Network bottleneck and the speed of learning

(DRAFT)

Jin F. Huang*

June, 2012

Abstract

We consider a learning model where agents update their opinions by taking an average of their neighbors' opinions, and the speed of learning is defined as the number of steps required for the updated behaviors to be arbitrarily close to the long-run behaviors. First, we bound the speed of learning by the bottleneck ratio. Second, we analyze this speed of learning and information exchange for a number of deterministic and random networks. Third, we consider network stability when agents are aware of network congestion, and by way of examples, we conclude that slow networks are not stable.

1. Introduction

1.1 Motivation

Our opinions are shaped by others. How we form our opinions, beliefs, and behaviors depends on whom we know, and so to understand the process of opinion formation, we model these social interactions in a network, which is represented by an undirected graph. Each agent has a set of neighbors whom she observes, talks to, and learns from. Because of this graph theoretical approach, models for opinion formation are intimately related to models for diffusion of innovation, aggregation of dispersed information, and the spread of epidemics in networks. For example, whether a person decides to use a social media platform depends on how many of her friends have already been using the technology; a political agent intending to learn the median voter's position aggregates dispersed signals by repeatedly requesting information from her friends, associates, and supporters; a disease spreads to different people through person-to-person contacts. A complete characterization of the random process of opinion formation would also allow us to better understand the dynamics of these examples.

Most studies on these interaction processes focus on establishing that in the long run, almost everyone in the society adopts the same behavior or opinion. In other words, there is a merging

*Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125. Email: jhuang@hss.caltech.edu

of opinion given that people are allowed enough time to communicate back and forth to reach a consensus, the long-run equilibrium. The opinion process at the tail end is non-random, and everyone agrees. But opinion dynamics before it reaches this very long-run limit is not well understood. In many models, we know that the opinion of each agent eventually settles down and converges[1], but we do not know much about the intermediate steps that the process goes through before the eventual limit occurs. This paper analyzes one aspect of the dynamics of opinion formation, giving an estimation of the amount of time the process takes to reach its equilibrium state.

We are going to analyze the speed of convergence in terms of the bottleneck ratio, a measure of the overall level of congestion that slows down the flow of information in a key region of a given network. We suppose that each agent's opinion is influenced only by her immediate neighbors at each time period. With this assumption, we find that the structure of the social network determines the dynamics of how opinion evolves and converges in the long run. The bottleneck ratio identifies the region of the network where agents are least likely to be influenced by those outside their group. We identify this inward-looking attitude as a key factor that slows down the learning process.

1.2 Roadmap

The study has three parts.[2] First, we describe the learning environment and the updating rule, the issue of convergence, and the connection between the speed of learning and the bottleneck ratio. We use the learning model that was first introduced by DeGroot (1974), in which agents update their opinions by taking an average of their neighbors' opinions. We investigate the dynamics of long-run opinion by the proxy measure of consensus time, defined as the number of periods that it requires for an opinion to be ϵ-close to the long-run opinion. Golub and Jackson (2011) study the relationship between homophily and speed of learning, where homophily is measured by the second largest eigenvalue in absolute value of a representative matrix. We extend their results by using a approach that relies on the estimation of the bottleneck ratio.

Second, we analyze the speed of learning in random networks. Opinions tend to converge more slowly in a large network than in a small network, and we focus on analyzing how consensus time grows with the size of the network. Because of Corollary 2.11, the behavior of the consensus time function depends crucially on the estimation of the bottleneck ratio. We show that the learning process in many of the common random networks is fast, because randomly connected neighbors are unlikely to cluster into a community that is almost isolated from the rest the network.

Third, we investigate network stability when agents are assumed to be aware of their contribution to congestion and be able to mitigate that by rewiring their connections. We look at three

[1]See Acemoglu and Ozdaglar (2010) for such a list of such models.

[2]I am planning to add an empirical and an applications section. The empirical section will require a dataset with social networks. The Add Health dataset, which contains many friendship networks, should be a natural fit. Applications will focus on comparing welfare implications between networks with the property of fast or slow speed of learning.

networks. The star network has a fast rate of convergence and is stable; the circle network is slow and not stable; and the dumbbell network is even slower than a circle, and is also not stable.

1.3 Related literature

There is a large and growing literature on mixing times of Markov chains. Our proxy for speed of learning is the notion of consensus time, which is mathematically equivalent to mixing time. Montenegro and Tetali (2006) and Levin, Peres, and Wilmer (2009) are excellent surveys on the topic. Cheeger (1969) introduces the bottleneck ratio to study networks, and Alon and Millman (1985) make the connection between mixing time and bottleneck ratio. Understanding mixing time in random networks is an active area of research in the field of theoretical computer science and applied probability, and we use results from those fields to study consensus time. For example, we adopt the estimation techniques in Fountoulakis and Reed (2008) and Benjamini, Kozma, and Wormald (2006), which analyze the mixing time in the giant component of a Erdös-Rényi random network, Cooper and Frieze (2003), which analyzes the Erdös-Rényi network when it is connected, and Addario-Berry and Lei (2012), which analyzes the Newman-Watts small world model.

Both Bayesian and non-Bayesian models have been used to study social learning and opinion dynamics. Bayesian learning processes assume that agents update their beliefs and act in a way that is statistically optimal, and non-Bayesian models impose restrictions on the learning process, explicitly stating what the boundedly rational agents can and cannot do. Acemoglu and Ozdaglar (2010) survey the literature for both types of models. Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992) use a simple Bayesian model to show a striking learning result, which states that there is a positive probability that agents do not properly aggregate the dispersed information in a network, a phenomenon known as herding. Bala and Goyal (1998) and Smith and Sorensen (2000) further this line of research of sequential observational learning models and study the conditions under which opinions converge. Acemoglu, Chernozhukov, and Yildiz (2007) and Acemoglu and Ozdaglar (2010) introduce additional structures (i.e., allowing uncertainty and communications) to illustrate that long-run opinions can diverge, leading to persistent disagreements.

On the other hand, most non-Bayesian approaches apply the DeGroot learning model (DeGroot, 1974). This learning process is a rule-of-thumb updating rule, where agents simply take the average of their neighbors' opinions as their own. Because of its simplicity, the DeGroot model is easy to analyze, and it provides insights into questions that are otherwise too difficult to answer in a setup with fully rational agents. Many variations of the model have been analyzed,[3] but most of them focus on the issue of whether learning occurs, i.e., whether long-run opinions converge.[4]

While there are many characterizations of how learning processes behave in the long run, how

[3]See Section 4 in Acemoglu and Ozdaglar (2010) for a review.

[4]See Golub and Jackson (2010) for a study of converging belief in discrete time and Lorenz (2005) in continuous time.

processes evolve or look like well before they are close to the limit is not well understood. In other words, what are the characteristics of a process when it is in between its initial and final stages? If opinion converges, we also want to know how fast it converges and what it is doing before it finally settles down. Bayesian learning models are difficult to analyze in this regard; Golub and Jackson (2011) make the first attempt in the DeGroot learning environment. They study the relationship between speed of learning and homophily, defined as the second largest eigenvalue of a representative matrix. This paper follows this theme, introducing the notion of bottleneck ratio and extending the analysis of speed of learning to a wider class of networks.

2. Measuring the speed of learning in a fixed network

2.1 The learning environment

Agents update their opinion by taking an average of their neighbors' behavior. Agents are represented by a node on a graph, and two agents are said to be connected if the two representative nodes are connected. We will use the terminology agent, connection, and graph interchangeably with node, edge, and network, respectively. The degree of a node is the number of edges connecting to that node. Let N be the set of agents with $n = |N|$. Denote i's degree by d_i. For a subset $S \subset N$, denote $d_S = d(S) = \sum_{i \in S} d_i$. A network is represented by its adjacency matrix A, a $(0,1)$-matrix of dimension $n \times n$ and with element $a_{ij} = 1$ if and only if the node i and j are connected, and $a_{ij} = 0$ otherwise. Let $\Delta = diag(d_1, d_2, d_3, \cdots, d_n)$ be the degree matrix.

Definition 2.1. The *influence matrix* T is the stochastic matrix,

$$T = 1/2(I + \Delta^{-1}A),$$

where I is a $n \times n$ identity matrix.

In the language of Markov chain theory, T describes a lazy random walk on a graph with the adjacency matrix A. The $\Delta^{-1}A$ term expresses the assumption that an agent is influenced by her neighbors uniformly, and the I term ensures that an agent weighs her own opinion as much as all of her neighbors' combined. The factor of $1/2$ normalizes T to be a stochastic matrix. Agents' initial *behaviors* (or *opinions*) are represented by a n-vector $b^0 \in \mathbb{R}^n$. b_i^0 is said to be agent i's behavior or opinion of at time $t = 0$. We denote the i,j-th element of the matrix T by either T_{ij} or $T(i,j)$. Behaviors are normalized so that $b_i^0 \in [0,1]$, for all $i \in N$. Agents update their behaviors at each time period t by the updating rule,

$$b_i^{t+1} = \sum_{j \in N} T_{ij} b_j^t, \tag{2.1}$$

and in matrix notation,[5] we write $b^{t+1} = Tb^t = T^t b^1$ and $b^1 = Tb^0$.

[5]Vectors are assumed to be column vectors unless they are multiplied by matrices from the left, in which case they

The updating rule states that agents repeatedly aggregate the information of their neighbors. As discussed by Golub and Jackson (2011), one interpretation of this updating process can be looked at as a myopic best-response in a pure coordination game if the influence matrix were to be defined as $T' \overset{\text{def}}{=} \Delta^{-1}A$; that is, agents do not weigh their own opinions from the previous period. With the influence matrix T', we assume that agents have utility functions of the following form,

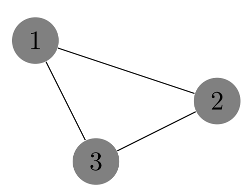
$$u_i(b^t) = -\sum_{j\in N} T'_{i,j}(b^t_j - b^t_i)^2.$$

Then agents would adopt the updating rule as defined by (2.1), only replacing T with T'. The focus of this paper is speed of learning, and as explained in the appendix, both updating rules as defined by T and T' lead to the same estimate of speed within the framework used in this study. There are technical advantages of using T in lieu of T', as the influence matrix T requires fewer assumptions to guarantee unique convergence.

The updating rule describes a simple behavioral pattern about how agents learn in a network environment. The time-invariant rule forces agents to learn from their neighbors the exact same way at each period. This can be viewed as a boundedly rational model in that agents do not have the flexibility to consider the overall structure of the network, even if they have that information. An agent is forced to considered that every one of her neighbors is equally important. An agent is not allowed to use the previous history to determine which one of her neighbors continues to receive new information and which one does not. Despite these shortcomings, the long-run behavior of this learning process can be shown to be equivalent to a fully rational learning model under further assumptions (DeMarzo, Vayanos, and Zwiebel, 2003; Golub and Jackson, 2010). There are also experimental evidence supporting the assumption that agents employ a naive learning rule, which is more similar to the process used here than a fully rational model (Chandrasekhar, Larreguy, and Xandri, 2010).

Example 2.2. (**Triangle network**)

This example and the next illustrate the mechanics of the learning environment described thus far. A triangle network has the following adjacency matrix A, Δ, and influence matrix T.



$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

are understood to be row vectors.

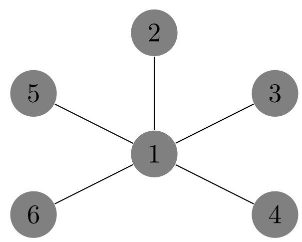
Let the initial behavior $b^0 = (1, 1/2, 1/3)$, then

$$b^1 = Tb^0 = (0.71, 0.58, 0.54).$$

$\square$

Example 2.3. (**Star network**)

A star network with five peripheral players and one central player has the following adjacency matrix A, Δ, and influence matrix T.



