i_{k+1}} > 0$ for each $k \in \{1, \dots, K-1\}$. A *cycle* is a walk i_1, i_2, \dots, i_k such that $i_1 = i_k$. The *length* of a cycle with K (not necessarily distinct) entries is defined to be $k-1$. A cycle is *simple* if the only node appearing twice in the sequence is the starting (and ending) node. The matrix \mathbf{A} is *connected* (also known as *irreducible*) if there is path in \mathbf{A} from any node to any other node. A matrix is *aperiodic* if the greatest common divisor of the lengths of all simple cycles is 1. In the undirected case where $A_{ij} = A_{ji}$, the matrix \mathbf{A} is aperiodic if and only if it is not possible to divide the nodes into two nonempty disjoint subsets so that links only go from one subset to the other (i.e., if and only if it is not bipartite).

2.2 The Learning Process

2.2.1 A Version of the DeGroot Model

Given a network \mathbf{A} , let $\mathbf{T}(\mathbf{A})$ be defined by $T_{ij}(\mathbf{A}) = A_{ij}/d_i$. Beginning with some initial belief vector $\mathbf{b}(0) \in [0, 1]^n$, let

$$\mathbf{b}(t) = \mathbf{T}(\mathbf{A})\mathbf{b}(t-1)$$

for all $t \geq 1$. That is, agents form today's beliefs by taking the average of neighbors' beliefs yesterday, where an agent can be his own neighbor. It is immediate that then

$$\mathbf{b}(t) = \mathbf{T}(\mathbf{A})^t \mathbf{b}(0).$$

The DeGroot model captures a boundedly rational updating process such that agents repeatedly form new beliefs (or behaviors) by averaging their neighbors' decisions in the past period. This process has many nice properties and can even lead agents to efficiently

¹Our definitions here follow closely the ones in Golub and Jackson (2010) and Golub and Jackson (2011). Indeed, all the basic notions in Section 2 of this paper are defined there, but we repeat the definitions here to keep the exposition as self-contained as possible. See those papers and also Jackson (2008), especially Chapters 2 and 8, for more detailed discussions of the basic concepts.

aggregate the signals in the society, despite its simplicity (e.g., see Golub and Jackson (2010) for more discussion on these points). This process also can be interpreted as choosing an action as a (myopic) best reply in a case where agents wish to match the average behavior of their neighbors, as discussed in Golub and Jackson (2011).

2.2.2 Convergence

The network \mathbf{A} is said to *converge to a consensus* if, for any $\mathbf{b}(0) \in [0, 1]^n$, there is some b so that $\lim_{t \rightarrow \infty} b_i(t) = b$ for every i .

A result from Markov chain theory can be adapted to this setting to conclude that this process converges to a consensus (for any initial beliefs) if and only if two conditions hold. The first condition is that the network \mathbf{A} is connected. This is an obvious condition, as otherwise differences can remain between separate components. The second is that the network is aperiodic, or — in this setting, equivalently — not bipartite. This prevents a situation in which beliefs can oscillate forever. Aperiodicity is a very weak condition, and it suffices for at least one agent i to be linked to him or herself, so that $A_{ii} = 1$ (which can be interpreted as placing some weight on his or her own past beliefs or behavior in the updating, a quite natural case). The result is stated formally as follows, and the (standard) proof follows directly from results in Golub and Jackson (2010).

PROPOSITION 1. *The network \mathbf{A} converges to a consensus if and only if it is strongly connected and aperiodic.² The steady state beliefs are a weighted average of initial beliefs equal to $\sum_i \frac{d_i(\mathbf{A})}{D(\mathbf{A})} b_i(0)$.*

2.2.3 The Speed of Convergence

We are primarily concerned with the speed of learning — how quickly do beliefs converge to their limit when they do converge.

To measure this, we study how many periods it takes for beliefs to get close to their limiting values. The measure of deviation from consensus (which is necessary to define “close”) has a simple interpretation. Imagine the agents talking to each other at each period. At each moment in time, there are two messages sent across each link. Let $\mathbf{m}(t)$ be this vector of messages for some ordering of the directed links. We define the distance from consensus at time t to be the root mean square distance of $\mathbf{m}(t)$ from its limit $\mathbf{m}(\infty)$.

Mathematically, we write this in a standard way. First, given two vectors \mathbf{v} and \mathbf{u} , as well as a vector \mathbf{w} of weights, let $\|\mathbf{v} - \mathbf{u}\|_{\mathbf{w}} = [\sum_i w_i (v_i - u_i)^2]^{1/2}$. The distance of beliefs

²Given the symmetry of links in our setting here, connectedness and strong connectedness are equivalent. Aperiodicity is still required in addition, as connectedness and symmetry together still do not imply that there exists an odd-length cycle.

at time t from consensus is then $\|\mathbf{T}(\mathbf{A})^t \mathbf{b} - \mathbf{T}(\mathbf{A})^\infty \mathbf{b}\|_{\mathbf{s}(\mathbf{A})}$, where we use the weights $\mathbf{s}(\mathbf{A})$ defined by $s_i(\mathbf{A}) = \frac{d_i(\mathbf{A})}{D(\mathbf{A})}$. This is equivalent to the “messages” interpretation because an agent with degree $d_i(\mathbf{A})$ sends a share of the messages in the network given exactly by his or her degree divided by the sum of degrees; this is related to his “influence” on final beliefs in the formula of Proposition 1 above.

