



# Spread of (mis)information in social networks<sup>☆</sup>

Daron Acemoglu<sup>a,\*</sup>, Asuman Ozdaglar<sup>b</sup>, Ali ParandehGheibi<sup>b</sup>

<sup>a</sup> Department of Economics, Massachusetts Institute of Technology, United States

<sup>b</sup> Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, United States

## ARTICLE INFO

### Article history:

Received 5 May 2009

Available online 1 February 2010

### JEL classification:

C72

D83

### Keywords:

Information aggregation

Learning

Misinformation

Social networks

## ABSTRACT

We provide a model to investigate the tension between information aggregation and spread of misinformation. Individuals meet pairwise and exchange information, which is modeled as both individuals adopting the average of their pre-meeting beliefs. “Forceful” agents influence the beliefs of (some of) the other individuals they meet, but do not change their own opinions. We characterize how the presence of forceful agents interferes with information aggregation. Under the assumption that even forceful agents obtain some information from others, we first show that all beliefs converge to a stochastic consensus. Our main results quantify the extent of misinformation by providing bounds or exact results on the gap between the consensus value and the benchmark without forceful agents (where there is efficient information aggregation). The worst outcomes obtain when there are several forceful agents who update their beliefs only on the basis of information from individuals that have been influenced by them.

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Individuals form beliefs on various economic, political and social variables (“state”) based on information they receive from others, including friends, neighbors and coworkers as well as local leaders, news sources and political actors. A key tradeoff faced by any society is whether this process of information exchange will lead to the formation of more accurate beliefs or to certain systematic biases and spread of misinformation. A famous idea going back to Condorcet’s Jury Theorem (now often emphasized in the context of ideas related to “wisdom of crowds”) encapsulates the idea that exchange of dispersed information will enable socially beneficial aggregation of information. However, as several examples ranging from the effects of the Swift Boat ads during the 2004 presidential campaign to the beliefs in the Middle East that 9/11 was a US or Israeli conspiracy illustrate, in practice social groups are often swayed by misleading ads, media outlets, and political leaders, and hold on to incorrect and inaccurate beliefs.

A central question for social science is to understand the conditions under which exchange of information will lead to the spread of misinformation (incorrect information) instead of aggregation of dispersed information. In this paper, we take a first step towards developing and analyzing a framework for providing answers to this question. While the issue of misinformation can also be studied using Bayesian models, non-Bayesian models appear to provide a more natural starting point.<sup>1</sup> Our modeling strategy is therefore to use a non-Bayesian model, approximating a Bayesian learning process in the

<sup>☆</sup> We would like to thank Pablo Parrilo, Devavrat Shah, Matthew Jackson and the anonymous referees for useful comments and discussions. This research is supported by NSF and AFOSR.

\* Corresponding author.

E-mail address: [Daron@MIT.edu](mailto:Daron@MIT.edu) (D. Acemoglu).

<sup>1</sup> In particular, misinformation can arise in a Bayesian model if an agent (receiver) is unsure of the type of another agent (sender) providing her with information and the sender happens to be of a type intending to mislead the receiver. Nevertheless, this type of misinformation will be limited since if the probability that the sender is of the misleading type is high, the receiver will not change her beliefs much on the basis of the sender’s communication.

absence of “forceful” agents (who are either trying to mislead or influence others or are, for various rational or irrational reasons, not interested in updating their opinions).

We consider a society envisaged as a social network of  $n$  agents, communicating and exchanging information. Specifically, each agent is interested in learning some underlying state  $\theta \in \mathbb{R}$ . For example, the underlying state  $\theta$  might represent the optimal value of some individual action or social choice, and the loss to individual (or society) is an increasing function of the gap between the actual choice and  $\theta$ . Each agent receives a signal  $x_i(0) \in \mathbb{R}$  in the beginning. We assume that  $\theta = 1/n \sum_{i=1}^n x_i(0)$ , for example, because each signal is an independent draw from some distribution with mean  $\theta$ . The important point for our purposes is that information about the relevant state  $\theta$  is dispersed but can be easily aggregated if the agents can communicate in a centralized or decentralized fashion.

Information exchange between agents results from bilateral communication. In particular, each individual is “recognized” according to a Poisson process in continuous time and conditional on this event, meets one of the individuals in her *social neighborhood* according to a pre-specified stochastic process. We think of this stochastic process as representing an underlying *social network* (for example, friendships, information networks, etc.). Following this meeting, there is a potential exchange of information between the two individuals, affecting the beliefs of one or both agents. We distinguish between two types of individuals: *regular* or *forceful*. When two regular agents meet, they update their beliefs to be equal to the average of their pre-meeting beliefs. This structure, though non-Bayesian, has a simple and appealing interpretation and ensures the convergence of beliefs to the underlying state  $\theta$  when the society consists only of regular agents.<sup>2</sup> In contrast, when an agent meets a forceful agent, this may result in the forceful agent “influencing” his beliefs so that this individual inherits the forceful agent’s belief except for an  $\epsilon$  weight on his pre-meeting belief.<sup>3</sup>

Our model of forceful agents is sufficiently general to nest several relevant cases. First, forceful agents might correspond to individuals that are stubborn and do not easily update their beliefs. Second, they may correspond to individuals (or media outlets) that purposefully wish to influence others with their opinion. Third, they may correspond to individuals, such as community leaders, that have greater influence with some subset of the population.<sup>4</sup> Fourth, the interaction between forceful and regular agents may represent potential “communication failures,” whereby instead of the typical exchange of information, these interactions introduce a degree of bias towards the beliefs of forceful agents. In all cases, a key assumption of our analysis is that even forceful agents engage in some updating of their beliefs (even if infrequently) as a result of exchange of information with their own social neighborhoods. This assumption captures the intuitive notion that “no man is an island” and thus receives some nontrivial input from the social context in which he or she is situated.<sup>5</sup> The influence pattern of social agents superimposed over the social network can be described by directed links, referred to as *forceful links*, and creates a richer stochastic process, representing the evolution of beliefs in the society. Both with and without forceful agents, the evolution of beliefs can be represented by a Markov chain and our analysis will exploit this connection. We will frequently distinguish the Markov chain representing the evolution of beliefs and the Markov chain induced by the underlying social network (i.e., just corresponding to the communication structure in the society, without taking into account the influence pattern) and properties of both will play a central role in our results.

Our objective is to characterize the evolution of beliefs and quantify the effect of forceful agents on public opinion in the context of this model. Our first result is that, despite the presence of forceful agents, the opinion of all agents in this social network converges to a common, though stochastic, value under weak regularity conditions. More formally, each agent’s opinion converges to a value given by  $\bar{x} = \pi'x(0)$ , where  $x(0)$  is the vector of initial beliefs and  $\pi$  is a random vector. The limiting value of opinions in this social network  $\bar{x}$  is a random variable because it depends on the stochastic order in which bilateral exchanges take place.

Given this convergence result, we define our measure of spread of misinformation in this society as the gap between the expected value of limiting opinions  $\bar{x}$  and  $\theta$ , that is,  $\bar{\pi}'x(0) - \theta = \sum_{i=1}^n (\bar{\pi}_i - 1/n)x_i(0)$ , where  $\bar{\pi}$  is the expected value of  $\pi$  and  $\bar{\pi}_i$  denotes its  $i$ th component.<sup>6</sup> The greater is this gap, the greater is the potential for misinformation in this society. Moreover, this formula also makes it clear that  $\bar{\pi}_i - 1/n$  gives the *excess influence* of agent  $i$ . Our strategy will be to develop bounds on the spread of misinformation in the society (as defined above) and on the excess influence of each agent for general social networks and also provide exact results for some special networks.

<sup>2</sup> The appealing interpretation is that this type of averaging would be optimal if both agents had independent beliefs drawn from a normal distribution with mean equal to the underlying state and equal precision. This interpretation is discussed in detail in DeMarzo et al. (2003) in a related context. This interpretation also provides another justification for thinking of the underlying state as corresponding to the average of the initial signals of the  $n$  agents in the society.

<sup>3</sup> When  $\epsilon = 1/2$ , then the individual treats the forceful agent just as any other regular agent (is not influenced by him over and above the information exchange) and the only difference from the interaction between two regular agents is that the forceful agent himself does not update his beliefs. All of our analysis is conducted for arbitrary  $\epsilon$ , so whether forceful agents are also “influential” in pairwise meetings is not important for any of our findings.

<sup>4</sup> What we do not allow are individuals who know the underlying state and try to convince others of some systematic bias relative to the underlying state, though the model could be modified to fit this possibility as well.

<sup>5</sup> When there are several forceful agents and none of them ever changes his or her opinion, then it is straightforward to see that opinions in this society will never settle into a “stationary” distribution provided that the initial opinions of such agents are different. While this case is also interesting to study, it is significantly more difficult to analyze and requires a different mathematical approach.

<sup>6</sup> Other moments of the stochastic distribution of the gap between  $\bar{x}$  and  $\theta$  are also of economic interest, but are more difficult to characterize.

The remainder of the paper provides three sets of results. First, using tools from matrix perturbation theory,<sup>7</sup> we provide global and general upper bounds on the extent of misinformation as a function of the properties of the underlying social network. In particular, the bounds relate to the *spectral gap* and the mixing properties of the Markov chain induced by the social network. Recall that a Markov chain is *fast-mixing* if it converges rapidly to its stationary distribution. It will do so when it has a large spectral gap, or loosely speaking, when it is highly connected and possesses many potential paths of communication between any pair of agents. Intuitively, societies represented by fast-mixing Markov chains have more limited room for misinformation because forceful agents themselves are influenced by the weighted opinion of the rest of the society before they can spread their own (potentially extreme) views. A corollary of these results is that for a special class of societies, corresponding to “expander graphs,” misinformation disappears in large societies provided that the number of forceful agents does not grow as fast as the total number of nodes and no forceful agent has global impact.<sup>8</sup> In contrast, the extent of misinformation can be substantial in slow-mixing Markov chains, also for an intuitive reason. Societies represented by such Markov chains would have a high degree of partitioning (multiple clusters with weak communication in between), so that forceful agents receive their information from others who previously were influenced by them, ensuring that their potentially extreme opinions are never moderated.<sup>9</sup>

Our second set of results exploit the local structure of the social network in the neighborhood of each forceful agent in order to provide a tighter characterization of the extent of misinformation and excess influence. Fast-mixing and spectral gap properties are global (and refer to the properties of the overall social network representing meeting and communication patterns among all agents). As such, they may reflect properties of a social network far from where the forceful agents are located. If so, our first set of bounds will not be tight. To redress this problem, we develop an alternative analysis using mean (first) passage times of the Markov chain and show how it is not only the global properties of the social network, but also the local social context in which forceful agents are situated that matter. For example, in a social network with a single dense cluster and several non-clustered pockets, it matters greatly whether forceful links are located inside the cluster or not. We illustrate this result sharply by first focusing on graphs with *forceful bridges*, that is, graphs representing societies in which a single forceful link connects two otherwise disconnected components. This, loosely speaking, represents a situation in which a forceful agent, for example a media outlet or a political party leader, obtains all of its (or his or her) information from a small group of individuals and influences the rest of the society. In this context, we establish the surprising result that all members of the small group will have the same excess influence, even though some of them may have much weaker links or no links to the forceful agent. This result is an implication of the society having a (single) forceful bridge and reflects the fact that the information among the small group of individuals who are the source of information of the forceful agent aggregates rapidly and thus it is the average of their beliefs that matter. We then generalize these results and intuitions to more general graphs using the notion of *information bottlenecks*.

Our third set of results are more technical in nature, and provide new conceptual tools and algorithms for characterizing the role of information bottlenecks. In particular, we introduce the concept of *relative cuts* and present several new results related to relative cuts and how these relate to mean first passage times. For our purposes, these new results are useful because they enable us to quantify the extent of local clustering around forceful agents. Using the notion of relative cuts, we provide improved bounds on the extent of misinformation in beliefs as a function of information bottlenecks in the social network.