$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Let the initial behavior $b^0 = (.1, .2, .3, .4, .5, .6)$, then

$$b^1 = Tb^0 = (0.25, 0.15, 0.2, 0.25, 0.3, 0.35).$$

$\square$

2.2 Convergence of long-run behavior

We want to understand the community's long-run behavior. We say that behavior b_t is *convergent* with respect to an initial behavior b^0 if $lim_{t\to\infty} T^t b^0$ exists, and that the network has the property of *converging behavior* if b_t is convergent with respect to any initial $b^0 \in [0, 1]^n$. The latter condition is equivalent to the existence of $T^\infty \overset{\text{def}}{=} lim_{n\to\infty} T^t$ because the rows of T^∞ are necessarily the same whenever T is a stochastic matrix. T is said to be convergent whenever T^{∞} exists. A probability distribution π over the set $\{1, \cdots, n\}$ is said to be a *stationary distribution* if $\pi = \pi T$. Throughout this paper, we maintain the assumption that the network is connected.[6] We collect some properties of the influence matrix T in the following lemma. All proofs are in the appendix.

Lemma 2.4.

i. *T is irreducible and aperiodic.*[7]

ii. *T is convergent.*

iii. *T has a unique stationary distribution π, with $\pi_i = d_i/d_N$. Further $T^{\infty} = \boldsymbol{e}\pi$, where $\boldsymbol{e} = (1, \cdots, 1)^{tr}$ is a all ones column vector.*[8]

iv. *T is time reversible.*

The long-run behavior of the community is defined by the individual initial behavior b_i^0 and the intrinsic influence of the agents hold in the network, represented by π_i. If T^{∞} is convergent, the long-run behavior exists and is unique, as defined by $b^{\infty} \stackrel{\text{def}}{=} T^{\infty} b^0$. Furthermore, it is necessarily true that all agents' opinions are exactly the same, i.e., $b_i^{\infty} = b_j^{\infty}, \forall i, j \in N$. Note by property *iii* of Lemma 2.4, $T^{\infty}(i, \cdot) = \pi$. We can interpret π_i as agent i's intrinsic influence in the network because the limiting behavior for any $i' \in N$, $b_{i'}^{\infty} = \sum_i \pi_i b_i^0$ is simply the initial opinion weighted by the stationary distribution π. For this reason, we say i's relative influence is π_i. The following two examples illustrate the key concepts in this section.

Example 2.5. (**Triangle network**)

The network is described in Example 2.2.

$$T^{\infty} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \qquad \pi = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

Let the initial behavior $b^0 = (1, 1/2, 1/3)$, then

$$b^{\infty} = T^{\infty} b^0 = (0.61, 0.61, 0.61).$$

□

Example 2.6. (**Star network**)

---

[6]Connectedness could be relaxed without compromising most of the results. Golub and Jackson (2010) give a thorough discussion about the issue of convergence.

[7]The notion of irreducibility, aperiodicity, and time reversibility of a stochastic matrix are standard and can be found in textbooks on finite Markov chains. In particular, see Levin, Peres, and Wilmer (2009). For a concise description in the context of social networks, see Jackson (2008).

[8]tr in the superscript denotes the transpose of a vector

The network is described in Example 2.3.

$$T^\infty = \begin{pmatrix} \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{pmatrix} \qquad \pi = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{10} \\ \frac{1}{10} \\ \frac{1}{10} \\ \frac{1}{10} \\ \frac{1}{10} \end{pmatrix}$$

Let the initial behavior $b_0 = (.1, .2, .3, .4, .5, .6)$, then

$$b^\infty = T^\infty b^0 = (0.25, 0.25, 0.25, 0.25, 0.25, 0.25).$$

□

2.3 Consensus time and spectral gap

In this subsection we develop the terminology to answer the question of how long it takes for a behavior b_t to converge in a network. A row vector $\mu \in [0,1]^n$ is said to be a probability distribution over the set $N = \{1, 2, 3, \cdots, n\}$ if $\sum_i \mu_i = 1$. For any two probability distributions μ and ν, we define the *total variance distance*

$$||\mu_1 - \mu_2||_{tv} \stackrel{\text{def}}{=} \max_{S \subset N} |\mu(S) - \nu(S)|,$$

where $\mu(S) = \sum_{i \in S} \mu_i$. This distance measures the largest difference between the probabilities of any set $S \subset N$. The notion will be used to compare the behavior at time b^t and the limiting behavior b^∞, where its existence is guaranteed by Lemma 2.4. The comparison will be made precise with the definition of consensus distance.

We interpret the i-th element of a probability measure μ_i as agent i's relative influence. For example, when $\mu = \pi$, μ_i is the ratio between the number of people who are i's neighbors and the total number of pairs of agents who are neighbors, i.e. $\pi_i = d_i / \sum_j d_j$. Further, we define the *consensus distance*

$$cd(t; T) \stackrel{\text{def}}{=} \max_{i \in N} ||T_i^t - \pi||_{tv}, \tag{2.2}$$

where $T_i^t = T^t(i, \cdot)$ denotes the i-th row of the matrix T^t. For any integer $t > 1$, the t-th power of the stochastic matrix T is again a stochastic matrix; hence $T^t(i, \cdot)$ is a probability distribution over N. While we are in fact interested in the distance between the limiting behavior b^∞ and the behavior b^t, we note that by bounding the consensus distance, we also bound the difference between them.

Suppose that $cd(t;T) < \epsilon$, then[9]

$$|b_i^t - b_i^\infty| = |\sum_j T_{ij}^t b_j^0 - \pi_j b_j^0| \leq \sum_j |T_{ij}^t - \pi_j| b_j^0 < \epsilon ||b^0||_1,$$

which yields an upper bound on the normalized distance between b_i^t and b_i^∞, i.e., $\frac{1}{||b^0||_1}|b_i^t - b_i^\infty| < \epsilon$, for any $i \in N$. Another advantage of using the distance measure in terms of T and π is that it is independent of the initial behavior and only depends on network characteristics.

The next definition, $ct(\epsilon;T)$, is used as the proxy for the speed of learning, and understanding the dynamics of consensus time is our main focus.

Definition 2.7. The *consensus time* with respect to an arbitrary error $\epsilon > 0$ is

$$ct(\epsilon;T) = min\ \{t \geq 0 : cd(t;T) < \epsilon\}.$$

In words, this is the time that guarantees that the behavior b^t is ϵ-close to the long-run behavior b^∞. Lemma 2.4 already guarantees that for any initial behavior b^0, agents' long-run opinions converge, and the agreement point is the weighted average, $\sum_j \pi_j b_j^0$. Hence, the questions of existence and uniqueness have been settled, we now turn our attention to dynamics. The measure of consensus time is only an approximation of the behavior process $(b^t)_{t\in\{1,2,3,\cdots\}}$. We know the process b^t is converging and its exact behavior can be studied through simulations, but analytical results are difficult to obtained because its exact structure depends on both network characteristics and initial behavior.

The measure of consensus time with respect to ϵ seeks to unpack this particular question: For a fixed target $\epsilon > 0$, how long does it take for the sequence of b^t to be less than ϵ distance away from the weighted average? We provide answers to this question in Proposition 2.8 and Corollary 2.11. In particular, applying Corollary 2.11 with additional estimation techniques gives precise answers for a variety of networks in Section 3.

The first result on consensus time is obtained through applying standard results from Markov Chain analysis, bounding $ct(\epsilon;T)$ in terms of the spectral properties of the matrix T. In particular, the second largest eigenvalue in absolute value is a key determinant of the speed of convergence. All of the eigenvalues λ_j of T are real positive numbers, and we order them by their magnitudes,

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_k \geq 0,$$

where $k \leq n$. The number of distinct eigenvalues can be strictly less than n. Symmetric matrices are diagonalizable, and so there always exists a complete set of eigenvectors. The strict inequality between λ_1 and λ_2 follows from T's irreducibility and that the Perron root λ_1 has an algebraic multiplicity of 1. Define the spectral gap $\gamma \stackrel{\text{def}}{=} \lambda_1 - \lambda_2 = 1 - \lambda_2$. The following proposition gives a lower and upper

[9] $||\cdot||_1$ is the norm defined as the absolute row sum of the elements.

bound for consensus time in terms of the spectral gap.[10]

Proposition 2.8. $-log(2\epsilon)(\frac{1}{\gamma}-1) \leq ct(\epsilon; T) \leq -log(\pi_{min}\epsilon)\frac{1}{\gamma}$

In words, the rate of convergence is roughly proportional to $1/\gamma$, the inverse of the spectral gap.

2.4 Consensus time and bottleneck ratio

In this subsection we bound the speed of learning by the bottleneck ratio, which identifies the subset of agents that are least influenced by those outside the subset relative to its size as measured by π. The bottleneck ratio controls the size of the spectral gap and hence the has consensus time. Golub and Jackson (2011) state an explicit relationship between homophily and the speed of learning in a similar setup that also uses an average-based updating rule. The results obtained from analyzing the bottleneck ratio does not necessarily improve the bounds of consensus time. The bounds in Proposition 2.8 are tighter than those in Proposition 2.10. But the later proposition provides a connection between the dynamics of learning and a geometric property of the network. Further, the bottleneck ratio can be more easily estimated if the learning environments have sufficient details.

The bottleneck ratio is defined by the subset of agents who are least likely to be influenced by others, and their collective unwillingness to be persuaded is the source of congestion. Let $q(i, j)$ denote the *network influence* that j has on i,

$$q(i, j) \stackrel{\text{def}}{=} \pi_i T_{i,j}.$$

It measures how much i values j's opinion relative to π_i.[11] Recall that the stationary distribution π_i measures how important agent i is relative to everyone else. We extend the notion of network influence to subsets of agents so that for any $S_1, S_2 \subset N$, $q(S_1, S_2) = \sum_{i \in S_1, j \in S_2} \pi_i T_{i,j}$. The quantity $q(S_1, S_2)$ is the one period network influence that agents in group S_2 exert on agents in group S_1. Note that $q(N, j) = \sum_{i \in N} \pi_i T_{i,j} = (\pi T)_j = \pi_j$. Group S is the most inward-looking group if $q(S, S^c)$ is the smallest. To get a measure that is sensible even if we compare a large group with a small group, we cannot look at the overall influence of a group alone but need to take into account the size of that group. The more agents in S the more influence S accumulates, and the group N would always have the most influence. Scaling $q(S, S^c)$ by π_S, we define a weighted measure of network influence as the *bottleneck ratio* of S, where

$$\Phi(S) \stackrel{\text{def}}{=} \frac{q(S, S^C)}{\pi(S)}.$$

We interpret the numerator in the expression as the total amount of influence the outsiders S^c have on the agents inside the set S. The larger this number, the more easily agents in the community S are

[10] These are not necessarily the best bounds.

[11] When T is interpreted as the transition matrix of a Markov chain, $q(i, j)$ is also known as the ergodic flow from i to j.

affected by external opinion swings. We interpret the denominator as the intrinsic volume of group S as measured by π. The ratio measures how much group S is isolated from the rest of the network. A group with a small $\Phi(S)$ creates a bottleneck in the network because its members tend to interact only internally and thus ignoring the opinions beyond its walls. This would slow down the learning process.

Definition 2.9. The bottleneck ratio of the network is

$$\Phi \stackrel{\text{def}}{=} min\{\Phi(S) : S \subset N, \pi(S) \leq 1/2\}.$$

In words, the bottleneck ratio of the network is defined by the subset of agents that contributes the most congestion. We also denotes $S^* = arg\,min\{\Phi(S) : S \subset N, \pi(S) \leq 1/2\}$. The factor $\pi(\cdot)$ makes sure that Φ is between 0 and $1/2$.[12] It is immediate that $\Phi > 0$ if and only if the network is connected. We provide two examples here. $\Phi = 1/n$ for a circle network of even size n with $n > 3$; this network is congested. And $\Phi = 1/2$ for a network of a triangle; this network is free from congestion.

We build on the result in Proposition 2.8 by relating the bottleneck ratio to the spectral gap. The following result is widely known as the Cheeger inequality.[13]

Proposition 2.10. $\frac{\Phi^2}{2} \leq \gamma \leq 2\Phi$

Corollary 2.11. $-log(2\epsilon)(\frac{1}{2\Phi} - 1) \leq ct(\epsilon; T) \leq -log(\pi_{min}\epsilon)\frac{2}{\Phi^2}$

The corollary follows directly from Proposition 2.8 and 2.10. These inequalities are the main tools when we estimate consensus time.