The consensus time is defined as the time it takes for the distance of beliefs from consensus to get below ε .

DEFINITION 1. The *consensus time* to $\varepsilon > 0$ of a connected network \mathbf{A} is

$$\text{CT}(\varepsilon; \mathbf{A}) = \sup_{\mathbf{b} \in [0,1]^n} \min\{t : \|\mathbf{T}(\mathbf{A})^t \mathbf{b} - \mathbf{T}(\mathbf{A})^\infty \mathbf{b}\|_{\mathbf{s}(\mathbf{A})} < \varepsilon\},$$

where $\mathbf{s}(\mathbf{A}) = \frac{d_i(\mathbf{A})}{D(\mathbf{A})}$.

We consider different potential starting belief vectors \mathbf{b} because some heterogeneity of beliefs is necessary for convergence not to be instant. Taking the supremum in this way is standard in defining convergence times³ and can be viewed as a “worst case” measure.

2.2.4 The Speed of Convergence and Second Eigenvalues

As is well-known, the speed of convergence of a linear process like this is closely related to the second eigenvalue of the updating matrix. The following lemma is the appropriate version of this result in our setting, and after stating it, we focus only on the second eigenvalue, since it can always be related back to consensus time using this result.

LEMMA 1 (Golub and Jackson (2011), Lemma 2). *Let \mathbf{A} be connected, $\lambda_2(\mathbf{T}(\mathbf{A}))$ be the second largest eigenvalue in magnitude of $\mathbf{T}(\mathbf{A})$, and $\underline{s} := \min_i d_i(\mathbf{A})/D(\mathbf{A})$ be the minimum relative degree. If $\lambda_2(\mathbf{T}(\mathbf{A})) \neq 0$, then for any $0 < \varepsilon \leq 1$:*

$$\left\lceil \frac{\log(1/(2\varepsilon)) - \log(1/\underline{s}^{1/2})}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}))|)} \right\rceil \leq \text{CT}(\varepsilon; \mathbf{A}) \leq \left\lceil \frac{\log(1/\varepsilon)}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}))|)} \right\rceil.$$

If $\lambda_2(\mathbf{T}) = 0$, then $\text{CT}(\varepsilon; \mathbf{A}) = 1$ for every $0 < \varepsilon < 1$.

From now on, we will not focus on consensus times but rather on second eigenvalues. Proposition 4 of Golub and Jackson (2011) shows that under mild restrictions on large-scale network structure, the relationship between the two quantities is even tighter than the one obtained in this lemma, and thus one can easily map results about eigenvalues into precise ones about consensus times.

³For background on mixing times of Markov processes, see Montenegro and Tetali (2006).

2.3 Degree-Weighted Homophily

2.3.1 Definition

The main focus of the paper is characterizing the speed of convergence, as introduced in the last section, in terms of homophily. The particular measure of homophily that we use is called *degree-weighted homophily*. In this section, we briefly recall the definition of degree-weighted homophily from Golub and Jackson (2011); more detail can be found in Section 4 of that paper.

A preliminary notion is that of the weight between two groups.

DEFINITION 2. Given $\mathbf{T} = \mathbf{T}(\mathbf{A})$ and two subsets of nodes, $B, C \subseteq N$, let

$$W_{B,C} = \frac{1}{|B||C|} \sum_{\substack{(i,j) \in B \times C \\ A_{ij}=1}} d_i^{-1} d_j^{-1}.$$

The summation takes into account the edges of the network, weighting an edge in inverse proportion to the product of degrees of the nodes it touches. This summation is divided by the product of the sizes of the groups involved.

DEFINITION 3. Given any $\emptyset \subsetneq M \subsetneq N$, let the *degree-weighted homophily* of the network \mathbf{A} relative to M be defined by

$$\text{DWH}(M; \mathbf{A}) = \frac{W_{M,M} + W_{M^c, M^c} - 2W_{M, M^c}}{\frac{1}{|M|^2} \sum_{i \in M} \frac{1}{d_i(\mathbf{A})} + \frac{1}{|M^c|^2} \sum_{i \in M^c} \frac{1}{d_i(\mathbf{A})}}, \quad (1)$$

where the W 's are computed relative to $\mathbf{T}(\mathbf{A})$.

It is clear that links within the group M or its complement M^c increase the degree weighted homophily and links between the two groups decrease it. The term in the denominator is a normalizing value which guarantees⁴ that this quantity is always between -1 and 1 .

For any adjacency matrix \mathbf{A} , let

$$\text{DWH}(\mathbf{A}) = \max_{\emptyset \subsetneq M \subsetneq N} |\text{DWH}(M; \mathbf{A})|.$$

Thus, the degree weighted homophily of a given network is the maximum level of degree homophily across all the different possible splits of the network.

⁴This can be verified by using the expression of DWH as a quadratic form in the proof of Lemma 3 in Golub and Jackson (2011) and then noting that the spectral norm of the matrix $\mathbf{T}(\mathbf{A})$ is 1.