Our paper is related to a large and growing learning literature. Much of this literature focuses on various Bayesian models of observational or communication-based learning; for example Bikchandani et al. (1992), Banerjee (1992), Smith and Sorensen (2000, 1998), Banerjee and Fudenberg (2004), Bala and Goyal (1998, 2001), Gale and Kariv (2003), and Celen and Kariv (2004b, 2004a). These papers develop models of social learning either using a Bayesian perspective or exploiting some plausible rule-of-thumb behavior. Acemoglu et al. (2008) provide an analysis of Bayesian learning over general social networks. Our paper is most closely related to DeGroot (1974), DeMarzo et al. (2003) and Golub and Jackson (forthcoming, 2008), who also consider non-Bayesian learning over a social network represented by a connected graph.<sup>10</sup> None of the papers mentioned above consider the issue of the spread of misinformation (or the tension between aggregation of information and spread of misinformation), though there are close parallels between Golub and Jackson's and our characterizations of influence. In particular, Golub and Jackson (2008) characterize the effects of homophily on learning and influence in two different models of learning in terms of mixing properties and the spectral gap of graphs. In one of their learning models, which builds on DeGroot (1974), DeMarzo et al. (2003) and Golub and Jackson (forthcoming), homophily has negative effects on learning (and speed of learning) for reasons related to our finding that in slow-mixing graphs, misinformation can spread more. In addition to our focus, the methods of analysis here, which develop bounds on the extent of misinformation

<sup>7</sup> In particular, we decompose the transition matrix of the Markov chain into a doubly stochastic matrix, representing the underlying social network, and a remainder matrix, representing a directed influence graph. Despite the term “perturbation,” this remainder matrix need not be “small” in any sense.

<sup>8</sup> Expander graphs are graphs whose spectral gap remains bounded away from zero as the number of nodes tends to infinity. Several social network models, such as small world graphs and preferential attachment graphs correspond to expander graphs.

<sup>9</sup> This result is related to Golub and Jackson (2008), where they link learning to homophily properties of the social network.

<sup>10</sup> An important distinction is that in contrast to the “averaging” model used in these papers, we have a model of pairwise interactions. We believe that this model has a more attractive economic interpretation, since it does not have the feature that neighbors' information will be averaged at each date (even though the same information was exchanged the previous period). In contrast, in the pairwise meeting model (without forceful agents), if a pair meets two periods in a row, in the second meeting there is no information to exchange and no change in beliefs takes place.

and provide exact characterization of excess influence in certain classes of social networks, are entirely new in the literature and also rely on the developments of new results in the analysis of Markov chains.

Our work is also related to other work in the economics of communication, in particular, to cheap-talk models based on Crawford and Sobel (1982) (see also Farrell and Gibbons, 1989 and Sobel, 2009), and some recent learning papers incorporating cheap-talk games into a network structure (see Ambrus and Takahashi, 2008; Hagenbach and Koessler, 2009; Galeotti et al., 2009).

In addition to the papers on learning mentioned above, our paper is related to work on consensus, which is motivated by different problems, but typically leads to a similar mathematical formulation (Tsitsiklis, 1984; Tsitsiklis et al., 1986; Jadbabaie et al., 2003; Olfati-Saber and Murray, 2004; Olshevsky and Tsitsiklis, forthcoming; Nedić and Ozdaglar, forthcoming; Fagnani and Zampieri, forthcoming). In consensus problems, the focus is on whether the beliefs or the values held by different units (which might correspond to individuals, sensors or distributed processors) converge to a common value. Our analysis here does not only focus on consensus, but also whether the consensus happens around the true value of the underlying state. There are also no parallels in this literature to our bounds on misinformation and characterization results.

The rest of this paper is organized as follows: In Section 2, we introduce our model of interaction between the agents and describe the resulting evolution of individual beliefs. We also state our assumptions on connectivity and information exchange between the agents. Section 3 presents our main convergence result on the evolution of agent beliefs over time. In Section 4, we provide bounds on the extent of misinformation as a function of the global network parameters. Section 5 focuses on the effects of location of forceful links on the spread of misinformation and provides bounds as a function of the local connectivity and location of forceful agents in the network. Section 6 contains our concluding remarks.

**Notation and terminology:** A vector is viewed as a column vector, unless clearly stated otherwise. We denote by  $x_i$  or  $[x]_i$  the  $i$ th component of a vector  $x$ . When  $x_i \geq 0$  for all components  $i$  of a vector  $x$ , we write  $x \geq 0$ . For a matrix  $A$ , we write  $A_{ij}$  or  $[A]_{ij}$  to denote the matrix entry in the  $i$ th row and  $j$ th column. We write  $x'$  to denote the transpose of a vector  $x$ . The scalar product of two vectors  $x, y \in \mathbb{R}^m$  is denoted by  $x'y$ . We use  $\|x\|_2$  to denote the standard Euclidean norm,  $\|x\|_2 = \sqrt{x'x}$ . We write  $\|x\|_\infty$  to denote the max norm,  $\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|$ . We use  $e_i$  to denote the vector with  $i$ th entry equal to 1 and all other entries equal to 0. We denote by  $e$  the vector with all entries equal to 1.

A vector  $a$  is said to be a *stochastic vector* when  $a_i \geq 0$  for all  $i$  and  $\sum_i a_i = 1$ . A square matrix  $A$  is said to be a (*row*) *stochastic matrix* when each row of  $A$  is a stochastic vector. The transpose of a matrix  $A$  is denoted by  $A'$ . A square matrix  $A$  is said to be a *doubly stochastic matrix* when both  $A$  and  $A'$  are stochastic matrices.

## 2. Belief evolution

### 2.1. Description of the environment

We consider a set  $\mathcal{N} = \{1, \dots, n\}$  of agents interacting over a social network. Each agent  $i$  starts with an initial belief about an underlying state, which we denote by  $x_i(0) \in \mathbb{R}$ . Agents exchange information with their neighbors and update their beliefs. We assume that there are two types of agents: *regular* and *forceful*. Regular agents exchange information with their neighbors (when they meet). In contrast, forceful agents influence others disproportionately.

We use an asynchronous continuous-time model to represent meetings between agents (also studied in Boyd et al., 2005 in the context of communication networks). In particular, we assume that each agent meets (communicates with) other agents at instances defined by a rate one Poisson process independent of other agents. This implies that the meeting instances (over all agents) occur according to a rate  $n$  Poisson process at times  $t_k$ ,  $k \geq 1$ . Note that in this model, by convention, at most one node is active (i.e., is meeting another) at a given time. We discretize time according to meeting instances (since these are the relevant instances at which the beliefs change), and refer to the interval  $[t_k, t_{k+1})$  as the  $k$ th time slot. On average, there are  $n$  meeting instances per unit of absolute time (see Boyd et al., 2005 for a precise relation between these instances and absolute time). Suppose that at time slot  $k$ , agent  $i$  is chosen to meet another agent (probability  $1/n$ ). In this case, agent  $i$  will meet agent  $j \in \mathcal{N}$  with probability  $p_{ij}$ . Following a meeting between  $i$  and  $j$ , there is a potential exchange of information. Throughout, we assume that all events that happen in a meeting are *independent of any other event that happened in the past*. Let  $x_i(k)$  denote the belief of agent  $i$  about the underlying state at time slot  $k$ . The agents update their beliefs according to one of the following three possibilities.

- (i) Agents  $i$  and  $j$  reach pairwise consensus and the beliefs are updated according to

$$x_i(k+1) = x_j(k+1) = \frac{x_i(k) + x_j(k)}{2}.$$

We denote the conditional probability of this event (conditional on  $i$  meeting  $j$ ) as  $\beta_{ij}$ .

- (ii) Agent  $j$  influences agent  $i$ , in which case for some  $\epsilon \in (0, 1/2]$ , beliefs change according to

$$x_i(k+1) = \epsilon x_i(k) + (1 - \epsilon)x_j(k) \quad \text{and} \quad x_j(k+1) = x_j(k). \quad (1)$$

In this case beliefs of agent  $j$  do not change.<sup>11</sup> We denote the conditional probability of this event as  $\alpha_{ij}$ , and refer to it as the *influence probability*. Note that we allow  $\epsilon = 1/2$ , so that agent  $i$  may be treating agent  $j$  just as a regular agent, except that agent  $j$  himself does not change his beliefs.

(iii) Agents  $i$  and  $j$  do not agree and stick to their beliefs, i.e.,

$$x_i(k+1) = x_i(k) \quad \text{and} \quad x_j(k+1) = x_j(k).$$

This event has probability  $\gamma_{ij} = 1 - \beta_{ij} - \alpha_{ij}$ .

Any agent  $j$  for whom the influence probability  $\alpha_{ij} > 0$  for some  $i \in \mathcal{N}$  is referred to as a *forceful agent*. Moreover, the directed link  $(j, i)$  is referred to as a *forceful link*.<sup>12</sup>

As discussed in the introduction, we can interpret forceful agents in multiple different ways. First, forceful agents may correspond to community leaders or news media, will have a disproportionate effect on the beliefs of their followers. In such cases, it is natural to consider  $\epsilon$  small and the leaders or media not updating their own beliefs as a result of others listening to their opinion. Second, forceful agents may be indistinguishable from regular agents, and thus regular agents engage in what they think is information exchange, but forceful agents, because of stubbornness or some other motive, do not incorporate the information of these agents in their own beliefs. In this case, it may be natural to think of  $\epsilon$  as equal to  $1/2$ . The results that follow remain valid with either interpretation.

The influence structure described above will determine the evolution of beliefs in the society. Below, we will give a more precise separation of this evolution into two components, one related to the underlying social network (communication and meeting structure), and the other to influence patterns.

## 2.2. Assumptions

We next state our assumptions on the belief evolution model among the agents. We have the following assumption on the agent meeting probabilities  $p_{ij}$ .

**Assumption 1** (*Meeting probabilities*).

- (a) For all  $i$ , the probabilities  $p_{ii}$  are equal to 0.
- (b) For all  $i$ , the probabilities  $p_{ij}$  are nonnegative for all  $j$  and they sum to 1 over  $j$ , i.e.,

$$p_{ij} \geq 0 \quad \text{for all } i, j, \quad \sum_{j=1}^n p_{ij} = 1 \quad \text{for all } i.$$

Assumption 1(a) imposes that “self-communication” is not a possibility, though this is just a convention, since, as stated above, we allow disagreement among agents, i.e.,  $\gamma_{ij}$  can be positive. We let  $P$  denote the matrix with entries  $p_{ij}$ . Under Assumption 1(b), the matrix  $P$  is a *stochastic matrix*.<sup>13</sup>

We next impose a connectivity assumption on the social network. This assumption is stated in terms of the directed graph  $(\mathcal{N}, \mathcal{E})$ , where  $\mathcal{E}$  is the set of directed links induced by the positive meeting probabilities  $p_{ij}$ , i.e.,

$$\mathcal{E} = \{(i, j) \mid p_{ij} > 0\}. \quad (2)$$

**Assumption 2** (*Connectivity*). The graph  $(\mathcal{N}, \mathcal{E})$  is strongly connected, i.e., for all  $i, j \in \mathcal{N}$ , there exists a directed path connecting  $i$  to  $j$  with links in the set  $\mathcal{E}$ .

Assumption 2 ensures that every agent “communicates” with every other agent (possibly through multiple links). This is not an innocuous assumption, since otherwise the graph  $(\mathcal{N}, \mathcal{E})$  (and the society that it represents) would segment into multiple non-communicating parts. Though not innocuous, this assumption is also natural for several reasons. First, the evidence suggests that most subsets of the society are not only connected, but are connected by means of several links (e.g., Watts, 2003; Jackson, 2008), and the same seems to be true for indirect linkages via the Internet. Second, if the society is segmented into multiple non-communication parts, the insights here would apply, with some modifications, to each of these parts.

<sup>11</sup> We could allow the own-belief weight  $\epsilon$  to be different for each agent  $i$ . This generality does not change the results or the economic intuitions, so for notational convenience, we assume this weight to be the same across all agents.

<sup>12</sup> We refer to directed links/edges as links and undirected ones as edges.

<sup>13</sup> That is, its row sums are equal to 1.

Let us also use  $d_{ij}$  to denote the length of the shortest path from  $i$  to  $j$  and  $d$  to denote the *maximum shortest path length* between any  $i, j \in \mathcal{N}$ , i.e.,

$$d = \max_{i, j \in \mathcal{N}} d_{ij}. \quad (3)$$

In view of Assumption 2, these are all well-defined objects.

Finally, we introduce the following assumption which ensures that there is positive probability that every agent (even if he is forceful) receives some information from an agent in his neighborhood.