3. Measuring the speed of learning in random networks

We characterize the speeds of learning in networks that we do not observe directly, the random networks. In particular, we investigate the dependence of consensus time $ct(\epsilon; T)$ on the size of the network. Random network models specify the process of how networks are generated according to parameters. For example, the Erdös-Rényi random network depends on n and p, the size of the network and the probability that any two agents are connected. Invariably, one of the parameters is n. We index $T(n)$ to emphasize its dependence on n. Throughout this section, we use C to denote a constant that is independent of n and ϵ, and C may vary from one expression from the next.

So far, we focus on the computation of the time it takes a behavior b_t to be ϵ-close to b_∞, and now we shift our attention to investigate how consensus time $ct(\epsilon; T(n))$ depends on n, for a fixed ϵ. We bound the growth rate of $ct(\epsilon; T(\cdot))$ as a function of n. To simplify notations and further the

[12] Φ is $\leq 1/2$ because the diagonal elements of T are $1/2$. The definition of Φ applies to all stochastic matrices, and in general, $0 \leq \Phi \leq 1$.

[13] The bottleneck ratio is also known as the Cheeger constant, isoperimetric constant, expansion parameter, and conductance.

point that ϵ is an auxiliary parameter, we denote $ct(n) \stackrel{\text{def}}{=} min\{t \geq 0 : cd(t; T(n)) < 1/(2e)\}$, where e is the natural number. The choice of $1/(2e)$ is arbitrary, and it is only important because $1/e < 1$. The next lemma allows us to upper-bound $ct(\epsilon, T(n))$ with an upper bound of $ct(n)$.

Lemma 3.1. *Let the influence matrix of a network be T, then*

$$ct(\epsilon; T(n)) \leq log(\epsilon^{-1})ct(n).$$

Before analyzing random models, we consider four deterministic networks. We estimate the bottleneck ratio by formula (A.2) in the appendix.

Example 3.2. The bottleneck ratio $\Phi(S)$ for a complete network is $\frac{1}{2}|\delta S|/\sum_{i \in S} d_i = \frac{1}{2}\frac{m(n-m)}{m(n-1)}$, and hence $\Phi = 1/4$ for a large n. The stationary weight π_{min} is $1/n$, and applying Corollary 2.11, we upper-bound $ct(n) \leq -1/8log(1/(2ne))$, and it follows that $ct(n) \leq C \log n$. The convergence rate of $\log n$ is a benchmark for fast convergence. □

Example 3.3. We look at a star network of size size $n + 1$, where n is even. It is not hard to see that $\Phi = 1/2$ because $\pi_i = 1/2$ if i is the central player. Any set S with $\pi_S \leq 1/2$ has as many cross edges as the total number of neighbors. The upper bound for $ct(n)$ is is again $C \log n$. □

Example 3.4. We estimate the bottleneck ratio of a circle network, where each agent is connected to exactly two neighbors. Let n be even, the $\Phi = 1/n$. Hence, $ct(n) \leq Cn^2 \log n$. The speed of learning on a circle network is slower than the learning rate in the previous two examples. □

Example 3.5. A dumbbell network is made up of two complete network of size n, connecting by only one pair of agents in the middle.

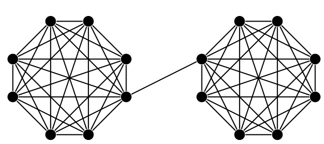


Figure 1: A dumbbell network with $2n$ nodes, where $n = 8$.

Its second largest eigenvalue[14] is at least $n^2/(n^2 + 1)$ and the spectral gap is approximately $1/n^2$. By Proposition 2.8, the consensus time is bounded at $ct(n) \leq Cn^2$. But bounding by the bottleneck ratio gives a different answer. We calculate $\Phi = 1/n^2$ for a large n, and it follows that $ct(n) \leq Cn^4 \log n$.□

The bounds derived via the spectral gap approach in Proposition 2.8 are always tighter than the bounds derived via Corollary 2.11. This is a drawback, but we will see shortly why the bottleneck ratio is a powerful analytical tool when we deal with more complex networks.

[14]The matrix of a simple random walk on dumbbell network has an eigenvalue of $n/(n + 1)$, and λ_2 of the lazy random walk on the same graph must be at least as large.

3.1 Erdös-Rényi networks

The Erdös-Rényi random network $G(n,p)$ is generated from a complete network with n agents by deleting each one of the $\binom{n}{2}$ connections with probability $1-p$ independently. A subset of agents $S \subset N$ is said to be connected if for any two agents $i, j \in S$, there is a sequence of neighbors who connect i to j. A component is the maximal connected subset. Let $\mathcal{C}$ denotes the largest component of $G(n,p)$. For $p = \frac{\lambda}{n}$, it is well-known that $\lambda = 1$ is the critical value for which a giant component emerges. If $\lambda > 1$, $G(n,p)$ is said to be supercritcal and the largest component $\mathcal{C}$ contains a positive fraction of all of the agents, *asymptotically almost surely* (abbreviated *a.s.s.*).[15] If$\lambda < 1$, $G(n,p)$ is subcritical and $|\mathcal{C}|$ *a.s.s.* of order $\log n$.

The following theorem gives an upper bound for the consensus time when the network has a giant component, i.e. $\lambda > 1$. This result is proved independently by Fountoulakis and Reed (2008) and Benjamini, Kozma, and Wormald (2006).

Theorem 3.6. *Let the network be the Erdös-Rényi $G(n, \lambda/n)$ with $\lambda > 1$. The consensus time in the largest component $\mathcal{C}$ is a.s.s.*

$$ct(n) \leq Clog^2(n).$$

If we allow the giant component of $G(n,p)$ to grow larger than what is assumed in the previous theorem so that $p = \lambda \log n/n$ with $\lambda > 1$, the network is connected almost surely, as $n \to \infty$.[16]

Theorem 3.7. *Let the network be the Erdös-Rényi $G(n, \lambda \log n/n)$ with $\lambda > 1$. The consensus time in the connected network is a.s.s.*

$$ct(n) \leq C \log n.$$

3.2 Preferential attachment network

The preferential attachment network is grown from a complete network with m_0 agents. Label these agents 1 to m_0. Let $t \in \{m_0+1, m_0+2, m_0+3, \cdots, n\}$ denotes time. An agent, also labeled by t, is added to the network at each time period t. The newly added agent t connects to exactly m existing agents, where $m_0 > m > 0$. The probability that the existing agent j connects to agent t is $\frac{m\, d_j(t)}{\sum_{k=1}^{k=t-1} d_k(t)}$, where $d_k(t)$ denotes the number of connections agent k has at time t. In other words, how likely j acquires one of the m connections from agent t depends on j's degree at time t. The highly connected agents are more likely to be connected with the new agent. The network at $t = n$ with parameter m_0 and m is said to be the preferential attachment network $PA(n,m)$. Note that parameter m_0 does not affect any asymptotic characteristics of the network when n is large.

Mihail, Papadimitrou, and Saberi (2004) analyze the preferential attachment model and find that it has a constant bottleneck ratio, which leads to the following estimate of consensus time.

[15] A random property that depends on n is said to be *a.s.s.* if the property holds almost surely as $n \to \infty$.

[16] All standard textbooks on random graph theory prove a version of this fact. In particular, seeBollobás (2000).

Theorem 3.8. *Let the network be the preferential attachment network $PA(n, m)$. The consensus time is a.s.s.*

$$ct(n) \leq C \log n.$$

3.3 Small world network

We will consider the small world model introduced by Newman-Watts. The network is modified version of a base network, the circle with n agents with each agent connecting to its closest $2k$ neighbors. A small world network is obtained by adding a connection between any pair of non-neighbors with probability $0 < p < 1$. We denote this network $NW(n, k, p)$.

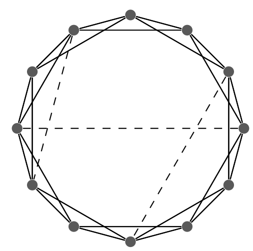


Figure 2: A $NW(12, 2, p)$ smal world network. Dashed lines are potential long-range connections.

The following theorem is due to Addario-Berry and Lei (2012). Note that this network has a slower speed of learning than the previous three models.

Theorem 3.9. *Let the network be $NW(n, k, \lambda/n)$ with $\lambda > 1$.*[17] *Then the following holds asymptotically almost surely,*

$$ct(n) \leq Clog^2(n).$$

A lower bound can be obtained via the following argument. For any interval of length $2l$ in the network, the probability that no agent in the interval acquire a long-range connection is

$$(1 - \lambda/n)^{(2l+1)[n-(2l+1)]} \geq (1 - \lambda/n)^{(2l+1)n} \geq exp(-2\lambda(2l + 1)).$$

The second inequality holds since $(1 - a/n)^n \geq e^{-2a}$, for any $a > 0$ when n is large enough. Letting $l = \log n/(8\lambda)$, the probability is greater than $e^{-(\log n/2+2c)}$, which is approximately $n^{-1/2}$ when n is sufficiently large. Then, the probability that all such intervals, which there are n of them, have at least one long-range connection is $(1 - n^{-1/2})^n$, which vanishes. That is, *a.s.s.*, there is at least one interval of length $\log n/(8\lambda)$ that has no long-range connection. This bounds the bottleneck ratio at $\Phi \leq C/\log n$, and by Corollary 2.11, we have $ct(n) \geq C \log n$.[18]

[17]Small world effect refers to the fact that network has a small diameter, the largest distance between any two agents in the network. For example, the diameter of the Erdös-Rényi network $G(n, \lambda/n)$ with $\lambda > 1$ is of order $\log n$.

[18]This bound can be tightened to $ct(n) \geq Clog^2(n)$ by a more elaborate, coupling argument.

3.4 Multi-type networks

The multi-type random network is described by the tuple $(\boldsymbol{n}, P)$, where $\boldsymbol{n} = (n_1, n_2, \cdots, n_K)$, where K denotes the total number of types, and n_k is the number of people that are of type k. Also let n denotes the total number of agents in the network, i.e. $n = \sum_{k=1}^{K} n_k$. The connection matrix P is of dimension $k \times k$, where its lk-th element, p_{lk}, represents the connecting probability between an agent of type l and an agent of type k. It is assumed that $p_{lk} = p_{kl}$. Let $\tilde{P}$ be an expansion of P such that $\tilde{P}$ is of dimension $n \times n$. Let $\tau(i)$ denotes i's type, then the ij-th element of $\tilde{P}$, $\tilde{p}_{ij}$, is simply be $p_{\tau(i)\tau(j)}$. The connecting probability between any two agent is determined by their types. We denote this network $MT(\boldsymbol{n}, P)$.

We restrict our attention to a simple version of the multi-type random model to show that this random model can give rise to networks that have fast speed of convergence and also networks of slow speed of convergence. This is not a surprise because the multi-type model can describe any network, the tuple $(\boldsymbol{n}, P)$ can has as many free parameters as a given network. To see this, consider an arbitrary network with n agents, then simply let the vector $\boldsymbol{n} = \{1, 2, \cdots, n\}$ and $P = A$. Hence, $(\boldsymbol{n}, P)$ deterministically replicate the given network.

We will consider only the *island model* (Marsden, 1987; Golub and Jackson, 2011). There are K types, and $\boldsymbol{n} = \{n_1, \cdots, n_K\}$ with $n_l = n_k$, for all $l, k \in \{1, \cdots, K\}$. For notational convenience, we also denotes $K \stackrel{\text{def}}{=} \{1, 2, \cdots, K\}$ whenever there is no ambiguity. Assume that $p_{ll} = p_s$, for all $l \in K$; $p_{lk} = p_d$, for all $l \neq k$; and $p_s > p_d$. In words, each type has the same number of agents, each agent of the same type is indistinguishable, and each island is also indistinguishable. Two agents connect with p_s if they come from the same group, and connect with p_d if they come from different groups. We note this random network by $IM(n, K, p_s, p_d)$ with the simplifying assumption that K divides n.