2.3.2 A Remark on Computation

The fact that quantity $DWH(\mathbf{A})$ is defined as a maximum over a potentially very large number of splits may raise questions about how practical it is to compute. Certainly, the “naive algorithm” of iterating over all possible splits would be very demanding. However, efficient ways to solve related problems have been developed in computer science (e.g. the Edmonds-Karp algorithm for the maximum flow/minimum cut problem) and similar approaches may be effective here. We have not explored this issue but believe it may be an interesting avenue for further inquiry.

Perhaps more importantly, using the formula with *any* particular choice of M provides a ready lower bound on DWH (and hence, as we show below, on the second eigenvalue and consensus time). Since we will often have a good idea of where the fault lines in the network lie based on knowledge about the types of agents in the network, there may be a shortcut to finding a good M . See also the comments in the next remark; more formal results along these lines can be found below in Section 3.

2.3.3 A Remark on Interpretation

As is clear from the definition, the DWH measure depends only on the network structure and does not take into account any information about the *types* of agents inhabiting it. In particular, consider the following thought experiment. Take a network having two types (e.g. races) of agents, with each agent having many connections to same-type agents and few connections to those of different types. This will be a typical segregated, homophilous network: there is homophily in the sense that agents are predominantly linked to those of the same type, and also segregation in the sense that there is a prominent “split” in the graph between two groups – a split that can be seen even if the type information is invisible. Now leave the network fixed and scramble the type labels randomly, so that each agent’s neighborhood has roughly the same number of same-type and different-type agents. After this, the network does not have homophily in the sense of people being predominantly linked to those of the same type. But it will still have a high DWH, because there is the same “split” in the network that there was before (since we did not change the graph structure in the label-scrambling). Thus, DWH captures the signature of homophily in the network structure; in principle, the underlying “types” of agents in the network could be anything. Nevertheless, we call our measure degree-weighted *homophily* (as opposed to degree weighted separation or something similar) because we believe that, in most economic or social settings, the structural split captured by DWH will indeed track homophily based on some underlying characteristic.

2.3.4 Bounding Speed of Convergence using DWH

The following theorem, in combination with Lemma 1 above, shows that DWH provides a lower bound on consensus time. We quote it because the main focus of the paper will be on investigating the quality of this bound and on providing a matching upper bound.

THEOREM 1 (Golub and Jackson (2011), Lemma 3). *If \mathbf{A} is connected, then*

$$|\lambda_2(\mathbf{T}(\mathbf{A}))| \geq |\text{DWH}(\mathbf{A})|.$$

3 The Quality of the DWH Bound

It is natural to ask how good the bound in Theorem 1 is. This section discusses this question, both heuristically and formally.

First, we show that the DWH bound is asymptotically tight by exhibiting a random network model in which the bound is asymptotically the best possible. Next, we compare the DWH bound with the famous Cheeger inequality and find that the DWH bound, in some important cases, is significantly tighter. Third, we show that the relationship between DWH and the consensus time uncovered in Theorem 1 can be strengthened with a complementary upper bound.

For this analysis, we work with a random graph model which allows us to naturally generate large graphs having a certain kind of structure. First we informally discuss a special case, and then we state the formal model more generally in Section 3.4 below.

3.1 A Simple Tightness Result

The set, N , of nodes is partitioned into two nonempty subsets, N_1 and N_2 . The matrix \mathbf{A} is formed as follows. Edges between nodes both in the same set N_i are created with probability p_s (s for “same”) and edges between subsets are created with probability p_d (d for “different”). All these zero/one edge random variables are independent of one another. From this one can compute the expected overall link density:

$$p = \frac{\left(\binom{|N_1|}{2} + \binom{|N_2|}{2} \right) p_s + |N_1||N_2|p_d}{\binom{n}{2}}.$$

When $p_s = p_d = p$, this is the Erdős-Rényi random graph with link probability p . Write $f_1 = |N_1|/n$ and $f_2 = 1 - f_1$. We will be considering asymptotics where n tends to infinity and assuming that f_1 and f_2 tend to limits not equal to 0 or 1.

We now show that when \mathbf{A} is a random graph of this type and the minimum expected degree grows at least as fast as $\log^2 n$, then $|\text{DWH}(\mathbf{A})|$ approaches λ_2 with high probability as n grows large.

PROPOSITION 2. *Consider the random graph model discussed above. Then, as long as the ratio $np/\log^2 n$ tends to ∞ , it follows that*

$$\text{plim}_{n \rightarrow \infty} \lambda_2(\mathbf{T}(\mathbf{A})) = \text{plim}_{n \rightarrow \infty} \text{DWH}(\mathbf{A}).$$

3.2 Heuristic Illustrations

We now give some heuristic calculations and illustrations to convey some of the intuition of the above tightness result.