**Assumption 3** (*Interaction probabilities*). For all  $(i, j) \in \mathcal{E}$ , the sum of the *averaging probability*  $\beta_{ij}$  and the *influence probability*  $\alpha_{ij}$  is positive, i.e.,

$$\beta_{ij} + \alpha_{ij} > 0 \quad \text{for all } (i, j) \in \mathcal{E}.$$

The connectivity assumption (Assumption 2) ensures that there is a path from any forceful agent to other agents in the network, implying that for any forceful agent  $i$ , there is a link  $(i, j) \in \mathcal{E}$  for some  $j \in \mathcal{N}$ . Then the main role of Assumption 3 is to guarantee that even the forceful agents at some point get information from the other agents in the network.<sup>14</sup> This assumption captures the idea that “no man is an island,” i.e., even the beliefs of forceful agents are affected by the beliefs of the society. In the absence of this assumption, any society consisting of several forceful agents may never settle into a stationary distribution of beliefs. While this is an interesting situation to investigate, it requires a very different approach. Since we view the “no man is an island” feature as plausible, we find Assumption 3 a useful starting point.

Throughout the rest of the paper, we assume that Assumptions 1, 2, and 3 hold.

### 2.3. Evolution of beliefs: Social network and influence matrices

We can express the preceding belief update model compactly as follows. Let  $x(k) = (x_1(k), \dots, x_n(k))$  denote the vector of agent beliefs at time slot  $k$ . The agent beliefs are updated according to the relation

$$x(k+1) = W(k)x(k), \quad (4)$$

where  $W(k)$  is a random matrix given by

$$W(k) = \begin{cases} A^{ij} \equiv I - \frac{(e_i - e_j)(e_i - e_j)'}{2} & \text{with probability } p_{ij}\beta_{ij}/n, \\ J^{ij} \equiv I - (1 - \epsilon)e_i(e_i - e_j)' & \text{with probability } p_{ij}\alpha_{ij}/n, \\ I & \text{with probability } p_{ij}\gamma_{ij}/n, \end{cases} \quad (5)$$

for all  $i, j \in \mathcal{N}$ . The preceding belief update model implies that the matrix  $W(k)$  is a stochastic matrix for all  $k$ , and is independent and identically distributed over all  $k$ .

Let us introduce the matrices

$$\Phi(k, s) = W(k)W(k-1) \cdots W(s+1)W(s) \quad \text{for all } k \text{ and } s \text{ with } k \geq s, \quad (6)$$

with  $\Phi(k, k) = W(k)$  for all  $k$ . We will refer to the matrices  $\Phi(k, s)$  as the *transition matrices*. We can now write the belief update rule (4) as follows: for all  $s$  and  $k$  with  $k \geq s \geq 0$  and all agents  $i \in \{1, \dots, n\}$ ,

$$x_i(k+1) = \sum_{j=1}^n [\Phi(k, s)]_{ij} x_j(s). \quad (7)$$

Given our assumptions, the random matrix  $W(k)$  is identically distributed over all  $k$ , and thus we have for some non-negative matrix  $\tilde{W}$ ,

$$E[W(k)] = \tilde{W} \quad \text{for all } k \geq 0. \quad (8)$$

The matrix  $\tilde{W}$ , which we refer to as the *mean interaction matrix*, represents the evolution of beliefs in the society. It incorporates elements from both the underlying social network (which determines the meeting patterns) and the influence structure. In what follows, it will be useful to separate these into two components, both for our mathematical analysis and to clarify the intuitions. For this purpose, let us use the belief update model (4)–(5) and write the mean interaction matrix  $\tilde{W}$  as follows<sup>15</sup>:

<sup>14</sup> This assumption is stated for all  $(i, j) \in \mathcal{E}$ , thus a forceful agent  $i$  receives some information from any  $j$  in his “neighborhood.” This is without any loss of generality, since we can always set  $p_{ij} = 0$  for those  $j$ 's that are in  $i$ 's neighborhood but from whom  $i$  never obtains information.

<sup>15</sup> In the sequel, the notation  $\sum_{i,j}$  will be used to denote the double sum  $\sum_{i=1}^n \sum_{j=1}^n$ .

$$\tilde{W} = \frac{1}{n} \sum_{i,j} p_{ij} [\beta_{ij} A^{ij} + \alpha_{ij} J^{ij} + \gamma_{ij} I] = \frac{1}{n} \sum_{i,j} p_{ij} [(1 - \gamma_{ij}) A^{ij} + \gamma_{ij} I] + \frac{1}{n} \sum_{i,j} p_{ij} \alpha_{ij} [J^{ij} - A^{ij}],$$

where  $A^{ij}$  and  $J^{ij}$  are matrices defined in Eq. (5), and the second equality follows from the fact that  $\beta_{ij} = 1 - \alpha_{ij} - \gamma_{ij}$  for all  $i, j \in \mathcal{N}$ . We use the notation

$$T = \frac{1}{n} \sum_{i,j} p_{ij} [(1 - \gamma_{ij}) A^{ij} + \gamma_{ij} I], \quad D = \frac{1}{n} \sum_{i,j} p_{ij} \alpha_{ij} [J^{ij} - A^{ij}] \quad (9)$$

to write the mean interaction matrix,  $\tilde{W}$ , as

$$\tilde{W} = T + D. \quad (10)$$

Here, the symmetric matrix  $T$  only depends on meeting probabilities (matrix  $P$ ) except that it also incorporates  $\gamma_{ij}$  (probability that following a meeting no exchange takes place). We can therefore think of the matrix  $T$  as representing the underlying *social network* (friendships, communication among coworkers, decisions about which news outlets to watch, etc.), and refer to it as the *social network matrix*. It will be useful below to represent the social interactions using an undirected (and weighted) graph induced by the social network matrix  $T$ . This graph is given by  $(\mathcal{N}, \mathcal{A})$ , where  $\mathcal{A}$  is the set of undirected edges given by

$$\mathcal{A} = \{\{i, j\} \mid T_{ij} > 0\}, \quad (11)$$

and the weight  $w_e$  of edge  $e = \{i, j\}$  is given by the entry  $T_{ij} = T_{ji}$  of the matrix  $T$ . We refer to this graph as the *social network graph*.

The matrix  $D$ , on the other hand, can be thought of as representing the *influence structure* in the society. It incorporates information about which individuals and links are forceful (i.e., which types of interactions will lead to one individual influencing the other without updating his own beliefs). We refer to matrix  $D$  as the *influence matrix*. It is also useful to note for interpreting the mathematical results below that  $T$  is a doubly stochastic matrix, while  $D$  is not. Therefore, Eq. (10) gives a decomposition of the mean connectivity matrix  $\tilde{W}$  into a doubly stochastic and a remainder component, and enables us to use tools from matrix perturbation theory (see Section 4).

### 3. Convergence

In this section, we provide our main convergence result. In particular, we show that despite the presence of forceful agents, with potentially very different opinions at the beginning, the society will ultimately converge to a consensus, in which all individuals share the same belief. This consensus value of beliefs itself is a random variable. We also provide a first characterization of the expected value of this consensus belief in terms of the mean interaction matrix (and thus social network and influence matrices). Our analysis essentially relies on showing that iterates of Eq. (4),  $x(k)$ , converge to a consensus with probability one, i.e.,  $x(k) \rightarrow \bar{x}e$ , where  $\bar{x}$  is a scalar random variable that depends on the initial beliefs and the random sequence of matrices  $\{W(k)\}$ , and  $e$  is the vector of all ones. The proof uses two lemmas which are presented in Appendix B.

**Theorem 1.** *The sequences  $\{x_i(k)\}$ ,  $i \in \mathcal{N}$ , generated by Eq. (4) converge to a consensus belief, i.e., there exists a scalar random variable  $\bar{x}$  such that*

$$\lim_{k \rightarrow \infty} x_i(k) = \bar{x} \quad \text{for all } i \text{ with probability one.}$$

Moreover, the random variable  $\bar{x}$  is a convex combination of initial agent beliefs, i.e.,

$$\bar{x} = \sum_{j=1}^n \pi_j x_j(0),$$

where  $\pi = [\pi_1, \dots, \pi_n]$  is a random vector that does not depend on the initial beliefs, and satisfies  $\pi_j \geq 0$  for all  $j$ , and  $\sum_{j=1}^n \pi_j = 1$ .

**Proof.** By Lemma 9 from Appendix B, we have

$$P \left\{ [\Phi(s + n^2 d - 1, s)]_{ij} \geq \frac{\eta^d}{2} \epsilon^{n^2 - 1} \text{ for all } i, j \right\} \geq \left( \frac{\eta^d}{2} \right)^{n^2} \quad \text{for all } s \geq 0,$$

where  $\Phi(s + n^2 d - 1, s)$  is a transition matrix [cf. Eq. (6)],  $d$  is the maximum shortest path length in graph  $(\mathcal{N}, E)$  [cf. Eq. (3)],  $\epsilon$  is the self belief weight against a forceful agent [cf. Eq. (1)], and  $\eta$  is a positive scalar defined in Eq. (40). This relation implies that over a window of length  $n^2 d$ , all entries of the transition matrix  $\Phi(s + n^2 d - 1, s)$  are strictly positive

with positive probability, which is uniformly bounded away from 0. Thus, we can use Lemma 6 (from Appendix A) with the identifications

$$H(k) = W(k), \quad B = n^2 d, \quad \mu = \frac{\eta^d}{2} \epsilon^{n^2-1}.$$

Letting

$$M(k) = \max_{i \in \mathcal{N}} x_i(k), \quad m(k) = \min_{i \in \mathcal{N}} x_i(k),$$

this implies that  $n \frac{\eta^d}{2} \epsilon^{n^2-1} \leq 1$  and for all  $s \geq 0$ ,

$$P\{M(s + n^2 d) - m(s + n^2 d) \leq (1 - n\eta^d/2\epsilon^{n^2-1})(M(s) - m(s))\} \geq \left(\frac{\eta^d}{2}\right)^{n^2}.$$

Moreover, by the stochasticity of the matrix  $W(k)$ , it follows that the sequence  $\{M(k) - m(k)\}$  is nonincreasing. Hence, we have for all  $s \geq 0$

$$E[M(s + n^2 d) - m(s + n^2 d)] \leq \left[1 - \left(\frac{\eta^d}{2}\right)^{n^2} + \left(\frac{\eta^d}{2}\right)^{n^2} (1 - n\eta^d/2\epsilon^{n^2-1})\right] (M(s) - m(s)),$$

from which, for any  $k \geq 0$ , we obtain

$$E[M(k) - m(k)] \leq \left[1 - \left(\frac{\eta^d}{2}\right)^{n^2} + \left(\frac{\eta^d}{2}\right)^{n^2} (1 - n\eta^d/2\epsilon^{n^2-1})\right]^{\lfloor \frac{k}{n^2 d} \rfloor} (M(0) - m(0)),$$

and therefore

$$\lim_{k \rightarrow \infty} E[M(k) - m(k)] = 0.$$

Since the random variables  $M(k) - m(k)$  are nonnegative for all  $k$ , we can use Fatou's lemma to obtain

$$0 \leq E\left[\liminf_{k \rightarrow \infty} M(k) - m(k)\right] \leq \liminf_{k \rightarrow \infty} E[M(k) - m(k)] = 0.$$

Thus, the nonnegative random variable  $\liminf_{k \rightarrow \infty} M(k) - m(k)$  has expectation equal to 0, which implies that

$$\liminf_{k \rightarrow \infty} M(k) - m(k) = 0 \quad \text{with probability one.}$$

Since the sequence  $\{M(k) - m(k)\}$  is nonincreasing and bounded from below, it follows that  $\{M(k) - m(k)\}$  converges, which combined with the preceding relation implies that

$$\lim_{k \rightarrow \infty} M(k) - m(k) = 0 \quad \text{with probability one.}$$

The stochasticity of the matrix  $W(k)$  further implies that the sequences  $\{M(k)\}$  and  $\{m(k)\}$  are bounded and monotone and therefore converge to the same limit, which we denote by  $\bar{x}$ . Since we have

$$m(k) \leq x_i(k) \leq M(k) \quad \text{for all } i \text{ and } k \geq 0,$$

it follows that

$$\lim_{k \rightarrow \infty} x_i(k) = \bar{x} \quad \text{for all } i \text{ with probability one,}$$

establishing the first result.