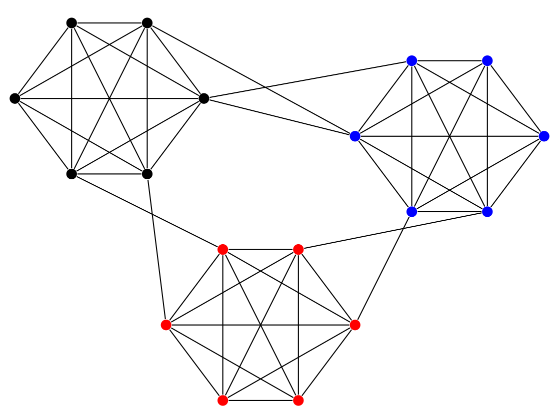


Figure 3: An island model where $p_s = 1$ and $p_d < 1$.

Assumptions 3.10. For the $IM(n, K, p_s, p_d)$ model, we make the following regularity assumptions:

i. Let $n_l = n_k$, for all $l, k \in K$. Hence, $n = n_k K$.
ii. Let $p_s = \frac{\lambda \log n_k}{n_k}$ with $\lambda > 1$.
iii. Let $p_s = \frac{\lambda}{n_k^{2-\epsilon}}$ with $\lambda > 1$ and $\epsilon > 0$.

Condition (i) is a simplifying assumption. Assumption (ii) ensures that each island is connected.[19] Assumption (iii) makes sure that any two islands are connected.[20]

Theorem 3.11. *Let $\frac{p_d}{p_s} = \frac{\lambda}{n^a}$. If $a = 0$ and $0 < \lambda < 1$, then $ct(n) \leq C \log n$. If $a, \lambda > 0 \geq 0$, then $ct(n) \leq Cn^{2a} \log n$.*

In words, p_d determines the speed of convergence of the island model. The bottleneck ratio precisely captures this notion: when $p_d/p_s \geq \lambda$, the bottleneck ratio is greater than some constant independent of n and the network, which indicates that the network is relatively free of potential congestion; when p_d/p_s decreases, each group becomes more isolated from each other, contributing to a slower speed of information exchange. The probability of linking across groups can be interpreted as the how much an agent is influenced by people outside her group than members of her own group. As p_d/p_s decreases, each group is becoming more and more of its own island.

4. Congestion avoidance and network stability

We define a notion of network stability with respect to network congestion, and illustrate how awareness of one's own contribution to the overall slowing down of the learning process can allow agents to avoid undesirable networks. We say that a network has a high level of congestion if the bottleneck ratio is small, i.e. the speed of convergence is slow. Each agent may unilaterally rewire her links to anyone in the network, but she is not allowed to add connections, or rewire them in such a way that some agents are left in isolation. We denote a modified network by T' and

$$\Phi^*(i;T) = \underset{i \in S,\, \pi_S \leq n/2}{min} \Phi(S;T),$$

where $\Phi(S;T)$ is the bottleneck ratio of the subset S with respect to the network represented by a matrix A that has the corresponding influence matrix T. And $S^*(i;T)$ is a set at which $\Phi^*(i;T)$ is attained. We denote $\{S^*(i;T)\} \stackrel{\text{def}}{=} \underset{i \in S,\, \pi_S \leq n/2}{arg\,min} \Phi(S;T)$. Agent i is said be able to *reduce congestion* if one of the following two conditions is satisfied:

$$i. \quad \Phi^*(i;T') > \Phi^*(i;T), \text{ or} \tag{4.1}$$

$$ii. \quad \Phi^*(i;T') = \Phi^*(i;T) \text{ and } |\{S^*(i;T')\}| < |\{S^*(i;T)\}|. \tag{4.2}$$

[19]An island is a Erdös-Rényi random graph $ER(n_k, p_s)$.

[20]Assumption (ii) and (iii) are stronger than what are needed to ensure that the island model is connected.

In words, agent i perceives that she would ease out the congestion of the network if she is able to either reduce the maximum level of congestion, or if she is able to reduce the number of ways that maximum level is encountered. A network is said to be *stable* if no agent is able to reduce congestion by rewirings. We use a few examples to illustrate this notion of stability.

Example 4.1. (**Star network**)

A star network of size $n+1$ has a fast convergence rate with $\Phi = 1/2$ and $ct(n) < C \log n$. It is stable. The central agent cannot modify her connections because any unilateral rewiring would disconnect at least one agent. A peripheral agent i may attempt to rewire her linking to the center to a peripheral site instead, but then if two peripheral agents i and j are neighbors, $\Phi^*(i; T') \leq \Phi^*(\{i,j\}; T') = 1/6 < 1/2 = \Phi^*(i; T)$. No agent is able to reduce congestion. □

Example 4.2. (**Circle network**)

A circle network of size $2n$ has a slow convergence rate, $\Phi = 1/(2n)$ and $ct(n) < Cn^2 \log n$. It is not stable because there is an agent who is able reduce congestion.

We pick agent i in the network, and label those on the clockwise direction in increasing order as 1 to $n-1$, and those on the counter-clockwise positions as $1'$ to n'. i's original neighbors are $1'$ and 1, and let her new neighbors be $n/2$ and $(n/2)'$. It is clear that $\Phi^*(i; T') = \Phi^*(i, T)$ because the i's two neighbors are just close enough so that there is no set S with $|S| \leq n/2$ that can internalize both of i's connections without also having at least two cross edges. But there are only 3 maximally congested sets $S^*(i; T')$ in the rewired network while there are $|\{S^*(i; T)\}| = n$.

First, we show that $\{S^*(i; T)\} = n$. It is easy to see that $S^*(i; T)$ must be a connected interval because if not, $S^*(i; T)$ would have at least at least 4 edges that cross into its complement, $N \backslash S^*(i; T)$. Also, $|S^*(i; T)|$ must be n for the denominator in A.2 to be the largest. The number of sets satisfying these two conditions and including i is exactly n. Second, we show that $|\{S^*(i; T')\}| = 3$. $S^*(i; T')$ cannot have both of i's edges cross into the complement; otherwise, the set would not be an interval. With this restriction, we divide the counting of the number of $\{S^*(i; T')\}$ into two cases. Case 1: $S^*(i; T')$ allows exactly one of i's two edges to be border cross. Then, $S^*(i; T')$ must include either node 1 or node $1'$ to limit the border cross edges in $S^*(i; T')$ to two. Hence, there are exactly two such $S^*(i; T')$ sets. Case 2: $S^*(i; T')$ internalize both of i's edges, but then both 1 and $1'$ must be included in the set. There is exactly one such set. We conclude that $|\{S^*(i; T')\}| = 3$, and we have constructed a way for agent i to rewire her edges to reduce congestion. □

Example 4.3. (**Dumbbell network**)

A dumbbell network of size $2n$ has a slow convergence rate with $\Phi = 1/(n^2 - n + 1)$ and $ct(n) \leq Cn^4 \log n$. Let n be even. This network is not stable.

We consider a modification (T') described as follows. Let us label the agent in the central position of the complete subnetwork on the left as i. We rewire $n/2 - 1$ of i's connections with the left subnetwork and connect them to agents on the right so that i has $n/2$ connections to the left and $n/2$ connections to the right.

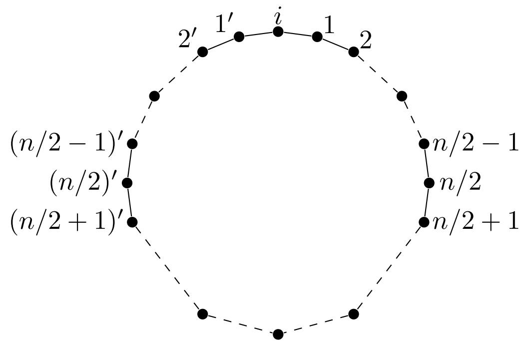
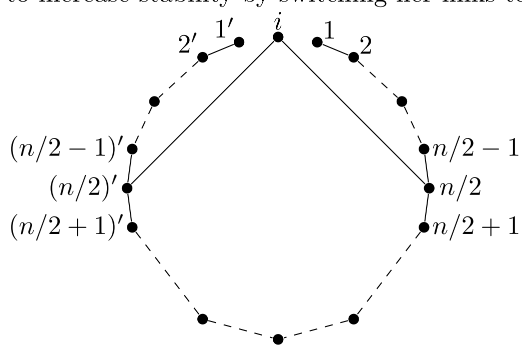


Figure 4: Agent i seeks to increase stability by switching her links to node $(n/2)$ and $(n/2)'$.



We want to show that $\Phi^*(i;T') > \Phi^*(i;T)$. We will use A.2 to compare $\Phi(S;T')$ and $\Phi(S;T')$ for any subset S with $i \in S$. We observe that for a sufficiently large n, $\pi_i \approx \pi_j$, for any $j \in N$, and so $\pi_S = \sum_{\in S} d_k / \sum_j d_j \propto |S|$. It is sufficient to show that for any size, $|S| \stackrel{\text{def}}{=} s$, with $1 \leq s \leq n$,

$$\underset{S\subset N:\,|S|=s}{min} e(S, S^c; T') > \underset{S\subset N:\,|S|=s}{min} e(S, S^c; T),$$

where $e(S, S^c; T)$ is the total number of cross edges between S and S^c with respect to T. We break the analysis down into four cases.

Case #1, $s = n$. This is obvious because $e(S, S^c; T) = 1$. The modification increases the number of cross edges, and it follows that $\Phi^*(i;T') > \Phi^*(i;T)$.

Case #2, $s = n - k$ with $1 \leq k \leq n/2 - 1$. With this criteria of S, we count the minimum number of cross edges in the modified network, $\frac{n^2}{4} + k + (n/2 - k)(n/2 - 1 + k)$, and for the original dumbbell, the minimum number $1 + k(n - k)$. Using the fact that $k \leq n/2 - 1$, we conclude that $min\,\{e(S, S^c; T') : |S| = s\} > min\,\{e(S, S^c; T) : |S| = s\}$ if and only if $n/2 + 2k - 1 > 0$, which holds for a large n.

Case #3, let $s = n/2$. This is the critical case where S includes exactly 1 agent from the left subnetwork, i, and $n/2 - 2$ agents from the right subnetwork. Then the minimum number of cross edges in T' is $(n - 2) + 1 + (n/2 - 2)(n/2 + 2)$, and with respect to T, $n^2/4 + 1$. And $min\,\{e(S, S^c; T') : |S| = s\} > min\,\{e(S, S^c; T) : |S| = s\}$ if and only if $n > 6$, which holds for a large

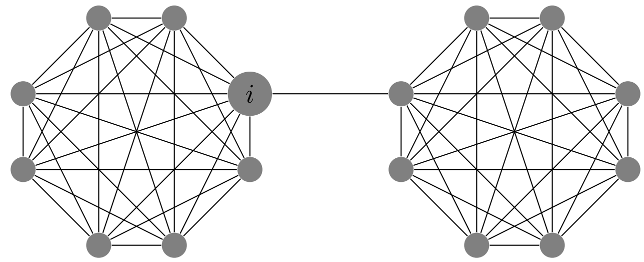
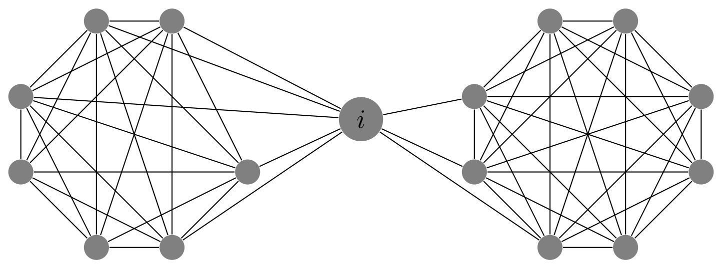


Figure 5: Agent i rewires her connections so that she evenly divides her links between the left and the right subnetwork.



n.