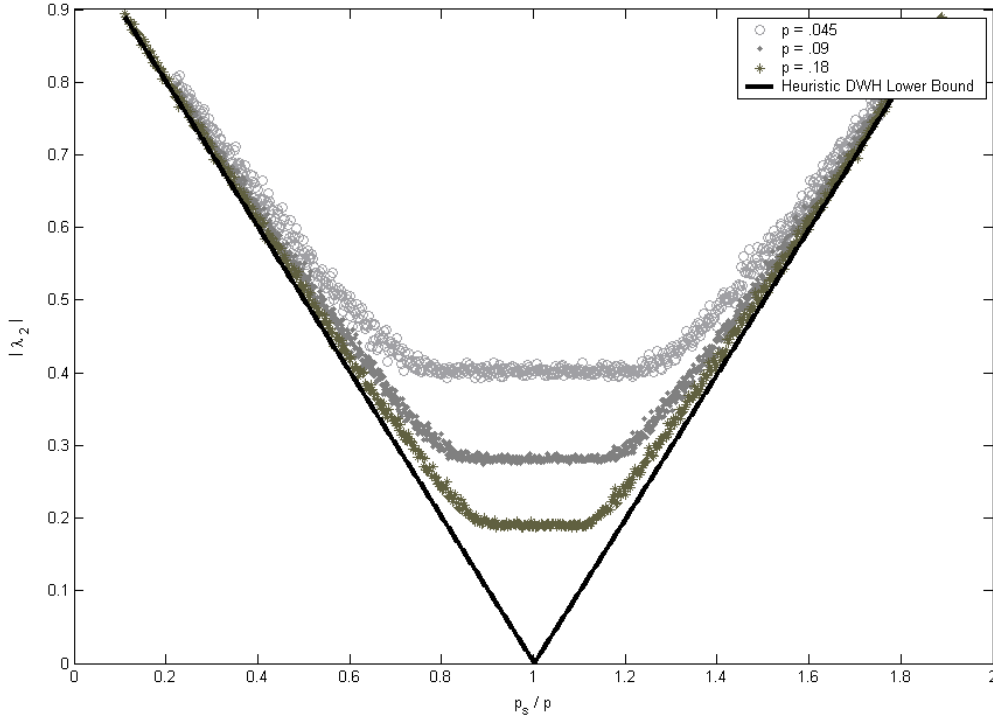


Figure 1: Our heuristic bound compared to the second eigenvalues of homophilous two-type random graphs with two equal-sized groups of 250 nodes each. The three sets of data correspond to $p = .045$, $p = 0.09$, and $p = 0.18$. For each of these, p_s was varied while p was held fixed.

For the heuristic calculation, let us pretend that all vertices in N_1 have the same degree, namely the expected degree of such vertices, and similarly for N_2 . Moreover, we assume that

the numbers of edges that enter the computation of DWH are all equal to their expectations.⁵

Example 1: Types of Equal Sizes First, consider the case where $|N_1| = |N_2|$, so that $f_1 = f_2 = 1/2$. If we simplify (1) under these heuristic assumptions and use the fact that p is very close to the average of p_s and p_d , we find

$$|\lambda_2| \gtrsim \frac{p_s}{p} - 1.$$

In Figure 1, we show the results of experiments in which we hold p fixed and vary p_s (correspondingly changing p_d). The experimental second eigenvalues are compared with the DWH bound.

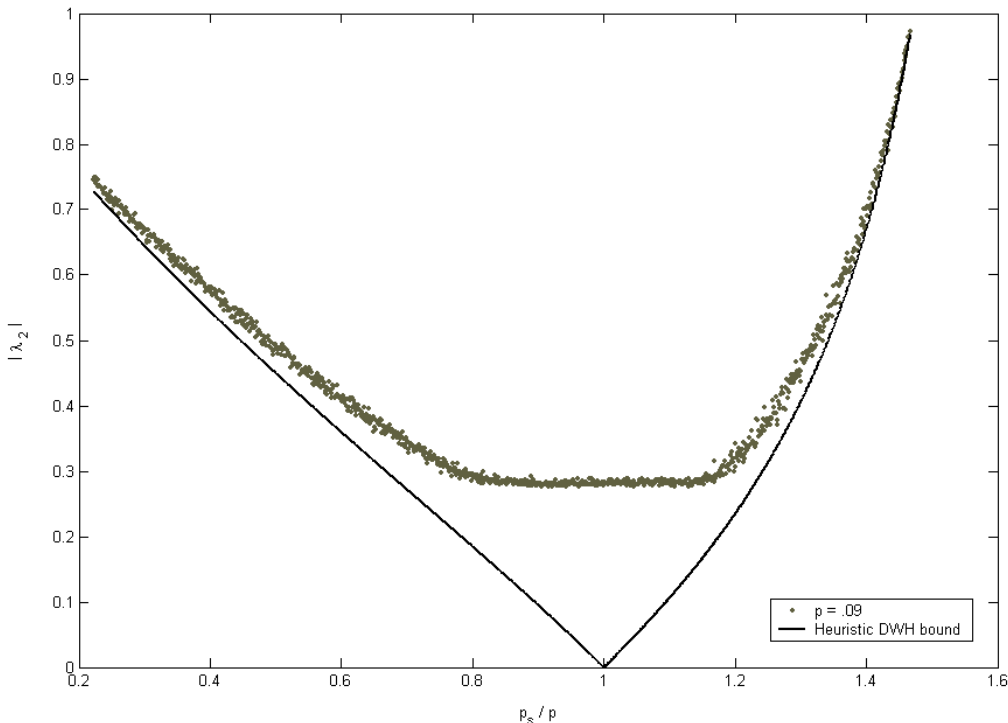


Figure 2: Our heuristic bound compared to the second eigenvalues of a homophilous two-type random graph with groups of 400 and 100 nodes. Here, p_s was varied while $p = 0.09$ was held fixed.