Letting  $s = 0$  in Eq. (7), we have for all  $i$

$$x_i(k) = \sum_{j=1}^n [\Phi(k-1, 0)]_{ij} x_j(0) \quad \text{for all } k \geq 0. \quad (12)$$

From the previous part, for any initial belief vector  $x(0)$ , the limit

$$\lim_{k \rightarrow \infty} x_i(k) = \sum_{j=1}^n \lim_{k \rightarrow \infty} [\Phi(k-1, 0)]_{ij} x_j(0)$$

exists and is independent of  $i$ . Hence, for any  $h$ , we can choose  $x(0) = e_h$ , i.e.,  $x_h(0) = 1$  and  $x_j(0) = 0$  for all  $j \neq h$ , implying that the limit

$$\lim_{k \rightarrow \infty} [\Phi(k-1, 0)]_{ih}$$



exists and is independent of  $i$ . Denoting this limit by  $\pi_h$  and using Eq. (12), we obtain the desired result, where the properties of the vector  $\pi = [\pi_1, \dots, \pi_n]$  follow from the stochasticity of matrix  $\Phi(k, 0)$  for all  $k$  (implying the stochasticity of its limit as  $k \rightarrow \infty$ ).  $\square$

The key implication of this result is that, despite the presence of forceful agents, the society will ultimately reach a consensus. Though surprising at first, this result is intuitive in light of our “no man is an island” assumption (Assumption 3). However, in contrast to “averaging models” used both in the engineering literature and recently in the learning literature, the consensus value here is a random variable even conditional on nonrandom initial beliefs. Moreover, the consensus value will depend on the order in which meetings have taken place. The main role of this result for us is that we can now conduct our analysis on quantifying the extent of the spread of misinformation by looking at this consensus value of beliefs.

The next theorem characterizes  $E[\bar{x}]$  in terms of the limiting behavior of the matrices  $\tilde{W}^k$  as  $k$  goes to infinity.

**Theorem 2.** Let  $\bar{x}$  be the limiting random variable of the sequences  $\{x_i(k)\}$ ,  $i \in \mathcal{N}$ , generated by Eq. (4) (cf. Theorem 1). Then we have:

(a) The matrix  $\tilde{W}^k$  converges to a stochastic matrix with identical rows  $\bar{\pi}$  as  $k$  goes to infinity, i.e.,

$$\lim_{k \rightarrow \infty} \tilde{W}^k = e\bar{\pi}'.$$

(b) The expected value of  $\bar{x}$  is given by a convex combination of the initial agent values  $x_i(0)$ , where the weights are given by the components of the probability vector  $\bar{\pi}$ , i.e.,

$$E[\bar{x}] = \sum_{i=1}^n \bar{\pi}_i x_i(0) = \bar{\pi}'x(0).$$

**Proof.** (a) This part relies on the properties of the mean interaction matrix established in Appendix B. In particular, by Lemma 7(a), the mean interaction matrix  $\tilde{W}$  is a primitive matrix. Therefore, the Markov chain with transition probability matrix  $\tilde{W}$  is regular (see Section 4.1 for a definition). The result follows immediately from Theorem 3(a).

(b) From Eq. (7), we have for all  $k \geq 0$

$$x(k) = \Phi(k-1, 0)x(0).$$

Moreover, since  $x(k) \rightarrow \bar{x}e$  as  $k \rightarrow \infty$ , we have

$$E[\bar{x}e] = E\left[\lim_{k \rightarrow \infty} x(k)\right] = \lim_{k \rightarrow \infty} E[x(k)],$$

where the second equality follows from the Lebesgue's Dominated Convergence Theorem (see Rudin, 1987). Combining the preceding two relations and using the assumption that the matrices  $W(k)$  are independent and identically distributed over all  $k \geq 0$ , we obtain

$$E[\bar{x}e] = \lim_{k \rightarrow \infty} E[\Phi(k-1, 0)x(0)] = \lim_{k \rightarrow \infty} \tilde{W}^k x(0),$$

which in view of part (a) implies

$$E[\bar{x}] = \bar{\pi}'x(0). \quad \square$$

Combining Theorem 1 and Theorem 2(a) (and using the fact that the results hold for any  $x(0)$ ), we have  $\bar{\pi} = E[\pi]$ . The stationary distribution  $\bar{\pi}$  is crucial in understanding the formation of opinions since it encapsulates the weight given to each agent (forceful or regular) in the (limiting) mean consensus value of the society. We refer to the vector  $\bar{\pi}$  as the *consensus distribution* corresponding to the mean interaction matrix  $\tilde{W}$  and its component  $\bar{\pi}_i$  as the *weight* of agent  $i$ .

It is also useful at this point to highlight how consensus will form around the correct value in the absence of forceful agents. Let  $\{x(k)\}$  be the belief sequence generated by the belief update rule of Eq. (4). When there are no forceful agents, i.e.,  $\alpha_{ij} = 0$  for all  $i, j$ , then the interaction matrix  $W(k)$  for all  $k$  is either equal to an averaging matrix  $A^{ij}$  for some  $i, j$  or equal to the identity matrix  $I$ ; hence,  $W(k)$  is a doubly stochastic matrix. This implies that the average value of  $x(k)$  remains constant at each iteration, i.e.,

$$\frac{1}{n} \sum_{i=1}^n x_i(k) = \frac{1}{n} \sum_{i=1}^n x_i(0) \quad \text{for all } k \geq 0.$$

Theorem 1 therefore shows that when there are no forceful agents, the sequences  $x_i(k)$  for all  $i$ , converge to the average of the initial beliefs with probability one, aggregating information. We state this result as a simple corollary.

**Corollary 1.** Assume that there are no forceful agents, i.e.,  $\alpha_{ij} = 0$  for all  $i, j \in \mathcal{N}$ . We have

$$\lim_{k \rightarrow \infty} x_i(k) = \frac{1}{n} \sum_{i=1}^n x_i(0) = \theta \quad \text{with probability one.}$$

Therefore, in the absence of forceful agents, the society is able to aggregate information effectively. Theorem 2 then also implies that in this case  $\pi = \bar{\pi}_i = 1/n$  for all  $i$  (i.e., beliefs converge to a deterministic value given the state  $\theta$ ), so that no individual has excess influence. These results no longer hold when there are forceful agents. In the next section, we investigate the effect of the forceful agents and the structure of the social network on the extent of misinformation and excess influence of individuals.

#### 4. Global limits on misinformation

In this section, we are interested in providing an upper bound on the expected value of the difference between the consensus belief  $\bar{x}$  (cf. Theorem 1) and the true underlying state  $\theta$  (or equivalently the average of the initial beliefs), i.e.,

$$E[\bar{x} - \theta] = E[\bar{x}] - \theta = \sum_{i \in \mathcal{N}} \left( \bar{\pi}_i - \frac{1}{n} \right) x_i(0) \quad (13)$$

(cf. Theorem 2). This is a natural measure of average misinformation in society and will be our focus throughout the rest of the paper.<sup>16</sup>

Our bound relies on a fundamental theorem from the perturbation theory of finite Markov chains. Before presenting the theorem, we first introduce some terminology and basic results related to Markov chains.

##### 4.1. Preliminary results

Consider a finite Markov chain with  $n$  states and transition probability matrix  $T$ .<sup>17</sup> We say that a finite Markov chain is *regular* if its transition probability matrix is a primitive matrix, i.e., there exists some integer  $k > 0$  such that all entries of the power matrix  $T^k$  are positive. The following theorem states basic results on the limiting behavior of products of transition matrices of Markov chains (see Theorems 4.1.4, 4.1.6, and 4.3.1 in Kemeny and Snell, 1960).

**Theorem 3.** Consider a regular Markov chain with  $n$  states and transition probability matrix  $T$ .

- (a) The  $k$ th power of the transition matrix  $T$ ,  $T^k$ , converges to a stochastic matrix  $T^\infty$  with all rows equal to the probability vector  $\pi$ , i.e.,

$$\lim_{k \rightarrow \infty} T^k = T^\infty = e\pi',$$

where  $e$  is the  $n$ -dimensional vector of all ones.

- (b) The probability vector  $\pi$  is a left eigenvector of the matrix  $T$ , i.e.,

$$\pi' T = \pi' \quad \text{and} \quad \pi' e = 1.$$

The vector  $\pi$  is referred to as the stationary distribution of the Markov chain.

- (c) The matrix  $Y = (I - T + T^\infty)^{-1} - T^\infty$  is well defined and is given by

$$Y = \sum_{k=0}^{\infty} (T^k - T^\infty).$$

The matrix  $Y$  is referred to as the fundamental matrix of the Markov chain.

The following theorem provides an exact perturbation result for the stationary distribution of a regular Markov chain in terms of its fundamental matrix. The theorem is based on a result due to Schweitzer (1968) (see also Haviv and Van Der Heyden, 1984).

<sup>16</sup> One might also want to study other moments of distribution of  $\bar{x} - \theta$  though these are mathematically more challenging.

<sup>17</sup> We use the same notation as in (10) here, given the close connection between the matrices introduced in the next two theorems and the ones in (10).

**Theorem 4.** Consider a regular Markov chain with  $n$  states and transition probability matrix  $T$ . Let  $\pi$  denote its unique stationary distribution and  $Y$  denote its fundamental matrix. Let  $D$  be an  $n \times n$  perturbation matrix such that the sum of the entries in each row is equal to 0, i.e.,

$$\sum_{j=1}^n [D]_{ij} = 0 \quad \text{for all } i.$$

Assume that the perturbed Markov chain with transition matrix  $\hat{T} = T + D$  is regular. Then, the perturbed Markov chain has a unique stationary distribution  $\hat{\pi}$ , and the matrix  $I - DY$  is nonsingular. Moreover, the change in the stationary distributions,  $\rho = \hat{\pi} - \pi$ , is given by

$$\rho' = \pi' DY (I - DY)^{-1}.$$

#### 4.2. Main results

This subsection provides bounds on the difference between the consensus distribution and the uniform distribution using the global properties of the underlying social network. Our method of analysis will rely on the decomposition of the mean interaction matrix  $\tilde{W}$  given in (10) into the social network matrix  $T$  and the influence matrix  $D$ . Recall that  $T$  is doubly stochastic.

The next theorem provides our first result on characterizing the extent of misinformation and establishes an upper bound on the  $l_\infty$ -norm of the difference between the stationary distribution  $\tilde{\pi}$  and the uniform distribution  $\frac{1}{n}e$ , which, from Eq. (13), also provides a bound on the deviation between expected beliefs and the true underlying state  $\theta$ .

#### Theorem 5.

(a) Let  $\tilde{\pi}$  denote the consensus distribution. The  $l_\infty$ -norm of the difference between  $\tilde{\pi}$  and  $\frac{1}{n}e$  is given by

$$\left\| \tilde{\pi} - \frac{1}{n}e \right\|_\infty \leq \frac{1}{1-\delta} \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{2n},$$

where  $\delta$  is a constant defined by

$$\delta = (1 - n\chi^d)^{\frac{1}{d}},$$

$$\chi = \min_{(i,j) \in \mathcal{E}} \left\{ \frac{1}{n} \left[ p_{ij} \frac{1 - \gamma_{ij}}{2} + p_{ji} \frac{1 - \gamma_{ji}}{2} \right] \right\},$$

and  $d$  is the maximum shortest path length in the graph  $(\mathcal{N}, \mathcal{E})$  [cf. Eq. (3)].

(b) Let  $\bar{x}$  be the limiting random variable of the sequences  $\{x_i(k)\}$ ,  $i \in \mathcal{N}$ , generated by Eq. (4) (cf. Theorem 1). We have

$$\left| E[\bar{x}] - \frac{1}{n} \sum_{i=1}^n x_i(0) \right| \leq \frac{1}{1-\delta} \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{2n} \|x(0)\|_\infty.$$

**Proof.** (a) Recall that the mean interaction matrix can be represented as

$$\tilde{W} = T + D$$

[cf. Eq. (10)], i.e.,  $\tilde{W}$  can be viewed as a perturbation of the social network matrix  $T$  by influence matrix  $D$ . By Lemma 10(a), the stationary distribution of the Markov chain with transition probability matrix  $T$  is given by the uniform distribution  $\frac{1}{n}e$ . By the definition of the matrix  $D$  [cf. Eq. (9)] and the fact that the matrices  $A^{ij}$  and  $J^{ij}$  are stochastic matrices with all row sums equal to one [cf. Eq. (5)], it follows that the sum of entries of each row of  $D$  is equal to 0. Moreover, by Theorem 2(a), the Markov chain with transition probability matrix  $\tilde{W}$  is regular and has a stationary distribution  $\tilde{\pi}$ . Therefore, we can use the exact perturbation result given in Theorem 4 to write the change in the stationary distributions  $\frac{1}{n}e$  and  $\tilde{\pi}$  as

$$\left( \tilde{\pi} - \frac{1}{n}e \right)' = \frac{1}{n} e' DY (I - DY)^{-1}, \quad (14)$$

where  $Y$  is the fundamental matrix of the Markov chain with transition probability matrix  $T$ , i.e.,

$$Y = \sum_{k=0}^{\infty} (T^k - T^\infty),$$

with  $T^\infty = \frac{1}{n}ee'$  [cf. Theorem 3(c)]. Algebraic manipulation of Eq. (14) yields

$$\left(\bar{\pi} - \frac{1}{n}e\right)' = \bar{\pi}'DY,$$

implying that

$$\left\|\bar{\pi} - \frac{1}{n}e\right\|_\infty \leq \|DY\|_\infty, \quad (15)$$

where  $\|DY\|_\infty$  denotes the matrix norm induced by the  $l_\infty$  vector norm.