Case #4, $s = n-k$ with $n/2 < k < n$. The subset S that minimizes the number of cross edges includes 1 agent from the left subnetwork, i, and $n-k-2$ agents from the right subnetwork. We count the cross edges between this S and S^c in the modified network, $2(n-1)+(k+2-n/2)+(k+2)(n-k-2)$, and for the original dumbbell, the minimum edges are $(n-k)k+1$ with an S that includes $|S|$ members of i's subnetwork. Then, $min\,\{e(S,S^c;T') : |S| = s\} > min\,\{e(S,S^c;T) : |S| = s\}$ if and only if $7n/2 - 3k - 5 > 0$, which hold when $n/2 < k < n$ and n is large. $\square$

These examples show how we can determine if a given network is stable or not, and now we consider existence of stable network. To be precise about what we mean about existence, we first observe it is easy to construct a stable network of size n. The star network is always stable because its bottleneck ratio is $1/2$ and any deviation from that only make Φ smaller. We want to know that starting from any unstable network, is there a finite sequence of action, each taken by a deviating agent, such that it modifies the network to be stable?

Definition 4.4. We say that two networks *differ by an one-step rewiring* if the set of edges of the two networks are identical for all but one agent.

In other words, one network can be obtained from another through some rewirings by one agent.

We are interested in finding a sequence of networks such that any two consecutive networks satisfy this property. The existence of such a sequence suggests that if agents are willing to coordinate in the order to which they act, they can always reach a stable network simply by changing their existing connections.

Proposition 4.5. *For any network, there is a finite sequence of one-step rewiring such that*

i. Each rewiring is performed by the agent who is reducing her congestion in the sense of (4.1) and (4.2).

ii. The last rewiring results in a stable network.

Further, it turns out that the order to which this sequence of rewiring happens does not matter if all we want to is a stable network in the end. In other words, agents do not need to overtly cooperate with each to assure that their actions will collectively lead to stability. Even if the order to which they are chosen to act is completely random, so long as they do not interfere with the connections of those who are already content with their stability, their selfish actions reduce overall congestion. The order of action only matters if there are different stable networks, which in general there are many, then one sequence may lead to a final network tjat has less congestion than another sequence may.

We conclude this section by making a few remarks about the limit of the analysis of stability presented here. First, the arguments for stability in the three examples do not yield much insight about how to analyze a general class of network, such as random networks. The reasoning of stability relies on explicit details of network structure, and so in applications, unless we observe all the connections, we would not be able say much about whether the network is stable or not. Second, for the same reason that we, as observers and outside analysts, need to have global information to assess stability, it is even more difficult to justify the assumption that the agents within the network should have the prescience to correctly compute their contribution. The agents need to be sophisticated enough to both identify the most congested subset as well as estimate their own contribution to the overall congestion. Third, even if agents have the information and the ability to do these complex calculations, why would they care about the efficiency of information exchanges in the network? To tackle this challenge, we need to extend the learning environment in at least two ways, endowing agents with payoff functions that are directly related to the speed of learning and giving them choices as they respond to those incentives.

5. Concluding remarks

We have shown that the technique of bottleneck ratio can be used in a variety of settings to give us bounds on the speed of convergence. The network is speed-constrained if there is a group in the society such that first, it does not interact much with those who are outside of the group, and second, it is relatively large. Information does not travel in or out of this insular group, creating an informational separation between two sections of the society. The bottleneck ratio measures this friction in a network.

The bottleneck ratio has an intuitive geometric interpretation in terms of network structures, and this measure can often be estimated with only a few details about the underlying network. This analytical tractability should allow the approach to be used for additional studies. We propose three such extensions. First, we allow strategic interactions. So far, we assume that agents use a mechanical learning rule, averaging the opinions of their neighbors. This naive updating process may be plausible under some qualifying assumptions, but if there is a range of actions available to agents, the speed would not only change with respect to network structures but also change with respect to strategic incentives. For example, a society at large may be able to provide benefits to some members the group that clogs up the network, enticing them to increase their amount of communication with members outside, and thus, creating a sufficient amount of information exchanges that squeeze through the tight bottleneck to counterbalance the slow convergence rate.

Second, we need to understand what type of networks are formed if agents agree that less congestion is more desirable. Section 3 considers stability with respect to network congestion, but it only analyzes which network is stable and which is not. It leaves open the question as to how agents choose their connections if they are starting afresh to form a network. Alternatively, if we allow a social planner to design a network, we want to know what is the most efficient network according to her consideration of network congestion as well as other constraints.

Third, we need to test our theory. We want to know the qualities of the bounds of speed of learning when they are compared with the exact speeds in real-world social and economic networks. We have shown that most random networks have low levels of congestion. This is not surprising given our interpretation of the bottleneck ratio, where friction is caused by a relatively isolated group. Randomness eliminates that tendency. In the case of the preferential attachment model, while there are agents who accumulate a disproportionate amount of connections, they act as information hub, speeding up the mixing process rather than becoming choke points. However, it is well-known that networks generated from random models do not look at all like real social and economic networks (Jackson, 2008). The results in Section 2 might give us a false sense of security if we believe that large networks generally have a fast rate of convergence. To get pass this, we envision a straightforward empirical study. We calculate the speeds of learning for as many empirical networks as possible in two ways, by way of estimating the bottleneck ratio and by direct simulation, and then we compare the two sets of speeds. The only challenge is to gather as many real social and economic networks as possible from the existing literature and databases.

A. Appendix

A.1 Spectral properties of the influence matrix

We make a few observations about the influence matrix T and the matrix $\widetilde{P} \overset{\text{def}}{=} \Delta^{-1}A$. $\widetilde{T}$ can a more natural candidate as the defining influence matrix because it does not require the assumption that agents strongly weigh their own opinion. T has two advantages over $\widetilde{T}$. T is always irreducible and aperiodic, as stated in Lemma 2.4. Further, the two candidate matrices have similar spectral properties. Let μ_i denote the eigenvalues of $\widetilde{T}$, then it follows from $T = 1/2(I + \widetilde{T})$ that

$$\lambda_i = 1/2(1 + \mu_i). \tag{A.1}$$

In particular, both set of eigenvalues have the same cardinalities and each pair of the corresponding eigenvalues have the same multiplicities. It is also important that the calculations of the bottleneck ratio for the two matrices are closely related. With the definition of Φ and the stationary distribution $\pi_i = d_i / \sum_j d_j$, the bottleneck ratio with respect to T can be expressed as

$$\Phi_T = 1/2 \min_{\pi(S) \leq 1/2} \frac{e(S, S^c)}{\sum_{i \in S} d_i} \tag{A.2}$$

where $e(S, S^c)$ is the number of pairs of neighbors that one agent resides in the set S and the other resides in the set S^c, the cross edges. On the other hand, the bottleneck ration with respect to $\widetilde{T}$ is $\Phi_P = \min_{\pi(S) \leq 1/2} \frac{e(S,S^c)}{\sum_{i \in S} d_i}$. Therefore, the two updating give identical bound on consensus time whenever the analytical approach makes use of Proposition 2.10.

The influence matrix T as defined in Section 2 is time reversible because it is a lazy random walk on an undirected graph. However, T may not be symmetric. By a similarity transform of $\widetilde{T}$ with $\Delta^{1/2}$,

$$R = \Delta^{1/2} \widetilde{T} \Delta^{-1/2} = \Delta^{-1/2} A \Delta^{-1/2}.$$

R is symmetric and hence all of its eigenvalues are real. $\widetilde{T}$ share those eigenvalues, and by (A.1), all of T's eigenvalues are real positive numbers. Further, T has a diagonalizable and has a spectral decomposition,[21]

$$T = \lambda_1 G_1 + \lambda_2 G_2 + \cdots + \lambda_k G_k,$$

where $G_j = x_j y_j / y_j x_j$, x_j is the right-hand eigenvector, and y_j is the left-hand eigenvector associated with the eigenvalue λ_j, $\sum_{j=1}^k G_j = I$, and $G_1 = \Pi = T^\infty$ with $\Pi \overset{\text{def}}{=} e\pi$. The spectral decomposition greatly simplifies the task of calculating the power of the matrix in that $T^t = \lambda_1^t G_1 + \lambda_2^t G_2 + \cdots + \lambda_k^t G_k$. The decomposition also illustrates that under the condition $1 = \lambda_1 < |\lambda_2|$, $|\lambda_2|$ is critical to the convergence rate of T^t. It is not hard to see that the rate of decay is $|\lambda_2|^t$.

The spectral decomposition can also be expressed in terms of the inner product with respect to π. Note that we sometimes write $f(i) = f_i$. For any $b, c \in \mathbb{R}^n$, define the inner product $\langle b, c \rangle \overset{\text{def}}{=} \sum_{i=1}^n b_i c_i \pi_i$. For a diagonalizable T there is a complete set of orthogonal basis, and by the Gram-Schimdt procedure, there is an orthonormal basis $\{x_k\}_{k=1}^n$ and corresponding $\{\lambda_k\}$ such that

$$Tb = \sum_{k=1}^n \langle b, x_k \rangle \lambda_k x_k. \tag{A.3}$$

[21]See Chapter 7 in Meyer (2000).

Note that each eigenvalue λ_j is semi-simple, so the total number of eigenvalues may be less than n. There are n eigenvectors. And any vector $b \in \mathbb{R}^n$ can be decomposed as $b = \sum_{k=1}^{n} \langle b, x_k \rangle x_k$. It follows from (A.3) and the definition of inner product that

$$T_{ij}^t = (T^t \delta_j)_i = \pi_j \sum_{k=1}^{n} \lambda_k^t x_k(i) x_k(j), \tag{A.4}$$

where $x_k(i)$ is the i-th component of x_k. Another useful observation is that

$$\pi_i = \langle \delta_i, \delta_i \rangle = \pi_i^2 \sum_{k=1}^{n} x_k(i)^2. \tag{A.5}$$

The first equality is by direct calculation and the second equality comes from applying the orthonormal decomposition.

A.2 Proofs of statements in Section 2

We give the proof of Lemma 2.4.

Proof. Irreducibility follows directly from connectedness, and aperiodicity follows from the definition of the influence matrix in that an agent considers her own opinion exactly as important as the sum of all of her neighbors' opinions. The Perron–Frobenius theory states that irreducible, aperiodic stochastic matrices are convergent.[22] which is property (*ii*). Irreducibility by itself is sufficient to ensure the existence and the uniqueness of the stationary distribution π. In particular,[23] for any $i \in N$,

$$\pi_i = \frac{d_i}{\sum_j d_j}. \tag{A.6}$$

To verify this, the i-th element of πT is

$$\sum_j \pi_j T_{ij} = \sum_j \frac{d_j}{\sum_k d_k} T_{ij} = \frac{1}{\sum_k d_k} \sum_{j:ij \text{ is a link}} d_j \frac{1}{d_j} = \frac{d_i}{\sum_j d_j}.$$

Time reversibility is equivalent to the condition of *detailed balance equations*, $T_{ij}\pi_i = T_{ji}\pi_j$. Using the formula of π in (A.6), we see that for any pair of agents $i, j \in N$,

$$T_{ij}\pi_i = \begin{cases} \frac{1}{\sum_j d_j} & \text{if } ij \text{ a link} \\ 0 & \text{otherwise} \end{cases}.$$

□

[22] See Chapter 8 in Meyer (2000).

[23] See Chapter 1 in Levin, Peres, and Wilmer (2009).

We give the proof of Proposition 2.8.