Example 2: Types of Unequal Sizes A similar calculation can be done in the case where group sizes are unequal. Suppose N_1 makes up a fraction f_1 of the population and N_2

⁵These replacements of realizations by expectations are not as unrealistic as they may first seem; see Lemma 4 of Golub and Jackson (2011).

makes up a fraction f_2 of the population. Then, letting $\delta_1 = f_1 p_s + f_2 p_d$ and $\delta_2 = f_2 p_s + f_1 p_d$ be normalized versions of the degrees of the two types of vertices yields a heuristic bound of

$$|\lambda_2| \gtrsim \frac{|p_s/\delta_1^2 + p_s/\delta_2^2 - 2p_d/(\delta_1\delta_2)|}{1/(f_1\delta_1) + 1/(f_2\delta_2)}.$$

In Figure 2 we plot the second eigenvalues and this heuristic bound for types of unequal sizes.

3.3 Comparison with the Cheeger Inequality

A famous result whose techniques date back to Polya and Szego (1951), which is commonly attributed in its continuous version to Cheeger (1970) and was developed in its discrete, graph-theoretic version by Alon and Milman (1985) states that, if we define $\text{Vol}(M) = \sum_{i \in M} d_i$ then

$$|\lambda_2| \geq 1 - \min_{\emptyset \subsetneq M \subsetneq N} \frac{2 \sum_{\substack{i \in M \\ j \notin M}} A_{ij}}{\min\{\text{Vol}(M), \text{Vol}(M^c)\}}.$$

Rewriting this as

$$|\lambda_2| \geq \max_{\emptyset \subsetneq M \subsetneq N} \frac{\min\{\text{Vol}(M), \text{Vol}(M^c)\} - 2 \sum_{\substack{i \in M \\ j \notin M}} A_{ij}}{\min\{\text{Vol}(M), \text{Vol}(M^c)\}}.$$

we see some resemblance to (1). However, it turns out that, at least in some cases, our bound is much stronger.

Continuing in the homophilous random graph setting with M and M^c equal to N_1 and N_2 from above, and continuing with the types of approximations used in the previous example, we find:

$$\text{Vol}(M) \approx n^2 f_1 (f_1 p_s + f_2 p_d)$$

and

$$\text{Vol}(M^c) \approx n^2 f_2 (f_2 p_s + f_1 p_d)$$

while

$$2 \sum_{\substack{i \in M \\ j \notin M}} A_{ij} \approx 2n^2 f_1 f_2 p_d.$$

Putting these together allows us to estimate the Cheeger bound in the homophilous Poisson random graph. (We are assuming that the best cut for the Cheeger inequality will be the one built into the random graph structure.) The result is plotted in Figure 3.

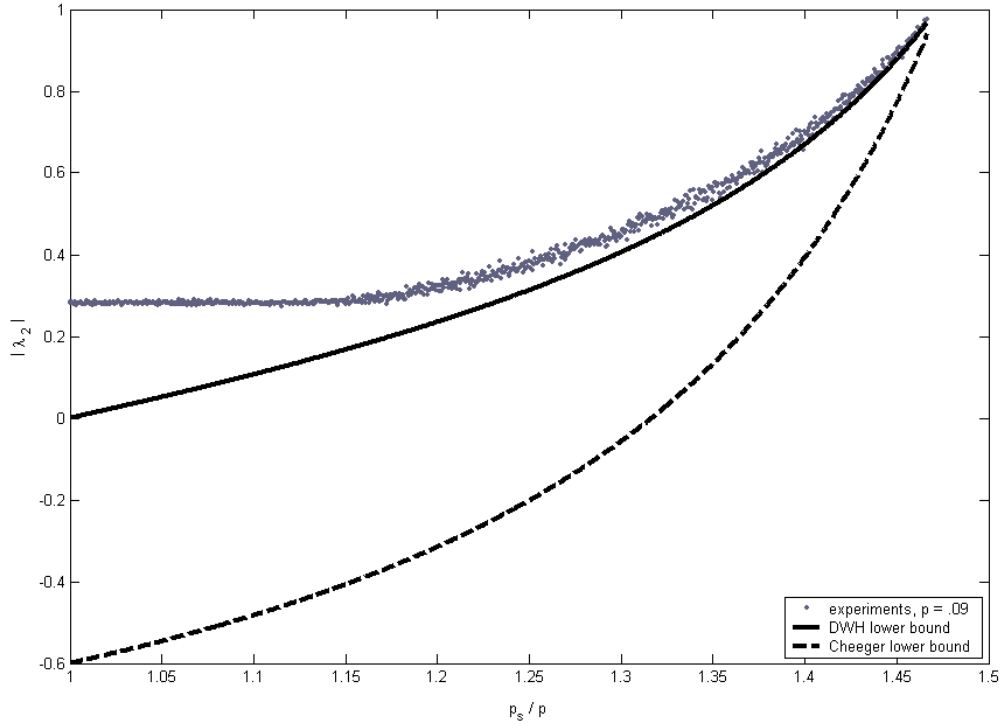


Figure 3: Our DWH bound compared to the Cheeger inequality. Here, as in Figure 2, $p = .09$ is held fixed as p_s is varied. Only the regime $p_s/p \geq 1$ is plotted because the Cheeger inequality is trivial for $p_s/p < 1$.