We next obtain an upper bound on the matrix norm  $\|DY\|_\infty$ . By the definition of the fundamental matrix  $Y$ , we have

$$DY = \sum_{k=0}^{\infty} D(T^k - T^\infty) = \sum_{k=0}^{\infty} DT^k, \quad (16)$$

where the second equality follows from the fact that the row sums of matrix  $D$  are equal to 0 and the matrix  $T^\infty$  is given by  $T^\infty = \frac{1}{n}ee'$ .

Given any  $z(0) \in \mathbb{R}^n$  with  $\|z(0)\|_\infty = 1$ , let  $\{z(k)\}$  denote the sequence generated by the linear update rule

$$z(k) = T^k z(0) \quad \text{for all } k \geq 0.$$

Then, for all  $k \geq 0$ , we have

$$DT^k z(0) = Dz(k),$$

which by the definition of the matrix  $D$  [cf. Eq. (9)] implies

$$DT^k z(0) = \frac{1}{n} \sum_{i,j} p_{ij} \alpha_{ij} z^{ij}(k), \quad (17)$$

where the vector  $z^{ij}(k) \in \mathbb{R}^n$  is defined as

$$z^{ij}(k) = [J^{ij} - A^{ij}]z(k) \quad \text{for all } i, j, \text{ and } k \geq 0.$$

By the definition of the matrices  $J^{ij}$  and  $A^{ij}$  [cf. Eq. (5)], the entries of the vector  $z^{ij}(k)$  are given by

$$[z^{ij}(k)]_l = \begin{cases} (\frac{1}{2} - \epsilon)(z_j(k) - z_i(k)) & \text{if } l = i, \\ \frac{1}{2}(z_j(k) - z_i(k)) & \text{if } l = j, \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

This implies that the vector norm  $\|z^{ij}(k)\|_\infty$  can be upper-bounded by

$$\|z^{ij}(k)\|_\infty \leq \frac{1}{2} \left[ \max_{l \in \mathcal{N}} z_l(k) - \min_{l \in \mathcal{N}} z_l(k) \right] \quad \text{for all } i, j, \text{ and } k \geq 0.$$

Defining  $M(k) = \max_{l \in \mathcal{N}} z_l(k)$  and  $m(k) = \min_{l \in \mathcal{N}} z_l(k)$  for all  $k \geq 0$ , this implies that

$$\|z^{ij}(k)\|_\infty \leq \frac{1}{2} (M(k) - m(k)) \leq \frac{1}{2} \delta^k (M(0) - m(0)) \quad \text{for all } i, j, \text{ and } k \geq 0,$$

where the second inequality follows from Lemma 10(b) in Appendix C. Combining the preceding relation with Eq. (17), we obtain

$$\|DT^k z(0)\|_\infty \leq \frac{1}{2n} \left( \sum_{i,j} p_{ij} \alpha_{ij} \right) \delta^k (M(0) - m(0)).$$

By Eq. (16), it follows that

$$\|DY z(0)\|_\infty \leq \sum_{k=0}^{\infty} \|DT^k z(0)\|_\infty \leq \sum_{k=0}^{\infty} \frac{1}{2n} \left( \sum_{i,j} p_{ij} \alpha_{ij} \right) \delta^k (M(0) - m(0)) \leq \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{2n(1 - \delta)},$$

where to get the last inequality, we used the fact that  $0 \leq \delta < 1$  and  $M(0) - m(0) \leq 1$ , which follows from  $\|z(0)\|_\infty = 1$ . Since  $z(0)$  is an arbitrary vector with  $\|z(0)\|_\infty = 1$ , this implies that

$$\|DY\|_\infty = \min_{\{z \mid \|z\|_\infty = 1\}} \|DY z\|_\infty \leq \frac{1}{2n(1 - \delta)} \left( \sum_{i,j} p_{ij} \alpha_{ij} \right).$$

Combining this bound with Eq. (15), we obtain

$$\left\| \bar{\pi} - \frac{1}{n}e \right\|_{\infty} \leq \frac{1}{1-\delta} \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{2n},$$

establishing the desired relation.

(b) By Lemma 2(b), we have

$$E[\bar{x}] = \bar{\pi}' x(0).$$

This implies that

$$\left| E[\bar{x}] - \frac{1}{n} \sum_{i=1}^n x_i(0) \right| = \left| \bar{\pi}' x(0) - \frac{1}{n} e' x(0) \right| \leq \left\| \bar{\pi} - \frac{1}{n} e \right\|_{\infty} \|x(0)\|_{\infty}.$$

The result follows by combining this relation with part (a).  $\square$

Before providing the intuition for the preceding theorem, we provide a related bound on the  $l_2$ -norm of the difference between  $\bar{\pi}$  and the uniform distribution  $\frac{1}{n}e$  in terms of the second largest eigenvalue of the social network matrix  $T$ , and then return to the intuition for both results.

**Theorem 6.** Let  $\bar{\pi}$  denote the consensus distribution (cf. Lemma 2). The  $l_2$ -norm of the difference between  $\bar{\pi}$  and  $\frac{1}{n}e$  is given by

$$\left\| \bar{\pi} - \frac{1}{n}e \right\|_2 \leq \frac{1}{1-\lambda_2(T)} \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n},$$

where  $\lambda_2(T)$  is the second largest eigenvalue of the matrix  $T$  defined in Eq. (9).

**Proof.** Following a similar argument as in the proof of Theorem 5, we obtain

$$\left\| \bar{\pi} - \frac{1}{n}e \right\|_2 \leq \|DY\|_2, \quad (19)$$

where  $\|DY\|_2$  is the matrix norm induced by the  $l_2$  vector norm. To obtain an upper bound on the matrix norm  $\|DY\|_2$ , we consider an initial vector  $z(0) \in \mathbb{R}^n$  with  $\|z(0)\|_2 = 1$  and the sequence generated by

$$z(k+1) = Tz(k) \quad \text{for all } k \geq 0.$$

Then, for all  $k \geq 0$ , we have

$$DT^k z(0) = \frac{1}{n} \sum_{i,j} p_{ij} \alpha_{ij} z^{ij}(k), \quad (20)$$

where the entries of the vector  $z^{ij}(k)$  are given by Eq. (18). We can provide an upper bound on the  $\|z^{ij}(k)\|_2^2$  as

$$\|z^{ij}(k)\|_2^2 \leq \frac{1}{2} (z_j(k) - z_i(k))^2 = \frac{1}{2} ((z_j(k) - \bar{z}) + (\bar{z} - z_i(k)))^2,$$

where  $\bar{z} = \frac{1}{n} \sum_{l=1}^n z_l(k)$  for all  $k$  (note that since  $T$  is a doubly stochastic matrix, the average of the entries of the vector  $z(k)$  is the same for all  $k$ ). Using the relation  $(a+b)^2 \leq 2(a^2 + b^2)$  for any scalars  $a$  and  $b$ , this yields

$$\|z^{ij}(k)\|_2^2 \leq \sum_{l=1}^n (z_l(k) - \bar{z})^2 = \|z(k) - \bar{z}e\|_2^2. \quad (21)$$

We have

$$z(k+1) - \bar{z}e = Tz(k) - \bar{z}e = T(z(k) - \bar{z}e),$$

where the second equality follows from the stochasticity of the matrix  $T$ , implying that  $Te = e$ . The vector  $z(k) - \bar{z}e$  is orthogonal to the vector  $e$ , which is the eigenvector corresponding to the largest eigenvalue  $\lambda_1 = 1$  of matrix  $T$  (note that  $\lambda_1 = 1$  since  $T$  is a primitive and stochastic matrix). Hence, using the variational characterization of eigenvalues, we obtain

$$\|z(k+1) - \bar{z}e\|_2^2 \leq (z(k) - \bar{z}e)' T^2 (z(k) - \bar{z}e) \leq \lambda_2(T)^2 \|z(k) - \bar{z}e\|_2^2,$$

where  $\lambda_2(T)$  is the second largest eigenvalue of matrix  $T$ , which implies

$$\|z(k) - \bar{z}e\|_2^2 \leq (\lambda_2(T)^2)^k \|z(0) - \bar{z}e\|_2^2 \leq \lambda_2(T)^{2k}.$$

Here the second inequality follows from the fact that  $\|z(0)\|_2 = 1$  and  $\bar{z}$  is the average of the entries of vector  $z(0)$ . Combining the preceding relation with Eq. (21), we obtain

$$\|z^{ij}(k)\|_2 \leq \lambda_2(T)^k \quad \text{for all } k \geq 0.$$

By Eq. (20), this implies that

$$\|DT^k z(0)\|_2 = \frac{1}{n} \left( \sum_{i,j} p_{ij} \alpha_{ij} \right) \lambda_2(T)^k \quad \text{for all } k \geq 0.$$

Using the definition of the fundamental matrix  $Y$ , we obtain

$$\|DY z(0)\|_2 \leq \sum_{k=0}^{\infty} \|DT^k z(0)\|_2 \leq \sum_{k=0}^{\infty} \frac{1}{n} \left( \sum_{i,j} p_{ij} \alpha_{ij} \right) \lambda_2(T)^k = \frac{1}{1 - \lambda_2(T)} \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n},$$

for any vector  $z(0)$  with  $\|z(0)\|_2 = 1$ . Combined with Eq. (19), this yields the desired result.  $\square$

Theorem 6 characterizes the variation of the consensus distribution in terms of the average influence,  $\frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n}$ , and the second largest eigenvalue of the social network matrix  $T$ ,  $\lambda_2(T)$ . As is well known, the difference  $1 - \lambda_2(T)$ , also referred to as the *spectral gap*, governs the rate of convergence of the Markov chain induced by the social network matrix  $T$  to its stationary distribution (see Bremaud, 1999). In particular, the larger  $1 - \lambda_2(T)$  is, the faster the  $k$ th power of the transition probability matrix converges to the stationary distribution matrix (cf. Theorem 3). When the Markov chain converges to its stationary distribution rapidly, we say that the Markov chain is *fast-mixing*.<sup>18</sup>

In this light, Theorem 6 shows that, in a fast-mixing graph, given a fixed average influence  $\frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n}$ , the consensus distribution is “closer” to the underlying state  $\theta = \frac{1}{n} \sum_{i=1}^n x_i(0)$  and the extent of misinformation is limited. This is intuitive. In a fast-mixing social network graph, there are several connections between any pair of agents. Now for any forceful agent, consider the set of agents who will have some influence on his beliefs. This set itself is connected to the rest of the agents and thus obtains information from the rest of the society. Therefore, in a fast-mixing graph (or in a society represented by such a graph), the beliefs of forceful agents will themselves be moderated by the rest of the society before they spread widely. In contrast, in a slowly-mixing graph, it is possible to have a high degree of clustering around forceful agents, so that forceful agents get their (already limited) information intake mostly from the same agents that they have influenced. If so, there will be only a very indirect connection from the rest of the society to the beliefs of forceful agents and forceful agents will spread their information widely before their opinions also adjust. As a result, the consensus is more likely to be much closer to the opinions of forceful agents, potentially quite different from the true underlying state  $\theta$ . Note that having a slow-mixing graph is a necessary condition for such scenarios to arise, but not sufficient. We shall discuss this in Section 5.

This discussion also gives intuition for Theorem 5 since the constant  $\delta$  in that result is closely linked to the mixing properties of the social network matrix and the social network graph. In particular, Theorem 5 clarifies that  $\delta$  is related to the maximum shortest path and the minimum probability of (indirect) communication between any two agents in the society. These two notions also crucially influence the spectral gap  $1 - \lambda_2(T_n)$ , which plays the key role in Theorem 6.

These intuitions are illustrated in the next example, which shows how in a certain class of graphs, misinformation becomes arbitrarily small as the social network grows.