Proof. To get the lower bound, we will find an lower estimate of $cd(t)$. For any right eigenvector of f with $b \neq \boldsymbol{e}$, where bfe is an all ones vector, $\sum_i \pi_i f_i = \langle f, \boldsymbol{e} \rangle = 0$ because any two distinct eigenvectors are orthogonal (see appendix). We calculate$|\lambda^t b_j| = \left|\sum_i T^t_{j,i} f_i - \pi_i f_i\right| \leq max_j f_j \sum_i |T^t_{j,i} - \pi_i|$. Noting that $||T^t_j - \pi|| = 1/2 \sum_i |T^t_{j,i} - \pi_i|$ and applying the definition of $cd(t)$, we obtain the bound $\lambda_2^t \leq 2cd(t)$. This concludes that

$$ct(\epsilon; T) \geq -log(2\epsilon)\frac{1}{log(\lambda^{-1})} \geq -log(2\epsilon)(\frac{1}{\gamma} - 1).$$

To get the upper bound, we use the spectral decomposition of T as in (A.4) to estimate the variation between T^t_{ij} and π_j,

$$\begin{aligned}
|T^t_{ij} - \pi_j| &= |\pi_j \sum_{k=1}^{n} \lambda_k x_k(i) x_k(j) - \pi_j| \\
&= \pi_j |\sum_{k=2}^{n} \lambda_k x_k(i) x_k(j)| \qquad \text{Because } \lambda_1 = 1 \text{ and } x_1 = (1, \cdots, 1) \\
&\leq \pi_j |\lambda_2|^t \sum_{k=2}^{n} |x_k(i) x_k(j)| \\
&\leq \pi_j |\lambda_2|^t \left(\sum_{k=2}^{n} x_k(i)^2 \sum_{k=2}^{n} x_k(j)^2 \right)^{1/2} \qquad \text{By the Cauchy–Schwarz inequality} \\
&\leq \pi_j |\lambda_2|^t \frac{1}{\pi_i} \frac{1}{\pi_j} \qquad \text{By (A.5)} \\
&= |\lambda_2|^t \frac{1}{\pi_i}.
\end{aligned}$$

Applying the definitions of consensus time, total variation distance, and spectral gap, we get

$$cd(t; T) \leq |\lambda_2|^t (1 - \pi_{min}) \leq (1 - \gamma)^t (1 - \pi_{min}) \leq (1 - \pi_{min}) e^{-\gamma t},$$

where $\pi_{min} = min_{i \in N}\{\pi_i\}$. We state the the result in terms of consensus time,

$$ct(\epsilon; T) \leq -log(\epsilon \pi_{min}) \frac{1}{\gamma}.$$

□

We introduce Dirichlet forms to relate the bottleneck ratio with consensus time. Let $f, g \in \mathbb{R}^n$ be a real value function. Define the expectation of f with respect to π as $\mathbb{E}_\pi f \stackrel{\text{def}}{=} \sum_i \pi_i f_i$ and the variation of f with respect to π as $\mathbb{V}_\pi f \stackrel{\text{def}}{=} \mathbb{E}(f - \mathbb{E}(f))^2$. We define the *Dirichlet form* of two real value functions as

$$\mathcal{E}^\pi_{f,g} = \langle f, (I - T) g \rangle.$$

We omit the dependence on π for $\mathbb{E}, \mathbb{V}, \mathcal{E}$ when there is no confusion about the underlying distribution. With the definition of the inner product, we can verify that $\mathcal{E}_{f,g} = 1/2 \sum_{i,j} \pi_i T_{i,j}(f_i - f_j)(g_i - g_j)$. The Dirichlet Form can be viewed as the local variation of f because the variation of f can be expressed as $\mathbb{V}f = 1/2 \sum_{i,,j} \pi_i \pi_j (f_i - f_j)^2$ and $\mathcal{E}_{f,f} = 1/2 \sum_{i,,j} \pi_i T_{i,j}(f_i - f_j)^2$. The main reason to introduce Dirichlet Form is that we can compute the spectral gap from it.

Lemma A.1. $\gamma = inf\ \{\mathcal{E}_{f,f} : \mathbb{V}f = 1\}$

Proof. Note that the Dirichlet form $\mathcal{E}_{f,f} = \mathcal{E}_{f-c,f-c}$ for any constant function c. Then, it is equivalent to show

$$\gamma = inf\ \{\mathcal{E}_{f,f} : \mathbb{E}f = 0 \text{ and } \mathbb{V}f = 1\}.$$

Letting $\mathbb{E}f = 0$, we calculate

$$\begin{aligned}
\langle f, (I-T)f \rangle &= \langle f, f \rangle - \langle f, Tf \rangle \\
&= \left\langle \sum_k \langle f, x_k \rangle x_k, \sum_k \langle f, x_k \rangle x_k \right\rangle - \left\langle \sum_k \langle f, x_k \rangle x_k, \sum_k \langle f, x_k \rangle \lambda_k x_k \right\rangle \\
&= \sum_k |\langle f, x_k \rangle|^2 - \sum_k \lambda_k |\langle f, x_k \rangle|^2 \\
&= \sum_k |\langle f, x_k \rangle|^2 (1 - \lambda_k) \\
&\geq 1 - \lambda_2.
\end{aligned}$$

The inequality is followed from $\sum_k |\langle f, x_k \rangle|^2 = \langle f, f \rangle = \mathbb{V}f = 1$, $\lambda_1 = 1$, and noting that $\gamma = 1 - \lambda$. Equality is obtained if f is picked such that $\langle f, x_k \rangle = 0$ except $k = 2$. $\square$

We give the proof of Proposition 2.10.

Proof. To get the upper bound, we pick any set of subset S and let $f_i = \mathbb{1}_{\{i \in S\}}$. Direct calculation gives $\mathcal{E}_{f,f} = 1/2 \sum_{i,,j} \pi_i T_{i,j}(f_i - f_j)^2 = \sum_{i \in S,, j \in S^c} \pi_i T_{i,j} = q(S, S^c)$. By lemma A.1, $\gamma \leq q(S, S^c) = \frac{q(S,S^c)}{\pi(S)} \pi(S)$. We restrict attention to $\pi(S) \leq 1/2$, and obtain the upper bound $\gamma \leq 2\Phi$.

To get the lower bound, we do it in three steps. Let g be the eigenvector associated with the second eigenvalue, i.e. $Tg = \lambda_2 g$. Let $f_i = (g_i)^+$ and $S = \{i : f_i > 0\}$, where $x^+ \stackrel{\text{def}}{=} max\ \{x, 0\}$. Note that $\pi_S \leq 1/2$; if not, we take $f = (g_i)^-$, where $^-$ represents min. Rearrange i such that $f(i) \geq f(i+1)$, for all $i \in N$. Then, $f^2(i) \geq f^2(j)$ if and only if $i \leq j$.

Step 1: Let's verify that $\gamma \geq \frac{\mathcal{E}_{f,f}}{\langle f, f \rangle}$. Note that $\mathcal{E}_{f,f} = \langle f, (I-T)f \rangle$. If $f_i > 0$, then $((I-T)f)_i \leq ((I-T)g)_i = (1 - \lambda_2) g_i$. By linearity of $\langle \cdot, \cdot \rangle$ and that $f_i \geq 0$ for all $i \in N$, we get $\gamma \langle f, f \rangle \geq \mathcal{E}_{f,f}$.

Step 2: We verify that $\langle f, f \rangle \leq \Phi^{-2} \sum_{i \leq j} (f^2(i) - f^2(j)) q(i,j)$. Let $A = \{i : f^2(i) > t\}$ with $t > 0$. Since $\pi_A \leq \pi_S \leq 1/2$,

$$\Phi \leq \frac{q(A, A^c)}{\pi(A)} \leq \frac{\sum_{i \in A\ j \in A^c} \pi_i q(i,j)}{\pi(S)} = \frac{\sum_{i,j} \mathbb{1}_{\{f^2(j) \leq t \leq f^2(i)\}} q(i,j)}{\pi(S)}.$$

Hence,

$$\pi\{i : f^2(i) > t\} \leq \Phi^{-1} \sum_{i<j} \mathbb{1}_{\{f^2(j) \leq t \leq f^2(i)\}} q(i,j)$$

Using $\int \pi\{i : f^2(i) > t\} = \mathbb{E} f^2 = \sum_i f^2(i)\pi_i = \langle f, f\rangle$ and $\int \mathbb{1}_{\{f^2(j) \leq t < f^2(i)\}} = f^2(i) - f^2(j)$, we have

$$\langle f, f\rangle \leq \Phi^{-1} \sum_{i<j} (f^2(i) - f^2(j)) q(i,j).$$

Step 3: From step 2 and applying Cauchy-Schwarz inequality,

$$\begin{aligned}
\langle f, f\rangle^2 &\leq \Phi^{-2} \sum_{i<j} (f(i) + f(j))^2 q(i,j) \sum_{i<j} (f(i) - f(j))^2 q(i,j) \\
&= \Phi^{-2} \sum_{i<j} \{2f^2(i) + 2f^2(j) - [f(i) + f(j)]^2\} q(i,j) \sum_{i<j} (f(i) - f(j))^2 q(i,j) \\
&= \Phi^{-2} \sum_{i<j} (2\langle f, f\rangle - \mathcal{E}_{f,f}) \mathcal{E}_{f,f}.
\end{aligned}$$

The first equality follows from the identity $(f(i) + f(j))^2 = 2f^2(i) + 2f^2(j) - (f(i) + f(j))^2$. The last equality follows from $\mathcal{E}_{f,f} = 1/2 \sum_{i,,j} \pi_i T_{i,j} (f_i - f_j)^2$. Re-arranging,

$$\frac{\mathcal{E}_{f,f}}{\langle f, f\rangle} \geq \frac{\Phi^2}{2} \frac{\langle f, f\rangle}{\langle f, f\rangle - \mathcal{E}_{f,f}}.$$

Applying step 1 one right hand side and the fact that $1/(1 - \frac{1}{2}\gamma) > 1$,

$$\frac{\mathcal{E}_{f,f}}{\langle f, f\rangle} \geq \frac{\Phi^2}{2}.$$

And by step 1 again, we have the desired result $\gamma \geq \frac{\Phi^2}{2}$. □

Lemma 3.1 is useful to express consensus time with respect to one fixed ϵ, namely $\epsilon = 1/(2e)4$. We give its proof here.

Proof. It is sufficient to show that $ct(\epsilon; T) \leq log(\epsilon^{-1}) ct(1/(2e); T)$. Let T be the influence matrix, and recall that $cd(t) = max_{i \in N} ||T_i^t - \pi||_{tv}$. We verify that $cd(t+s) \leq 2cd(t)cd(s)$ directly from definitions of total variation distance and consensus distance.

$$\begin{aligned}
cd(t+s) &= \max_{i \in N} ||T_i^{t+s} - \pi||_{tv} \\
&= \max_{i \in N} \max_{S \subset N} |T_i^{t+s}(S) - \pi(S)| \\
&= \max_{i \in N} \frac{1}{2} \sum_{j \in N} |T^{t+s}(i,j) - \pi(j)| \\
&= \max_{i \in N} \frac{1}{2} \sum_{j \in N} |[\sum_{k \in N} T^t(i,k) T^s(k,j)] - (\pi T^s)(j)|
\end{aligned}$$

$$
\begin{aligned}
&= \max_{i \in N} \frac{1}{2} \sum_{j \in N} | \sum_{k \in N} T^t(i,k) T^s(k,j) - T^s(k,j)\pi(k)| \\
&\le 2 \max_{i \in N} \frac{1}{2} \sum_{j \in N} \max_{k \in N} T^s(k,j) \frac{1}{2} \sum_{k \in N} |T^t(i,k) - \pi(k)| \\
&= 2 \left[\max_{k \in N} 1/2 \sum_{j \in N} T^s(k,j) \right] \left[\max_{i \in N} 1/2 \sum_{k \in N} |T^t(i,k) - \pi(k)| \right] \\
&\le 2cd(t)cd(s).
\end{aligned}
$$

Let m be a sufficiently large integer, then for any ϵ',

$$
\begin{aligned}
cd(m\, cd(\epsilon'; T)) &\le (2cd(ct(\epsilon'; T)))^m \\
&\le (2\epsilon')^m.
\end{aligned}
$$

Letting $\epsilon' = 1/(2e)$, we have $cd(m\, cd(1/(2e); T)) \le e^{-m}$. And choosing an arbitrary small $\epsilon = e^{-m}$ by increasing m, we have the desired result

$$
ct(\epsilon; T) \le m\, ct(1/(2e); T) = log(\epsilon^{-1}) ct(1/(2e); T).
$$

□

A.3 Proofs of statements in Section 3

We will not provide the details of the proof of Theorem 3.6. Both of the proofs (Fountoulakis and Reed, 2008; Benjamini, Kozma, and Wormald, 2006) are quite involved, and we will briefly sketch out the steps taken by Fountoulakis and Reed.