3.4 Multi-Type Random Graphs and a Matching Upper Bound on the Second Eigenvalue in Terms of DWH

We now work with a generalization of the random-graph model discussed above. A *multi-type random network* is defined by the partition into types $\mathbf{n} = (N_1, \dots, N_m)$ together with a symmetric m -by- m matrix \mathbf{P} , whose entries in $[0, 1]$ describe the probabilities of links between various types. The resulting random network is captured via its adjacency matrix, which is denoted $\mathbf{A}(\mathbf{P}, \mathbf{n})$ and is random. In particular, $\mathbf{A}(\mathbf{P}, \mathbf{n})$ is built by letting the entries A_{ij} with $i > j$ be independent Bernoulli random variables that take on a value of 1 with probability $P_{k\ell}$ if $i \in N_k$ and $j \in N_\ell$. That is, the entry $P_{k\ell}$ is the probability that an agent of type k links to an agent of type ℓ . We fill in the remaining entries of \mathbf{A} by symmetry: $A_{ij} = A_{ji}$. We denote by n the total number of agents in the network, $\sum_k |N_k|$.

Now let us define group-level analogs of links and degrees. Set $B_{k\ell}(\mathbf{P}, \mathbf{n}) = n_k n_\ell P_{k\ell}$. This is the expected number of links group k has to group ℓ . Define $d_k(\mathbf{B}) = \sum_\ell B_{k\ell}$. This is the expected number of links that group k has – a group-level analogue of degree.

One of the main results of Golub and Jackson (2011), Theorem 2, is that the second

eigenvalue of $\mathbf{T}(\mathbf{A}(\mathbf{P}, \mathbf{n}))$ converges to that of $\mathbf{T}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$, a much smaller matrix, as society grows large, as long as the minimum expected degree $d_{\min}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$ grows⁶ faster than $\log^2 n$.

In this section, we show that instead of computing the second eigenvalue of $\mathbf{T}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$, we can instead focus on the DWH of $\mathbf{B}(\mathbf{P}, \mathbf{n})$, once again reducing an abstract eigenvalue invariant into a concrete structural quantity. We already know from Lemma 1 that the DWH of \mathbf{B} furnishes a lower bound on its second eigenvalue. Here we show that, with some adjustment factors, it also imposes an upper bound. For this result, we impose a natural homophily condition whereby groups place most of their weight inside.

Before stating the proposition, a little bit of additional notation is required. Define $d_{\max}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$ to be the maximum entry of $\mathbf{d}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$, and $d_{\min}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$ to be the minimum entry.

THEOREM 2. *Suppose that for every k*

$$B_{kk}(\mathbf{P}, \mathbf{n}) \geq \sum_{\ell \neq k} B_{k\ell}(\mathbf{P}, \mathbf{n}).$$

Then

$$\lambda_2(\mathbf{T}(\mathbf{B}(\mathbf{P}, \mathbf{n}))) \leq 2(m-1) \cdot \frac{d_{\max}(\mathbf{B}(\mathbf{P}, \mathbf{n}))}{d_{\min}(\mathbf{B}(\mathbf{P}, \mathbf{n}))} \cdot \text{DWH}(\mathbf{B}(\mathbf{P}, \mathbf{n})).$$

REMARK 1. It is interesting to ask: does a close approximation of the same result hold, with high probability, when $\lambda_2(\mathbf{T}(\mathbf{B}(\mathbf{P}, \mathbf{n})))$ is replaced by $\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P}, \mathbf{n})))$ and $\text{DWH}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$ is replaced by $\text{DWH}(\mathbf{A}(\mathbf{P}, \mathbf{n}))$? We conjecture that the answer is yes. We already know from Theorem 2 of Golub and Jackson (2011) that $\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P}, \mathbf{n})))$ gets arbitrarily close to the left-hand side of the formula above under the expected-degree growth condition stated above. As for the right-hand side, the key is to show that $\text{DWH}(\mathbf{A}(\mathbf{P}, \mathbf{n}))$ converges to $\text{DWH}(\mathbf{B}(\mathbf{P}, \mathbf{n}))$. We believe this can be done by using concentration inequalities (see McDiarmid (1998)) to replace realizations (of weights, degrees, etc.) in the formula defining $\text{DWH}(\mathbf{A}(\mathbf{P}, \mathbf{n}))$ by the expectations of those quantities, which can be computed from $\mathbf{B}(\mathbf{P}, \mathbf{n})$. Nevertheless, the problem is far from trivial, because one needs to control the DWH of many cuts all at once in order to control their maximum.