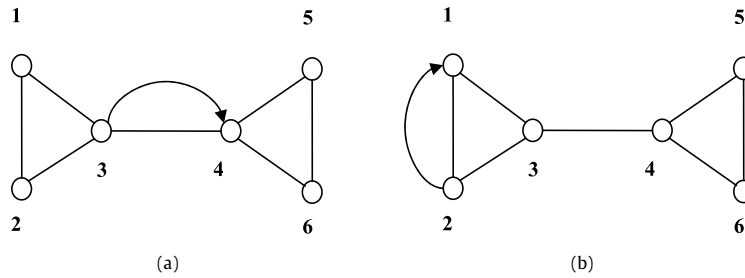
**Example 1 (Expander graphs).** Consider a sequence of social network graphs  $\mathcal{G}_n = (\mathcal{N}_n, \mathcal{A}_n)$  induced by symmetric  $n \times n$  matrices  $T_n$  [cf. Eq. (11)]. Assume that this sequence of graphs is a *family of expander graphs*, i.e., there exists a positive constant  $\gamma > 0$  such that the spectral gap  $1 - \lambda_2(T_n)$  of the graph is uniformly bounded away from 0, independent of the number of nodes  $n$  in the graph, i.e.,

$$\gamma \leq 1 - \lambda_2(T_n) \quad \text{for all } n$$

(see Chung, 1997). Social network models such as small world and preferential attachment models are expander graphs. Expander graphs have high connectivity properties and are fast-mixing.

We consider the following influence structure superimposed on the social network graph  $\mathcal{G}_n$ . We define an agent  $j$  to be *locally forceful* if he influences a constant number of agents in the society, i.e., his total influence, given by  $\sum_i p_{ij} \alpha_{ij}$ , is a constant independent of  $n$ . We define an agent  $j$  to be *non-globally forceful* if his total influence satisfies  $\lim_{n \rightarrow \infty} \frac{\sum_i p_{ij} \alpha_{ij}}{n} = 0$ , i.e., he influences a number of agents growing slower than  $n$ . We assume that the total influence is such that  $\lim_{n \rightarrow \infty} \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n} = 0$ , i.e., the total influence grows slower than  $n$ . This will be satisfied when there is a constant number of non-globally forceful agents or a number of locally forceful agents growing slower than  $n$ . Let  $\bar{\pi}_n$  denote the

<sup>18</sup> We use the terms “spectral gap of the Markov chain” and “spectral gap of the (induced) graph,” and “fast-mixing Markov chain” and “fast-mixing graph” interchangeably in the sequel.



**Fig. 1.** Impact of location of forceful agents on the stationary distribution. (a) Misinformation over the bottleneck. (b) Misinformation inside a cluster.

stationary distribution of the Markov chain with transition probability matrix given by the mean interaction matrix  $\tilde{W}$  [cf. Eq. (8)]. Then, it follows from Theorem 6 that

$$\left\| \bar{\pi}_n - \frac{1}{n} e \right\|_2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This shows that if the social network graph is fast-mixing and the total influence grows slower than  $n$ , then the difference between the consensus belief and the average of the initial beliefs vanishes. It is important to note that for this to hold, the graph sequence should be such that  $\lambda_2(T_n)$  is bounded away from 0 as  $n \rightarrow \infty$ . Intuitively, in expander graphs, as  $n$  grows large, the set of individuals who are the source of information of forceful agents become highly connected, and thus rapidly inherit the average of the information of the rest of the society. Provided that the total influence grows slower than  $n$ , their influence becomes arbitrarily small as  $n$  increases.

## 5. Connectivity of forceful agents and misinformation

The results provided so far exploit the decomposition of the evolution of beliefs into the social network component (matrix  $T$ ) and the influence component (matrix  $D$ ). This decomposition does not exploit the interactions between the structure of the social network and the location of forceful agents within it. For example, forceful agents located in different parts of the same social network will have different impacts on the extent of misinformation in the society, but our results so far do not capture this aspect. The following example illustrates these issues in a sharp way.

**Example 2.** Consider a society consisting of six agents and represented by the (undirected) social network graph shown in Fig. 1. The weight of each edge  $\{i, j\}$  is given by

$$[T]_{ij} = \frac{1}{2n}(p_{ij} + p_{ji}),$$

where, for illustration, we choose  $p_{ij}$  to be inversely proportional to the degree of node  $i$ , for all  $j$ . The self-loops are not shown in Fig. 1.

We distinguish two different cases as illustrated in Fig. 1. In each case, there is a single forceful agent and  $\alpha = 1/2$ . This is represented by a directed forceful link. The two cases differ by the location of the forceful link, i.e., the forceful link is over the bottleneck of the connectivity graph in part (a) and inside the left cluster in part (b). The corresponding consensus distributions can be computed as

$$\bar{\pi}^a = \frac{1}{6}(1.25, 1.25, 1.25, 0.75, 0.75, 0.75)', \quad \bar{\pi}^b = \frac{1}{6}(0.82, 1.18, 1, 1, 1, 1)'.$$

Even though the social network matrix  $T$  (and the corresponding graph) is the same in both cases, the consensus distributions are different. In particular, in part (a), each agent in the left cluster has a higher weight compared to the agents in the right cluster, while in part (b), the weights of all agents, except for the forceful and influenced agents, are equal and given by  $1/6$ . This is intuitive since when the forceful link is over a bottleneck, the misinformation of a forceful agent can spread and influence a larger portion of the society before his opinions can be moderated by the opinions of the other agents.

This example shows how the extent of the spread of misinformation varies depending on the location of the forceful agent. The rest of this section provides a more detailed analysis of how the location and connectivity of forceful agents affect the formation of opinions in the network. We proceed as follows. First, we provide an alternative exact characterization of excess influence using mean first passage times. We then introduce the concept of a *bridge*, similar to the situation depicted in Example 2, and provide sharper exact results for graphs in which forceful links coincide with bridges. We then generalize these notions to more general networks by introducing the concept of *information bottlenecks*, and finally, we develop new techniques for determining tighter upper bounds on excess influence by using ideas from graph clustering.

### 5.1. Characterization in terms of mean first passage times

Our next main result provides an exact characterization of the excess influence of agent  $i$  in terms of the mean passage times of the Markov chain with transition probability matrix  $T$ . This result, and those that follow later in this section, will be useful both to provide more informative bounds on the extent of misinformation and also to highlight the sources of excess influence for certain agents in the society.

We start with presenting some basic definitions and relations (see Chapter 2 of Aldous and Fill, 2002).

**Definition 1.** Let  $(X_t, t = 0, 1, 2, \dots)$  denote a discrete-time Markov chain. We denote the *first hitting time* of state  $i$  by

$$\tau_i = \inf\{t \geq 0 \mid X_t = i\},$$

which is a random variable that depends on the starting state of the Markov chain. We define the *mean first passage time* from state  $i$  to state  $j$  as

$$m_{ij} = \mathbb{E}[\tau_j \mid X_0 = i],$$

and the *mean commute time* between state  $i$  and state  $j$  as  $m_{ij} + m_{ji}$ . Moreover, we define the *mean first return time* to a particular state  $i$  as

$$m_i^+ = \mathbb{E}[\tau_i^+ \mid X_0 = i],$$

where

$$\tau_i^+ = \inf\{t \geq 1 \mid X_t = i\}.$$

**Lemma 1.** Consider a Markov chain with transition matrix  $Z$  and stationary distribution  $\pi$ . We have:

- (i) The mean first return time from state  $i$  to  $i$  is given by  $m_i^+ = 1/\pi_i$ .
- (ii) The mean first passage time from  $i$  to  $j$  is given by

$$m_{ij} = \frac{Y_{jj} - Y_{ij}}{\pi_j},$$

where  $Y = \sum_{k=0}^{\infty} (Z^k - Z^{\infty})$  is the fundamental matrix of the Markov chain.

We use the relations in the preceding lemma between the fundamental matrix of a Markov chain and the mean first passage times between states to provide an exact characterization of the excess influence of agent  $k$ .

**Theorem 7.** Let  $\bar{\pi}$  denote the consensus distribution. We have:

- (a) For every agent  $k$

$$\bar{\pi}_k - \frac{1}{n} = \frac{1}{2n^2} \sum_{i,j} p_{ij} \alpha_{ij} ((1 - 2\epsilon)\bar{\pi}_i + \bar{\pi}_j)(m_{ik} - m_{jk}).$$

- (b) Assume that there exists a single pair  $(i, j)$  for which the influence probability  $\alpha_{ij} > 0$ . Then, for every agent  $k$

$$\bar{\pi}_k - \frac{1}{n} = \frac{1}{n^3} \cdot \frac{p_{ij} \alpha_{ij} (1 - \epsilon)}{1 - \zeta_{ij}/n^2} (m_{ik} - m_{jk}), \quad (22)$$

where

$$\zeta_{ij} = \frac{p_{ij} \alpha_{ij}}{2} [m_{ij} - (1 - 2\epsilon)m_{ji}]$$

and  $m_{ij}$  is the mean first passage time from state  $i$  to state  $j$  of a Markov chain with transition matrix given by the social network matrix  $T$  [cf. Eq. (9)].

**Proof.** (a) Following the same line of argument as in the proof of Theorem 5, we can use the perturbation results of Theorem 4 to write the excess influence of agent  $k$  as

$$\bar{\pi}_k - \frac{1}{n} = \bar{\pi}' D[Y]^k, \quad (23)$$



where  $Y$  is the fundamental matrix of a Markov chain with transition matrix  $T$ . Using (5), and the definition of  $D$  in (9) we have

$$[D[Y]^k]_l = \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{n} \begin{cases} (\frac{1}{2} - \epsilon)(Y_{jk} - Y_{ik}) & \text{if } l = i, \\ \frac{1}{2}(Y_{jk} - Y_{ik}) & \text{if } l = j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, we can write right-hand side of Eq. (23) as follows:

$$\bar{\pi}_k - \frac{1}{n} = \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{2n} ((1 - 2\epsilon)\bar{\pi}_i + \bar{\pi}_j)(Y_{jk} - Y_{ik}). \quad (24)$$

By Lemma 1(ii), we have

$$Y_{jk} - Y_{ik} = \frac{1}{n}(m_{ik} - m_{jk}), \quad (25)$$

where  $Y$  is the fundamental matrix of the Markov chain with transition matrix  $T$ . The desired result follows by substituting the preceding relation in Eq. (24).

(b) In view of the assumption that there exists a single forceful link, one can write the following for the perturbation matrix  $D$  using Eq. (9):

$$D = \frac{p_{ij}\alpha_{ij}}{n} [J^{ij} - A^{ij}] = \frac{p_{ij}\alpha_{ij}}{n} \left[ \left( \epsilon - \frac{1}{2} \right) e_i - \frac{1}{2} e_j \right] (e_i - e_j)'. \quad (26)$$

By Lemma 11 of Appendix C, one can write

$$DY(I - DY)^{-1} = \left( 1 - \frac{\zeta_{ij}}{n^2} \right)^{-1} DY.$$

Combined with the exact perturbation result in Theorem 4, this implies

$$\begin{aligned} \bar{\pi}_k - \frac{1}{n} &= \frac{1}{n} [e' DY(I - DY)^{-1}]_k \\ &= \frac{1}{n} \left( 1 - \frac{\zeta_{ij}}{n^2} \right)^{-1} [e' DY]_k \\ &= \frac{p_{ij}\alpha_{ij}}{n^2} \cdot \left( 1 - \frac{\zeta_{ij}}{n^2} \right)^{-1} \left[ e' \left[ \left( \epsilon - \frac{1}{2} \right) e_i - \frac{1}{2} e_j \right] (e_i - e_j)' Y \right]_k \\ &= \frac{p_{ij}\alpha_{ij}}{n^2} \cdot \left( 1 - \frac{\zeta_{ij}}{n^2} \right)^{-1} [(\epsilon - 1)(e_i - e_j)' Y]_k \\ &= \frac{1}{n^2} \cdot \frac{p_{ij}\alpha_{ij}(1 - \epsilon)}{1 - \zeta_{ij}/n^2} (Y_{jk} - Y_{ik}). \end{aligned}$$

The main result follows by substituting Eq. (25) in the above equation.  $\square$

Part (a) of Theorem 7 provides an exact expression for the excess influence of agent  $k$  as a function of the mean first passage times from agent (or state)  $k$  to the forceful and influenced agents. The excess influence of each agent therefore depends on the difference between the *distance* of that agent to the forceful and the influenced agent. To provide an intuition for this result, let us consider the special case in which there is a single forceful link  $(j, i)$  in the society (i.e., only one pair of agents  $i$  and  $j$  with  $\alpha_{ij} > 0$ ) and thus a single forceful agent  $j$ . Then for any agent  $k$ , his or her only source of excess influence can come from their (potentially indirect) impact on the beliefs of the forceful agent  $j$ . This is why  $m_{jk}$ , which, loosely speaking, measures the distance between  $j$  and  $k$ , enters negatively in to the expression for the excess influence of agent  $k$ . In addition, any agent who meets (communicates) with agent  $i$  with a high probability will be indirectly influenced by the opinions of the forceful agent  $j$ . Therefore, the excess influence of agent  $k$  is increasing in his distance to  $i$ , thus in  $m_{ik}$ . In particular, in the extreme case where  $m_{ik}$  is small, agent  $k$  will have negative excess influence (because he is very close to the heavily “influenced” agent  $i$ ) and in the polar extreme, where  $m_{jk}$  is small, he will have positive excessive influence (because his views will be quickly heard by the forceful agent  $j$ ). The general expression in part (a) of the theorem simply generalizes this reasoning to general social networks with multiple forceful agents and several forceful links.