Proof. The key is, as every other result in the study, to estimate the bottleneck ratio. But instead of direct estimation, the proofs first show that the giant component is a α-AN expander graph, define as follows

Definition A.2. A connected graph G is a α-AN expander if there is subgraph S such that

i. $\Phi_S \ge \alpha$, where Φ_S is the bottleneck ratio of the subgraph S.
ii. For any component $(S^c_k)$of the subgraph $G \backslash S$ is small in that $|k : e(S, S^c_i) > \lambda| \le \frac{d(G)}{2} e^{-\lambda\alpha}$.
iii. For any $i \in S$, i has edges with at most $1/\alpha$ different distinct S^c_i-s.

It is shown that in a α-AN graph G, $ct(n) \le C\alpha^{-6} \log^2(d(G)/2)$. Then, it is left to show that the giant component of the Erdös-Rényi graph is in fact a α-AN expander, for some constant $\alpha > 0$. □

The proof of Theorem 3.7 follows Cooper and Frieze (2003) and Durrett (2007). We observe two facts. First, the Erdös-Rényi network $G(n, p)$ with $p = \lambda \log n/n$ and $\lambda > 1$ is connected. A proof of this classical result can be found in Bollobás (2000). Second, both the minimum and maximum degree of the network is, *a.s.s.*, of order $\log n$; that is, there is a c_1 and c_2 such that smallest degree of a node is at least $c_1 \log n$ and the largest degree of a node is at most $c_2 \log n$. We will use formula (A.2) to estimate Φ. We further denote $d(S) \stackrel{\text{def}}{=} \sum_{i \in S} d_i$.

Proof. By Corollary 2.11, it is sufficient to show that exists a constant C such that $\Phi \geq C$. Using the fact that the degree of any vertex is of order $\log n$, we see that $\pi_{min} \geq \frac{c_1 \log n}{nc_2 \log n} \geq \frac{c_1}{c_2 n}$. Combining the constants, we get

$$ct(n) \leq C \log n.$$

We separate the estimations into two cases. Let $s \stackrel{\text{def}}{=} |S|$, where $S \subset N$. Case 1: $1 \leq s \leq \frac{n}{\lambda \log n}$. Let event $A = \{\exists S \text{ s.t. } e(S,S) \geq s \log \log n\}$. We bound the probability of A as follows,

$$\begin{aligned}
\mathbb{P}\{A\} &\leq \mathbb{P}\{\exists S \text{ s.t. } e(S,S) = s \log \log n\} \\
&= \binom{n}{s} \binom{\binom{s}{2}}{s \log \log n} p^{s \log \log n} \\
&\leq \left(\frac{en}{s}\right)^s \left(\frac{e\left(\frac{es}{2}\right)^2}{s \log \log n}\right)^{s \log \log n} p^{s \log \log n} \qquad \text{Using } \binom{n}{m} \leq (en/m)^m \\
&= \left(\frac{en}{s}\right)^s \left(\frac{1/4se^3 \lambda \log n}{s \log \log n}\right)^{s \log \log n} \\
&= C \exp\left\{\log(en) + s \log \log n \cdot \log\left[\frac{1/4\lambda e \log n}{n \log \log n}\right]\right\}.
\end{aligned}$$

Note that the function is decreasing in s, which can be verified by taking the derivative with respect to s. The right hand side is maximized at $s = 1$. The leading term is $\exp(-\log n(\log \log n)^2)$, which vanishes as $n \to \infty$. Finally, we use the fact that $d(S) \leq sc_2 \log n$, and $e(S,S^c) = [e(S,S^c) + e(S,S)] - e(S,S)$, which gives $e(S,S^c) \geq sc_1 \log n - s \log \log n \geq sC \log n$. Hence

$$\Phi = 1/2 \frac{e(S,S^c}{d(S)} \geq C.$$

Case 2: $\frac{n}{\lambda \log n} \leq s \leq \alpha n$, where $0 < \alpha < 1$. Note that we are not considering any set S with $s > \alpha n$ because for such a set S, $\pi(S) > 1/2$. To see that, consider the total degree of the network $d(N)$, which has a binomial distribution $2B(\binom{n}{2}, p)$, which has a mean $\lambda(n-1)\log n$ and a finite variance. And for a sufficiently large n, $d(S)$ is close to $\lambda(n-1)\log n$. Because the maximum degree is bounded above by $c_2 \log n$, there is a α such that for any S with $s > \alpha n$, $1/2d(S)/d(N) = \pi(S) > 1/2$.

The total degree $d(S) \leq sc_2 \log n$. And for the edges, note that the expected value

$$\mu \stackrel{\text{def}}{=} \mathbb{E}e(S,S^c) = \binom{n}{2} p \leq Cn \log n.$$

Using the Chernoff bound by considering each edge as a Bernoulli random variable, we bound the probability of the event $A = \{\exists S : e(S, S^c) \leq \mu/2\}$ by

$$\begin{aligned}\mathbb{P}(A) &\leq \binom{n}{s} e^{-C\mu} \\ &\leq C \left(\frac{en}{s}\right)^s \exp(-n\log n) \\ &= C\exp\{n\log(\frac{en}{n\log n}) - n\log n\},\end{aligned}$$

which tends to 0 as $n \to \infty$. Thus, we bound the the bottleneck ratio $\Phi = 1/2\frac{e(S,S^c)}{d(S)} \geq C$. □

We give the proof of Theorem 3.8.

Proof. The minimal degree of a vertex is exactly m because the last node note has only has m connections, and π_{min} is approximately $1/(2n)$, for large n (in fact, it is exactly $1/(2n)$ if $m = m_0$. By Corollary 2.11 and Lemma 3.1, if the Φ is bounded from below by a constant, then the $ct(n) \leq C\log n$.

Fix a $S \subset N$ with $\pi(S) \leq 1/2$. We verify the largest such set is at most a fraction of all the nodes, i.e. there is a constant $rho < 1$ such that $|S| \leq \rho|N|$. To see this, note that there is a constant p_k with $n^{1/15} > k \leq m$, the fraction of nodes with of degree k is about a non-vanishing constant, almost surely. This is the main result in Bollobás (2001), which explicitly calculates that $p_k = \frac{2m(m+1)}{(k+2)(k+1)k}$. The largest set S with $pi(S) \leq 1/2$ is constructed by taking the all the nodes with degree m, then those with degree $m+1$, $m+2$, and so forth, until the sum of degree of these nodes is close nd. Let K the constant be such that $\sum_{k=m}^{K} kp_k \geq nd$, then $\rho = \sum_{k=m}^{K} p_k$.

Observe an inequality between the total degree of set and its size: for any set $S \in N$, $\sum_{j\in S} d_j \leq 2m|S| + e(S, S^c)$, which can be verified by counting $2m|S|$ is the every node in S only connects its birth connection to nodes inside S and $e(S, S^c)$ is the number of additional degrees these nodes acquire from others. Let us define the *edge expansion constant* as $\alpha = min_{|S|\leq n/2}\frac{e(S,S^c)}{|S|}$. Then,

$$\begin{aligned}\Phi &= \min_{\pi(S)\leq 1/2} \frac{e(S,S^c)}{\sum_{j\in S} d_j} \\ &\geq \min_{|S|\leq n/2} \frac{e(S,S^c)}{2m|S| + e(S,S^c)} \\ &= \frac{\alpha}{m+\alpha}.\end{aligned}$$

The first inequality uses $\sum_{j\in S} d_j \leq 2m|S| + e(S, S^c)$. Then it is sufficient to bound α. Theorem 2 in Mihail, Papadimitrou, and Saberi (2004) shows that for k such that $m - 1 - 2k > 0$ and $m \geq 2$,

$$\mathbb{P}[\alpha < k] \to 0, \text{ as } n \to \infty.$$

This shows that α is bounded from below by a constant. By the observation above, Φ is also bounded from below by a constant. □

The proof of Theorem 3.9 will not be presented in full here. Interested readers can find the details in Addario-Berry and Lei (2012). We sketch out the overall argument.

Proof. This proof requires a different, more complicated consensus time bound than the one given by Corollary 2.11. The main result of Fountoulakis and Reed (2007) gives such a bound.

Theorem A.3. *Suppose that the network is connected. Let* $E \overset{\text{def}}{=} |e(N,N)|$ *be the number of connections in the network. Then for any small* $\epsilon > 0$*,*

$$ct(\epsilon; T) \leq C \log(\epsilon^{-1}) \sum_{k=1}^{\lceil \log_2 E \rceil} [\Phi(2^{-k})]^{-2},$$

where $\Phi(z) = \min\{\Phi(S) : S$ *is connected and* $zE \leq e(S,S) \leq 2zE\}$*.*

The proof of Theorem 3.9 divides the task of estimating the sequence of $\Phi(2^{-k})$ into two cases. Let $k \geq 1$, x_k be the solution to the equation $x/720 - \log(4(x+2k)) = 5$, and $M = k + 1 + 10\max\{x_k, \lambda\}$. Case 1 considers the bounds of the bottleneck ratio when $\lambda \geq x_k$, i.e. when the network has a large number of shortcuts. It is shown that for a large n, the following holds almost surely.

$$\sum_{k=1}^{\lceil \log_2 E \rceil} [\Phi(2^{-k})]^{-2} \leq \frac{4^6 M^2}{3} \log(n)^2 + \frac{6^4 M^2}{\lambda^2} \log_2 n,$$

which leads to

$$ct(n) \leq C \left(\frac{4^6 M^2}{3} \log(n)^2 + \frac{6^4 M^2}{\lambda^2} \log_2 n \right).$$

Case 2 considers the bounds of the bottleneck ratio when $0 < \lambda < x_k$. Let $R = \lceil \max\{k, \frac{2x_1}{\lambda}\} \rceil$. Let $\beta > 0$ be such that

$$\mathbb{P}\{\exists S \in N : |S| > (1-\beta)n \text{ and } e(S,S) \leq E) \leq (1-\beta)^n\}.$$

Then the following result is a key step toward bounding Φ. Let $i \in N$, and $B_{s,i}$ be the set of all $S \in N$ such that $i \in S$ and $|S| = s$. Define $B_s \overset{\text{def}}{=} \cup_{i \in N} B_{s,i}$.

Lemma A.4. *There is a* α *such that*

$$\mathbb{P}\{\exists S \in \cup_{R \log n \leq j \leq (1-\beta)n} B_j, e(S, S^c) \leq \alpha |S|\} \leq \frac{3R^3}{n^3}$$

Proof. See Lemma 15 in Addario-Berry and Lei (2012). □

Lemma A.4 is used to estimate the quantity $\sum_{k=1}^{\lceil \log_2 E \rceil} [\Phi(2^{-k})]^{-2}$. Similar to the case where $\lambda \geq c_k$, it is then shown

$$ct(n) \leq \sum_{k=1}^{\lceil \log_2 E \rceil} [\Phi(2^{-k})]^{-2} \leq C \log(n)^2,$$

for sufficiently large n. Combining the two cases give the desired result. $\square$

The representative agent theorem in Golub and Jackson (2011) (Theorem 2) offers a key insight about the dynamics of opinion formation. It purports that in the DeGroot process, opinions between agents of the same type converge quickly and the main obstacle to the overall convergence is the difference of opinion between agents of different types. To precisely state this notion, let is introduce a stochastic matrix of dimension $K \times K$ that is analogous to T. Let $e_{lk} = n_l n_k p_{lk}$, the expected number of connections between type l and type k. Let F be a stochastic matrix of dimension $K \times K$, where the lk-th element is

$$f_{lk} = \frac{e_{lk}}{\sum_{l \in K} e_{lk}}.$$

f_{lk} is the ratio of expected number of connections between l and k and the expected total number of connections that l has. In other words, f_{lk} measures how much influence type k has on type l. Recall that our definition of T allows each agent gives its own opinion the weight of $1/2$, so we are going to compare F with $\widetilde{T}$ (as defined in Section A.1).

The following proposition describes the spectral relationship between the group-level and individual-level influence matrix. The next two propositions are the main results in Golub and Jackson (2011). The $MT(\boldsymbol{n}, P)$ network are analyzed under a set of connectivity assumptions that are stronger than those assumed here. In fact, those assumptions are so stringent so that the resulting asymptotic networks always have the same convergence, up to a constant.