4 Concluding Remarks

We have studied a measure of segregation in networks called degree-weighted homophily and have shown that it can be used to control, fairly precisely, the time it takes for a certain well-

⁶To make such statements formal, we must consider a sequence $\mathbf{P}(q)$ and $\mathbf{n}(q)$, and consider $q \rightarrow \infty$. The simplest way is to have \mathbf{P} fixed and only the sizes of the groups changing with q . When there is no risk of confusion, we suppress this indexing, but it is always in the background when we speak of asymptotics.

studied information propagation process (the DeGroot process) to operate. For a class of random graphs with two groups, this measure exactly characterizes the time to convergence.

The measure of segregation is intuitive and is based on the idea of counting links within and between groups. Within-group links increase the DWH measure and between-group links decrease it. How much each link matters depends on the degrees of the agents on its ends: links between low-degree agents count for more than links between high-degree agents.

Our analysis makes two main points. First, homophily and the segregation it induces in networks has important consequences for processes of interest, particularly ones of information flow. Homophily slows down information aggregation via repeated local averaging, and DWH captures a great deal about how it does so. Second, thinking about the substantive consequences of homophily or segregation can provide quite sharp guidance on how to measure it, and can motivate new measures of important network phenomena. This idea extends to other settings: while the precise measures relevant in any context are likely to be different, we suggest that modeling the substantive consequences of a network phenomenon in a particular setting can be quite useful in determining how best to quantify the phenomenon itself.

References

- ALON, N. AND V. D. MILMAN (1985): “Isoperimetric inequalities for graphs and super-concentrators,” *Journal of Combinatorial Theory Series B*, 38, 73–88.
- CHEEGER, J. (1970): “A lower bound for the smallest eigenvalue of the Laplacian,” in *Problems in Analysis*, ed. by R. C. Gunning, Princeton: Princeton University Press, vol. 27, 195–199.
- CHUNG, F., L. LU, AND V. VU (2004): “The Spectra of Random Graphs with Given Expected Degrees,” *Internet Mathematics*, 1, 257–275.
- COLEMAN, J. S. (1985): “Relational Analysis: The Study of Social Organizations With Survey Methods,” *Human Organization*, 17, 28–36.
- DEGROOT, M. H. (1974): “Reaching a Consensus,” *Journal of the American Statistical Association*, 69, 118–121.
- DEMARZO, P., D. VAYANOS, AND J. ZWIEBEL (2003): “Persuasion Bias, Social Influence, and Uni-Dimensional Opinions,” *Quarterly Journal of Economics*, 118, 909–968.
- DIACONIS, P. AND D. STROOCK (1991): “Geometric Bounds for Eigenvalues of Markov Chains,” *The Annals of Applied Probability*, 1, 36–61.

- ECHENIQUE, F. AND R. G. FRYER (2007): “A Measure of Segregation Based on Social Interactions,” *Quarterly Journal of Economics*, 122, 441–485.
- FRENCH, J. (1956): “A Formal Theory of Social Power,” *Psychological Review*, 63, 181–194.
- GOLUB, B. AND M. O. JACKSON (2010): “Naïve Learning in Social Networks: Convergence, Influence, and the Wisdom of Crowds,” *American Economic Journal: Microeconomics*, 2.
- (2011): “How Homophily Affects the Speed of Learning and Best-Response Dynamics,” Preprint, available at <http://www.stanford.edu/~jacksonm/homophily.pdf>
- HARARY, F. (1959): “A Criterion for Unanimity in French’s Theory of Social Power,” in *Studies in Social Power*, ed. by D. Cartwright, Ann Arbor: University of Michigan Press.
- JACKSON, M. O. (2007): “Social Structure, Segregation, and Economic Behavior,” Presented as the Nancy Schwartz Memorial Lecture, Available at SSRN: <http://ssrn.com/abstract=1530885>
- (2008): *Social and Economic Networks*, Princeton, N.J.: Princeton University Press.
- MCDIARMID, C. (1998): “Concentration,” Available at <http://cgm.cs.mcgill.ca/~breed/conc/colin.pdf>
- MCPHERSON, M., L. SMITH-LOVIN, AND J. COOK (2001): “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–44.
- MONTENEGRO, R. AND P. TETALI (2006): “Mathematical Aspects of Mixing Times in Markov Chains,” *Foundations and Trends in Theoretical Computer Science*, 1, 237–354.
- POLYA, G. AND S. SZEGO (1951): *Isoperimetric Inequalities in Mathematical Physics*, vol. 27 of *Annals of Mathematics Studies*, Princeton: Princeton University Press.
- VARGA, R. S. (2004): *Gerögorin and His Circles*, Berlin: Springer-Verlag.