Part (b) provides an alternative expression [cf. Eq. (22)], with a similar intuition for the special case in which there exists a single forceful link. The main advantage of the expression in part (b) is that, though more complicated, is not in terms of the expected consensus distribution  $\bar{\pi}$  (which is endogenous). The parameter  $\zeta_{ij}$  in Eq. (22) captures the asymmetry

between the locations of agents  $i$  and  $j$  in the underlying social network graph. Although the expression for excess influence in part (a) of Theorem 7 is a function of the consensus distribution  $\bar{\pi}$ , each element of this vector (distribution) can be bounded by 1 to obtain an upper bound for the excess influence of agent  $k$ . This is explicitly stated in the following corollary of Theorem 7.

**Corollary 2.** *The magnitude of the excess influence of agent  $k$  is bounded from above as*

$$\left| \bar{\pi}_k - \frac{1}{n} \right| \leq \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{2n^2} |m_{ik} - m_{jk}|.$$

**Proof.** By Theorem 7(a) and triangle inequality we have for every agent  $k$

$$\left| \bar{\pi}_k - \frac{1}{n} \right| \leq \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{2n^2} ((1-2\epsilon)\bar{\pi}_i + \bar{\pi}_j) |m_{ik} - m_{jk}| \leq \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{2n^2} |m_{ik} - m_{jk}|,$$

where the second inequality follows from the fact that  $\bar{\pi}$  is a stochastic vector.  $\square$

Using the results in Theorem 7, the difference between the consensus distributions discussed in Example 2 can be explained as follows. In Example 2(a), the mean first passage time from agent 4 to any agent  $k$  in the left cluster is strictly larger than that of agent 3 to agent  $k$ , because every path from agent 4 to the left cluster should pass through agent 3. Therefore,  $m_{4k} > m_{3k}$  for  $k = 1, 2, 3$ , and agents in the left cluster have a higher consensus weight. In Example 2(b), due to the symmetry of the social network graph, the mean first passage times of agents 1 and 2 to any agent  $k \neq 1, 2$  are the same, hence establishing by Theorem 7 the uniform weights in the consensus distribution.

In the following we study the effect of the location of a forceful link on the excess influence of each agent by characterizing the relative mean first passage time  $|m_{ik} - m_{jk}|$ , in terms of the properties of the social network graph.

## 5.2. Forceful bridges

In this subsection, we provide an exact characterization of the excess influence of agent  $k$  explicitly in terms of the properties of the social network graph. We focus on the special case of a single forceful link where the undirected edge between the forceful and the influenced agent is a *bridge* of the social network graph, in the sense that without this edge the graph would be disconnected. We refer to such edges as *forceful bridges*. Graphs with forceful bridges approximate situations in which a forceful agent, for example a media outlet or political leader, itself obtains all of its information from an insular community.

We first give the definition of a bridge of an undirected graph.

**Definition 2.** Let  $\mathcal{G} = (\mathcal{N}, \mathcal{A})$  be an undirected connected graph. An edge  $\{i, j\} \in \mathcal{A}$  is a *bridge* of the graph  $\mathcal{G} = (\mathcal{N}, \mathcal{A})$  if its removal would disconnect the graph into two sets  $N(i, j) \subset \mathcal{N}$  with  $i \in N(i, j)$ , and  $N(j, i) \subset \mathcal{N}$  with  $j \in N(j, i)$ .

The following lemma provides an exact characterization of the mean first passage time from state  $i$  to state  $j$ , where  $i$  and  $j$  are the end nodes of a bridge  $\{i, j\}$ .

**Lemma 2.** *Consider a Markov chain with a symmetric transition probability matrix  $T$ . Let  $\{i, j\}$  be a bridge of the social network graph induced by matrix  $T$ .*

(a) *We have*

$$m_{ij} = \frac{|N(i, j)|}{T_{ij}}.$$

(b) *For every  $k \in N(j, i)$ ,*

$$m_{ik} - m_{jk} = m_{ij}.$$

**Proof.** Consider a Markov chain over the set of states  $\mathcal{N}' = N(i, j) \cup \{j\}$ , with transition probabilities

$$\hat{T}_{kl} = \begin{cases} \sum_{r \neq i} T_{jr} & \text{if } k = l = j, \\ T_{kl} & \text{otherwise.} \end{cases}$$

It is straightforward to check that the transition matrix  $\hat{T}$  corresponding to the new chain is doubly stochastic. Hence, it has a uniform stationary distribution  $\hat{\pi}$ . In particular,

$$\hat{\pi}_j = \frac{1}{|N(i, j)| + 1}.$$

Since  $\{i, j\}$  is a bridge, every path from  $i$  to  $j$  should pass through  $\{i, j\}$ . Moreover, because of equivalent transition probabilities between the new Markov chain and the original one on  $\mathcal{N}'$ , the mean passage time  $m_{ij}$  of the original Markov chain is equal to the mean first passage time  $\hat{m}_{ij}$  of the new chain. On the other hand, for the new chain, we can write the mean return time to  $j$  as

$$\hat{m}_j^+ = 1 + \hat{T}_{ji}\hat{m}_{ij} = 1 + T_{ji}m_{ij} = 1 + T_{ij}m_{ij},$$

where the last equality follows from symmetry of  $T$ . This implies [cf. Lemma 1(i)]

$$m_{ij} = \frac{1}{T_{ij}} \left( \frac{1}{\hat{\pi}_j} - 1 \right) = \frac{|N(i, j)|}{T_{ij}}.$$

The second part of the claim follows from the fact that all of the paths from  $i$  to  $k$  must pass through  $\{i, j\}$ , because it is the only edge connecting  $N(i, j)$  to  $N(j, i)$ . Thus, we conclude

$$m_{ik} = m_{ij} + m_{jk}. \quad \square$$

We use the relation in Lemma 2 to study the effect of a single forceful link over a bridge on the excess influence of each agent.

**Theorem 8.** Let  $\bar{\pi}$  denote the consensus distribution. Assume that there exists a single pair  $(i, j)$  for which the influence probability  $\alpha_{ij} > 0$ . Assume that the edge  $\{i, j\}$  is a bridge of the social network graph. Then, we have for all  $k$ ,

$$\bar{\pi}_k - \frac{1}{n} = \frac{2}{n^2} \frac{\theta_{ij}(1 - \epsilon)}{1 - \frac{\theta_{ij}}{n}(|N(i, j)| - (1 - 2\epsilon)|N(j, i)|)} \Psi_{ij}(k),$$

where

$$\theta_{ij} = \frac{p_{ij}\alpha_{ij}}{p_{ij}(1 - \gamma_{ij}) + p_{ji}(1 - \gamma_{ji})}$$

and

$$\Psi_{ij}(k) = \begin{cases} |N(i, j)|, & k \in N(j, i), \\ -|N(j, i)|, & k \in N(i, j). \end{cases}$$

**Proof.** Since edge  $\{i, j\}$  is a bridge, by Lemma 2 we have for every  $k \in N(j, i)$

$$m_{ik} - m_{jk} = m_{ij} = \frac{|N(i, j)|}{T_{ij}} = \frac{2n|N(i, j)|}{p_{ij}(1 - \gamma_{ij}) + p_{ji}(1 - \gamma_{ji})}.$$

Similarly, for every  $k \in N(i, j)$ , we obtain

$$m_{ik} - m_{jk} = -m_{ji} = -\frac{2n|N(j, i)|}{p_{ij}(1 - \gamma_{ij}) + p_{ji}(1 - \gamma_{ji})}.$$

Combining the preceding relations, we can write for the relative mean passage time  $m_{ik} - m_{jk} = \frac{2n\theta_{ij}}{p_{ij}\alpha_{ij}} \Psi_{ij}(k)$ . Since  $(i, j)$  is the only forceful link, we can apply Theorem 7(b) to get

$$\bar{\pi}_k - \frac{1}{n} = \left( \frac{2}{n^2} \right) \frac{\theta_{ij}(1 - \epsilon)}{1 - \zeta_{ij}/n^2} \cdot \Psi_{ij}(k),$$

where  $\zeta_{ij}$  is given by

$$\zeta_{ij} = \frac{p_{ij}\alpha_{ij}}{2} [m_{ij} - (1 - 2\epsilon)m_{ji}].$$

Combining the above relations with Lemma 2(a) establishes the desired result.  $\square$

Theorem 8 shows that if two clusters of agents, e.g., two communities, are connected via a bridge over which there is a forceful link, then the excess influences of all agents within the same cluster are equal (even when the cluster does not have any symmetry properties). This implies that the opinions of all agents that are in the same cluster as the forceful agent affect the consensus opinion of the society with the same strength. This property is observed in part (a) of Example 2, in which edge  $\{3, 4\}$  is a bridge. Intuitively, all of the agents in that cluster will ultimately shape the opinions of the forceful agent and this is the source of their excess influence. The interesting and surprising feature is that they all have the same excess influence, even if only some of them are directly connected to the forceful agent. Loosely speaking, this can be explained using the fact that, in the limiting distribution, it is the consensus among this cluster of agents that will impact the beliefs of the forceful agent, and since within this cluster there are no other forceful agents, the consensus value among them puts equal weight on each of them (recall Corollary 1).

### 5.3. Information bottlenecks

We now extend the ideas in Theorem 8 to more general societies. We observed in Example 2 and Section 5.2 that influence over a bridge can have global effects on the consensus distribution since bridges are “bottlenecks” of the information flow in the network. In this subsection we generalize this idea to influential links over bottlenecks that are not necessarily bridges as defined in Definition 2. Our goal is to study the impact of influential links over bottlenecks on the consensus distribution.

To achieve this goal, we first state a corollary of Theorem 7 to bound the excess influence of each agent in terms of the *mean commute time* between pairs of agents each connected via an influential link. Then by a series of (successively tighter) upper bounds on the key term  $(m_{ij} + m_{ji})$ , we show how the effect of a particular influential link  $(i, j)$  on the excess influence throughout the society is related to the information bottleneck between agents  $i$  and  $j$ .

**Corollary 3.** Let  $\bar{\pi}$  denote the consensus distribution. Then, for every agent  $k$  we have

$$\left| \bar{\pi}_k - \frac{1}{n} \right| \leq \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{2n^2} (m_{ij} + m_{ji}).$$

**Proof.** The claim is a direct result of Corollary 2 by noting that

$$|m_{ik} - m_{jk}| \leq \max(m_{ij}, m_{ji}) \leq m_{ij} + m_{ji}.$$

The first inequality holds because  $m_{ik} \leq m_{ij} + m_{jk}$  and  $m_{jk} \leq m_{ji} + m_{ik}$ .  $\square$

Our first bound on this object will be in terms of the *minimum normalized cut* of a Markov chain (induced by an undirected weighted graph), which is introduced in the next definition. We will use the term *cut of a Markov chain* (or *cut of an undirected graph*) to denote a partition of the set of states of a Markov chain (or equivalently the nodes of the corresponding graph) into two sets.

**Definition 3.** Consider a reversible Markov chain with set of states  $\mathcal{N}$ , transition probability matrix  $Z$ , and stationary distribution  $\pi$ . The *minimum normalized cut value* (or *conductance*) of the Markov chain, denoted by  $\rho$ , is defined as

$$\rho = \min_{S \subset \mathcal{N}} \frac{Q(S, S^c)}{\pi(S)\pi(S^c)}, \quad (27)$$

where  $Q(A, B) = \sum_{i \in A, j \in B} \pi_i Z_{ij}$ , and  $\pi(S) = \sum_{i \in S} \pi_i$ . We refer to a cut that achieves the minimum in this optimization problem as a *minimum normalized cut*.