Proposition A.5. *Assume the connectivity assumptions as described in Definition 3 in Golub and Jackson (2011). Let the Let $\lambda_2(\widetilde{T}_n)$ and $\lambda_2(F)$ denote the second largest eigenvalue in absolute value for the respective matrix $\widetilde{T}_n$ and F. Then asymptotically almost surely, $\lambda_2(\widetilde{T}_n) \to \lambda_2(F)$.*

Proof. See Theorem 2 in Golub and Jackson (2011). $\square$

Proposition A.6. *Assume the connectivity assumptions as described in Definition 3 in Golub and Jackson (2011). Let the network be multi-type $MT(\boldsymbol{n}, P)$. The consensus time is a.s.s.*

$$\begin{aligned} i. \quad & ct(n) \le C \log n \frac{1}{1 - \lambda_2(F)}, \\ ii. \quad & ct(n) \le C \log n. \end{aligned}$$

Proof. By Proposition A.5 and formula A.1, we get the *a.s.s.* relation

$$\gamma(n) \to \frac{1}{1 - \frac{1}{2}(1 + \lambda_2(F))},$$

where $\gamma(n)$ denotes the spectral gap of the $MT(\boldsymbol{n}, P)$ network. By Proposition 2.8 and Lemma 3.1, we can bound consensus time as

$$
\begin{aligned}
ct(n) &\leq C \log(\pi_{min})\gamma(n) \\
&\leq C \log(\pi_{min}) \frac{1}{1 - \frac{1}{2}(1 + \lambda_2(F))}.
\end{aligned}
$$

Note that for any $k \in K$, island k is just a $ER(n_k, p_s)$ and by assumption ii, the minimum and maximum degree of each node is of order $\log n$ for the connected Erdös-Rényi network.[24] Then, $\log(\pi_{min}^{-1}) \leq C \log n$. By combining the constants, we get statement i. Observe that $\lambda_2(F)$ does not depend on n, so as the network grows, the term $\frac{1}{1-\lambda_2(F)}$ remains constant. Statement ii follows. $\square$

While statement i captures the intuition that the information in a representative influence matrix is sufficient to determine the speed of learning of realized networks, for a large n, its lack of dependence on n is troubling. In fact, we state statement i of Proposition A.6 because we want to argue that the assumptions made by Golub and Jackson restrict the type of speed of convergence that large networks can have. Their result cleanly links the second eigenvalue of the representative matrix to the random matrix, but in essence, they make enough assumptions so that almost all of the random networks do not have informational congestion, for sufficiently large n.

To prove the Theorem 3.11, we start with a lemma to show that we only need to consider the subsets that is an island, a subset that includes all agents of their own type.

Lemma A.7. *Consider the network In the $IM(n, K, p_s, p_d)$. For a sufficiently large n, with probability one Φ is attained at the critical subgroup S such that S includes all members of type k and only members of type k, for some $k \in K$. In other words, any island forms a critical subgroup.*

Proof. Let us use use the formula (A.2), where $\Phi_T = 1/2 \underset{\pi(S)\leq 1/2}{min} \frac{e(S,S^c)}{\sum_{i\in S} d_i}$. We need to consider two cases. Case #1, let S be the group that includes every member of type k, and $\bar{S}$ be the group that includes only a fraction (β) of agents of type k, for some type $k \in K$, where $0 < \beta < 1$. Then, $\Phi(\bar{S}) < \Phi(S)$ if and only if $\frac{e(\bar{S},\bar{S}^c)}{d(\bar{S})} < \frac{e(S,S^c)}{d(S)}$. By the law of large number, both the number of cross edges $(e(S, S^c))$ and total number of degree in S $(d(S))$ approach the number their expected value because each agent in the island model $IM(n, K, p_s, p_d)$ forms connections independently from each other with the probability law. That is, with probability one,

$$
\begin{aligned}
e(\bar{S}, \bar{S}^c) &= \beta n_k(1 - \beta)n_k p_s + \beta n_k(K - 1)n_k p_d\,,\ d(\bar{S}) = \beta n_k L\,, \\
e(S, S^c) &= \binom{n_k}{2} p_s + n_k(K - 1)n_k p_d\,,\ \text{and } d(S) = n_k L\,,
\end{aligned}
$$

where L is the expected number of edges of one agent. It follows that $\frac{e(\bar{S},\bar{S}^c)}{d(\bar{S})} > \frac{e(S,S^c)}{d(S)}$ with probability one. Case #2, we compare a mixed group with a group with only one type of agents. Let $\bar{S}$ be

[24] See Bollobás (2000) for a proof.

composed of exactly two types of agent, of fraction β_1 and β_2 of their respective group. Note that the island model assumes that $n_l = n_k$, for $l, k \in K$. Let $\beta_1 > \beta_2$. As argued in the previous case, with probability one,

$$\begin{aligned} e(\bar{S}, \bar{S}^c) &= \beta_1 n_k (k - \beta_1) n_k p_s + \beta_1 n_k [(K-2)n_k + (1-\beta_2)n_k] p_d \\ &\quad + \beta_2 n_k (k - \beta_2) n_k p_s + \beta_2 n_k [(K-2)n_k + (1-\beta_1)n_k] p_d \,, \\ d(\bar{S}) &= (\beta_1 + \beta_2) n_k L \,, \\ e(S, S^c) &= [\beta_1 n_k (1 - beta_1) n_k] p_s + \beta_1 n_k (K-1) n_k p_d \,, \\ d(\bar{S}) &= \beta_1 n_k L \,. \end{aligned}$$

Then, $\frac{e(\bar{S}, \bar{S}^c)}{d(\bar{S})} > \frac{e(S, S^c)}{d(S)}$, and so restricting to one unique type always induce a $\Phi(S)$ that is less than $\Phi(\bar{S})$, where $\bar{S}$ is a group of two mixed types. If the group has more than two types, then repeat the argument to show that group (S) with a unique type always lead to the smaller bottleneck ratio. Since number of types are assumed to be constant (K), unique type is a necessary condition for the group S to be critical. The two cases completes the proof of the lemma. □

Now we tackle the theorem.

Proof. For the case of $\frac{p_d}{p_s} = \lambda$, with $0 < \lambda < 1$. We simply compute the following,

$$\begin{aligned} e(S, S^c 0 &= \frac{1}{2} \frac{e(S, S^c)}{d(S)} \\ &= \frac{1}{2} \frac{n_k (K-1) n_k p_d}{2\binom{n_k}{2} p_s + n_k (K-1) n_k p_d} \\ &= \frac{1}{2} \frac{n_k^2}{C n_k^2 \frac{p_s}{p_d} + n_k^2} \\ &\geq C \frac{1}{\frac{1}{\lambda} + 1} . \end{aligned}$$

The last inequality is obtained from the assumption $\frac{p_d}{p_s} = \lambda$ and by absorbing some constants. This shows that Φ is bounded from below by a constant. And by the same argument in the proof of Theorem 3.6, this is sufficient to show that the upper bound for $ct(n)$ is of order $\log n$.

For the case of $\frac{p_d}{p_s} = \frac{\lambda}{n^a}$, with $\lambda > 0$ and $a > 0$. A similar calculation shows that

$$\begin{aligned} e(S, S^c 0 &= \frac{1}{2} \frac{n_k^2}{C n_k^2 \frac{p_s}{p_d} + n_k^2} \\ &\geq C \frac{1}{\frac{n^a}{\lambda} + 1} \\ &\geq C n^{-a} . \end{aligned}$$

By Corollary 2.11, we get $ct(n) \leq C n^{2a} \log n$. This completes the proof. □

A.4 Proofs of statements in Section 4

The proof of Proposition 4.5 also shows that there is no need to arrange the sequence of rewiring in any special way in order to achieve a stable network. Any sequence of congestion-reducing rewiring lead to a stable outcome.

Proof. T is not stable, and let $S \subset N$ be the set of agents who is able reduce congestion. $S \neq \varnothing$. We want to show that we can always remove one agent from S by a rewiring, thus downsizing $|S|$. The following lemma helps us to get to that goal. Let $i \in S$.

Lemma A.8. *Let T' be modification of T by i's rewiring. Then, i does not rewire any of the connection with an agent $k \in S^c$.*

Proof. Suppose not, then there is a subset $A \subset N$ such that $i, k \in A$, $\pi_A \leq 1/2$, and

$$\Phi(A; T') > \Phi(A; T) = \Phi^*(i; T).$$

But then note that $k \in A$, and it follows that k is also able to reduce congestion. □

With this observation, we see that if there are congestion reducing rewirings available, then by one rewiring, we also remove one agent from the set S. The network is finite, and so there is a stable network at the end of the rewiring process. □

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2007): "Learning and disagreement in an uncertain world," Working paper.

Acemoglu, D., and A. Ozdaglar (2010): "Opinion dynamics and learning in social networks," *Dynamic games application*, 1, 3–49.

Addario-Berry, L., and T. Lei (2012): "The mixing time of the Newman-Watts small world," ArXiv math.PR.

Alon, N., and V. Millman (1985): "λ_1,isoperimetric inequalities for graphs, and superconcentrators," *Journal of Combinatorial Theory, Series B*, 38(1), 73–88.

Bala, V., and S. Goyal (1998): "Learning from neighbours," *The Review of Economic Studies*, 65(3), 595–621.

Banerjee, A. V. (1992): "A simple model of herd behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

Benjamini, I., G. Kozma, and N. Wormald (2006): "The mixing time of the giant component of a random graph," ArXiv math.PR.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 100(5), 992–1026.

Bollobás, B. (2000): *Modern Graph Theory.* Springer.

——— (2001): "The degree sequence of a scale-free random graph process," *Random Structures and Algorithms*, 18(3), 279–290.

Chandrasekhar, A., H. Larreguy, and J. P. Xandri (2010): "Testing models of social learning on networks: evidence from a lab experiment in the field," Working paper.

Cheeger, J. (1969): "A lower bound for the smallest eigenvalue of the Laplacian," *Problems in Analysis*, Papers dedicated to Salomon Bochner, 195–199.

Cooper, C., and A. Frieze (2003): "The cover time of sparse random graphs," Working paper.

DeGroot, M. (1974): "Reaching a consensus," *Journal of the American Statistical Association*, 69(345), 118–121.

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Uni-Dimensional Opinions," *Quarterly Journal of Economics*, 118(3), 909–968.

Durrett, R. (2007): *Random Graph Dynamics.* Cambridge University Press. New York, NY.

Fountoulakis, N., and B. Reed (2007): "Faster mixing and small bottleneck," *Probability Theory and Related Fields*, 137, 475–486.

——— (2008): "The Evolution of the Mixing Rate," *Random Structures and Algorithms*, 33, 68–86.

Golub, B., and M. Jackson (2010): "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2(1), 112–149.

——— (2011): "How homophily affects the speed of learning and best-response dynamics," Working paper.

Jackson, M. (2008): *Social and Economic Networks.* Princeton University Press.

Levin, D., Y. Peres, and E. Wilmer (2009): *Markov Chains and Mixing Times.* AMS.

Lorenz, J. (2005): "A stabilization theorem for dynamics of continuous opinions," *Physica A: Statistical Mechanics and its Applications*, 355(1), 217–223.

Marsden, P. (1987): "Core Discussion Networks of Americans," *American Sociological Review*, 52, 122–131.

MEYER, C. D. (2000): *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA: Scoiety for Inudstrial and Applied Mathematics.

MIHAIL, M., C. PAPADIMITROU, AND A. SABERI (2004): "On certain connectivity properties of the Internet topology," *In Proceedings of 44th Annual IEEE Symposium on Foundation of Computer Science 2003*, pp. 28–35.

MONTENEGRO, R., AND P. TETALI (2006): "Mathematical Aspects of Mixing Times in Markov Chains," *Foundations and Trends in Theoretical Computer Science*, 1(3), 237–354.

SMITH, L., AND P. SORENSEN (2000): "Pathological outcomes of observational learning," *Econometrica*, 68(2), 371–398.