5 Appendix: Proofs

Proof of Proposition 2: First, we show

$$\text{plim}_{n \rightarrow \infty} \lambda_2(\mathbf{T}(\mathbf{A})) = \text{plim}_{n \rightarrow \infty} \text{DWH}(N_1; \mathbf{A}).$$

We use the notation of Golub and Jackson (2011), which is introduced in Section 3.4 of this paper. For the left hand side, apply Theorem 2 of Golub and Jackson (2011) and then compute the second eigenvalue of

$$\mathbf{F}(\mathbf{P}, \mathbf{n}) = \begin{bmatrix} \frac{f_1 p_s}{f_1 p_s + f_2 p_d} & \frac{f_2 p_d}{f_1 p_s + f_2 p_d} \\ \frac{f_1 p_d}{f_1 p_d + f_2 p_s} & \frac{f_2 p_s}{f_1 p_d + f_2 p_s} \end{bmatrix},$$

(the result appears in Jackson (2008), Section 8.3.6). For the right hand side, first use the definition of DWH; then apply Lemma 4 of Golub and Jackson (2011) to show the degrees in the denominators in the DWH formula are arbitrarily close to their expectations⁷ with probability tending to 1 as n grows; then use the strong law of large numbers to conclude that the ratios appearing in the formula converge to their expectations. ■

We know from the definition of DWH and from Theorem 1 that

$$\text{DWH}(N_1; \mathbf{A}) \leq \text{DWH}(\mathbf{A}) \leq \lambda_2(\mathbf{T}(\mathbf{A})).$$

The above argument shows that the leftmost and rightmost quantities tend to the same value as $n \rightarrow \infty$, from which we conclude that DWH also tends to this value.

Proof of Theorem 2: Throughout, we hold \mathbf{P} and \mathbf{n} fixed and drop the corresponding arguments. We will write simply d_i for $d_i(\mathbf{B})$ and drop the arguments on d_{\max} and d_{\min} as well. Finally, λ_k in this proof will mean $\lambda_k(\mathbf{T}(\mathbf{B}))$.

Let \mathbf{v} be a right eigenvector of $\mathbf{T}(\mathbf{B})$ corresponding to the second-largest eigenvalue in magnitude, and assume \mathbf{v} is normalized so that $\|\mathbf{v}\|_{\mathbf{d}} = 1$. Define \mathbf{w} by

$$w_i = \begin{cases} \frac{1}{r d_i} & \text{if } v_i \geq 0 \\ -\frac{1}{(m-r) d_i} & \text{otherwise.} \end{cases}$$

where r is the number of nonnegative entries of \mathbf{v} .

Now, the key observation is that we can write

$$\mathbf{w} = a_2 \mathbf{v} + \sum_{k=3}^m a_k \mathbf{z}_k,$$

where \mathbf{z}_k is the eigenvector of $\mathbf{T}(\mathbf{B})$ corresponding to λ_k and $a_2 = \langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{d}}$. We are just computing the projection of \mathbf{w} on the subspace spanned by \mathbf{v} . Further, we may assume the eigenvectors are chosen so that each $a_k \geq 0$. Since $\mathbf{T}(\mathbf{B})$ satisfies the diagonal-dominance assumption in the proposition, all its eigenvalues (which are real because the matrix is

⁷In the sense that the ratio of the realized degree to the expectation can be guaranteed to be within δ of 1 for all nodes simultaneously with a probability at least $1 - \delta$ for sufficiently large n .

symmetric) are nonnegative by the Geršgorin circle theorem (Varga, 2004, Theorem 1.1). Thus,

$$\langle \mathbf{w}, \mathbf{T}(\mathbf{B})\mathbf{w} \rangle_{\mathbf{d}} \geq a\lambda_2. \quad (2)$$

In the proof of Lemma 3 of Golub and Jackson (2011), we saw that

$$\text{DWH}(\mathbf{B}) = \frac{\langle \mathbf{w}, \mathbf{T}(\mathbf{B})\mathbf{w} \rangle_{\mathbf{d}}}{\langle \mathbf{w}, \mathbf{w} \rangle_{\mathbf{d}}}.$$

Next, observe that

$$\langle \mathbf{w}, \mathbf{w} \rangle_{\mathbf{d}} = \sum_{v_i \geq 0} \frac{d_i}{d_i^2 r^2} + \sum_{v_i < 0} \frac{d_i}{d_i^2 (m-r)^2} \leq d_{\min}^{-1} \left[\frac{1}{r} + \frac{1}{m-r} \right] \leq 2d_{\min}^{-1}.$$

So $\text{DWH}(\mathbf{B}) \geq a_2 \cdot d_{\min} \cdot \lambda_2/2$.

It only remains to compute a_2 .

$$\begin{aligned} a_2 &= \langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{d}} \\ &= \sum_{i:v_i \geq 0} \frac{|v_i|}{r} + \sum_{i:v_i < 0} \frac{|v_i|}{(m-r)} \\ &\geq \sum_i \frac{|v_i|}{m-1} && r \leq m-1 \text{ and } m-r \leq m-1 \\ &= \frac{1}{m-1} \sum_i |v_i|^2 && \text{all the } v_i \leq 1 \\ &\geq \frac{1}{d_{\max}(m-1)} && \text{normalization of } v. \end{aligned}$$

Putting everything together, we conclude that $\text{DWH}(\mathbf{B}) \geq \frac{\lambda_2}{2(m-1)} \cdot \frac{d_{\min}}{d_{\max}}$. ■