The objective in the optimization problem in (27) is the (normalized) conditional probability that the Markov chain makes a transition from a state in set  $S$  to a state in set  $S^c$  given that the initial state is in  $S$ . The minimum normalized cut therefore characterizes how fast the Markov chain will escape from any part of the state space, hence is an appropriate measure of information bottlenecks or the mixing time of the underlying graph. Clearly, the minimum normalized cut value is larger in more connected graphs.

The next lemma provides a relation between the maximum mean commute time of a Markov chain (induced by an undirected graph) and the minimum normalized cut of the chain, which is presented in Section 5.3 of Chapter 4 of Aldous and Fill (2002). This result will then be used in the next theorem to provide an improved bound on the excess influences.

**Lemma 3.** Consider an  $n$ -state reversible Markov chain with transition matrix  $Z$  and stationary distribution  $\pi$ . Let  $\rho$  denote the minimum normalized cut value of the Markov chain (cf. Definition 3). The maximum mean commute time satisfies the following relation:

$$\max_{i,j} \{m_{ij} + m_{ji}\} \leq \frac{4(1 + \log n)}{\rho \min \pi_k}. \quad (28)$$

We use the preceding relation together with our characterization of excess influence in terms of mean first passage times in Corollary 3 to obtain a tighter upper bound on the  $l_\infty$ -norm of excess influence than in Theorem 5. This result, which is stated next, both gives a more directly interpretable limit on the extent of misinformation in the society and also shows the usefulness of the characterization in terms of mean first passage times in Theorem 7.

**Theorem 9.** Let  $\bar{\pi}$  denote the consensus distribution. Then, we have

$$\left\| \bar{\pi} - \frac{1}{n} e \right\|_\infty \leq \sum_{i,j} \frac{2p_{ij}\alpha_{ij}}{n} \left( \frac{1 + \log n}{\rho} \right),$$

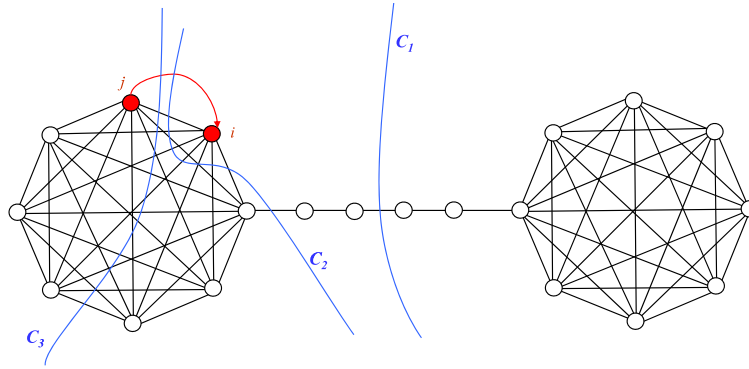


Fig. 2. The barbell graph with  $n_1 = 8$  nodes in each bell and  $n_2 = 4$ . There is a single forceful link, represented by a directed link in the left bell.

where  $\rho$  is the minimum normalized cut value of the Markov chain with transition probability matrix given by the social network matrix  $T$  (cf. Definition 3).

**Proof.** Note that the stationary distribution  $\pi$  of the social matrix  $T$  is uniform. The claim immediately follows from plugging the result of Lemma 3 into Corollary 3.  $\square$

One advantage of the result in this theorem is that the bound is in terms of  $\rho$ , the minimum normalized cut of the social network graph. As emphasized in Definition 3, this notion is related to the strength of (indirect) communication links in the society. Although the bound in Theorem 9 is tighter than the one we provided in Theorem 5, it still leaves some local information unexploited because it focuses on the maximum mean commute times between all states of a Markov chain. The following example shows how this bound may be improved further by focusing on the mean commute time between the forceful and the influenced agents.

**Example 3 (Barbell graph).** The barbell graph consists of two complete graphs each with  $n_1$  nodes that are connected via a path that consists of  $n_2$  nodes (cf. Fig. 2). Consider the asymptotic behavior

$$n \rightarrow \infty, \quad n_1/n \rightarrow \nu, \quad n_2/n \rightarrow 1 - 2\nu,$$

where  $n = 2n_1 + n_2$  denotes the total number of nodes in the barbell graph, and  $0 < \nu < \frac{1}{2}$ . The mean first passage time from a particular node in the left bell to a node in the right bell is  $O(n^3)$  as  $n \rightarrow \infty$ , while the mean passage time between any two nodes in each bell is  $O(n)$  (see Chapter 5 of Aldous and Fill, 2002 for exact results). Consider a situation where there is a single forceful link in the left bell.

The minimum normalized cut for this example is given by cut  $C_1$ , with normalized cut value  $O(1/n)$ , which captures the bottleneck in the global network structure. Since the only forceful agent is within the left bell in this example, we expect the flow of information to be limited by cuts that separate the forceful and the influenced agent, and partition the left bell. Since the left bell is a complete graph, the cuts associated with this part of the graph will have higher normalized cut values, thus yielding tighter bounds on the excess influence of the agents. In what follows, we consider bounds in terms of “relative cuts” in the social network graph that separate forceful and influenced agents in order to capture bottlenecks in the spread of misinformation (for example, cuts  $C_2$  and  $C_3$  in Fig. 2).

#### 5.4. Relative cuts

The objective of this section is to improve our characterization of the extent of misinformation in terms of information bottlenecks. To achieve this objective, we introduce a new concept, *relative normalized cuts*, and then show how this new concept is useful to derive improved upper bounds on the excess influence of different individuals and on the extent of misinformation. Our strategy is to develop tighter bounds on the mean commute times between the forceful and influenced agents in terms of relative normalized cut values. Together with Corollary 3, this enables us to provide bounds on the excess influence as a function of the properties of the social network graph and the location of the forceful agents within it.

**Definition 4.** Consider a Markov chain with set of states  $\mathcal{N}$ , transition probability matrix  $Z$ , and stationary distribution  $\pi$ . The *minimum normalized relative cut value between  $a$  and  $b$* , denoted by  $\rho_{ab}$ , is defined as

$$\rho_{ab} = \min_{S \subset \mathcal{N}} \left\{ \frac{Q(S, S^c)}{\pi(S)\pi(S^c)} \mid a \in S, b \notin S \right\},$$

where  $Q(A, B) = \sum_{i \in A, j \in B} \pi_i Z_{ij}$ , and  $\pi(S) = \sum_{i \in S} \pi_i$ . We refer to a cut that achieves the minimum in this optimization problem as a *minimum normalized relative cut*.

The next theorem provides a bound on the mean commute time between two nodes  $a$  and  $b$  as a function of the minimum normalized relative cut value between  $a$  and  $b$ .

**Theorem 10.** Consider a Markov chain with set of states  $\mathcal{N}$ , transition probability matrix  $Z$ , and uniform stationary distribution. For any  $a, b \in \mathcal{N}$ , we have

$$m_{ab} + m_{ba} \leq \frac{3n \log n}{\rho_{ab}},$$

where  $\rho_{ab}$  is the minimum normalized relative cut value between  $a$  and  $b$  (cf. Definition 4).

**Proof.** We present a generalization of the proof of Lemma 3 by Aldous and Fill (2002), for the notion of normalized relative cuts. The proof relies on the characterization of the mean commute time given by Lemma 12 in Appendix D. For a function  $0 \leq g \leq 1$  with  $g(a) = 0$  and  $g(b) = 1$ , order the nodes as  $a = 1, 2, \dots, n = b$  so that  $g$  is increasing. The Dirichlet form (cf. Definition 6) can be written as

$$\begin{aligned} \mathcal{E}(g, g) &= \sum_i \sum_{k>i} \pi_i Z_{ik} (g(k) - g(i))^2 \\ &\geq \sum_i \sum_{k>i} \sum_{i \leq j < k} \pi_i Z_{ik} (g(j+1) - g(j))^2 \\ &= \sum_{j=1}^{n-1} (g(j+1) - g(j))^2 Q(A_j, A_j^c) \\ &\geq \sum_{j=1}^{n-1} (g(j+1) - g(j))^2 \rho_{ab} \pi(A_j) \pi(A_j^c), \end{aligned} \quad (29)$$

where  $A_j = \{1, 2, \dots, j\}$ , and the last inequality is true by Definition 4. On the other hand, we have

$$1 = g(b) - g(a) = \sum_{j=1}^{n-1} (g(j+1) - g(j)) (\rho_{ab} \pi(A_j) \pi(A_j^c))^{\frac{1}{2}} (\rho_{ab} \pi(A_j) \pi(A_j^c))^{-\frac{1}{2}}.$$

Using the Cauchy–Schwartz inequality and Eq. (29), we obtain

$$\frac{1}{\mathcal{E}(g, g)} \leq \frac{1}{\rho_{ab}} \sum_{j=1}^{n-1} \frac{1}{\pi(A_j) \pi(A_j^c)}. \quad (30)$$

But  $\pi(A_j) = j/n$ , because the stationary distribution of the Markov chain is uniform. Thus,

$$\sum_{j=1}^{n-1} \frac{1}{\pi(A_j) \pi(A_j^c)} = \sum_{j=1}^{n-1} \frac{n^2}{j(n-j)} \leq 3n \log n.$$

Therefore, by applying the above relation to Eq. (30) we conclude

$$\frac{1}{\mathcal{E}(g, g)} \leq \frac{3n \log n}{\rho_{ab}}.$$

The above relation is valid for every function  $g$  feasible for the maximization problem in Eq. (50). Hence, the desired result follows from the extremal characterization of the mean commute time given by Lemma 12.  $\square$

Note that the minimum normalized cut value of a Markov chain in Definition 3 can be related to normalized relative cut values as follows:

$$\rho = \min_{a \neq b \in \mathcal{N}} \{\rho_{ab}\}.$$

Therefore, the upper bound given in Theorem 10 for the mean commute time is always tighter than that provided in Lemma 3.

Let us now examine our new characterization in the context of Example 3. The minimum normalized relative cut is given by cut  $C_2$  with (normalized relative cut) value  $O(1/n)$ , which gives a bound of  $O(n^2 \log n)$  on the mean commute time between the nodes in the left bell. Note that the definition of a minimum normalized cut results in *balanced cuts*, i.e., the cuts that partition the graph into two parts with almost the same number of nodes. Despite the fact that  $C_2$  is a

balanced cut with respect to the entire graph, it is not a balanced cut in the left bell. Nevertheless, there is a cut  $C_3$  which is balanced in the left bell, but the notion of normalized relative cut of the whole graph does not yield  $C_3$  as a minimum normalized relative cut, and we cannot benefit from  $C_3$  (which has a larger value) in the current form of Theorem 10. These considerations motivate us to consider balanced cuts within subsets of the original graph. In the following we obtain tighter bounds on the mean commute times by considering relative cuts in a subset of the original graph.

**Definition 5.** Consider a weighted undirected graph  $(\mathcal{N}, \mathcal{A})$ , with edge  $\{i, j\}$  weight given by  $w_{ij}$ . For any  $S \subseteq \mathcal{N}$ , we define the subgraph of  $(\mathcal{N}, \mathcal{A})$  induced by  $S$  as a weighted undirected graph, denoted by  $(S, \mathcal{A}_S)$ , where  $\mathcal{A}_S$  contains all edges of the original graph connecting nodes in  $S$  with the following weights

$$\bar{w}_{ij} = \bar{w}_{ji} = \begin{cases} w_{ij}, & i \neq j, \\ w_{ii} + \sum_{k \in S^c} w_{ik}, & i = j. \end{cases}$$

The next lemma uses the Monotonicity law presented in Appendix D, Lemma 13 to relate the mean commute times within an induced subgraph to the mean commute times of the original graph.

**Lemma 4.** Let  $G = (\mathcal{N}, \mathcal{A})$  be an undirected graph with edge  $\{i, j\}$  weight given by  $w_{ij}$ . Consider a Markov chain induced by this graph and denote the mean first passage times between states  $i$  and  $j$  by  $m_{ij}$ . We fix nodes  $a, b \in \mathcal{N}$ , and  $S \subseteq \mathcal{N}$  containing  $a$  and  $b$ . Consider a subgraph of  $(\mathcal{N}, \mathcal{A})$  induced by  $S$  (cf. Definition 5) and let  $\bar{m}_{ij}$  denote the mean first passage time between states  $i$  and  $j$  for the Markov chain induced by this subgraph. We have

$$m_{ab} + m_{ba} \leq \frac{w}{w(S)} (\bar{m}_{ab} + \bar{m}_{ba}),$$

where  $w$  is the total edge weight of the original graph, and  $w(S)$  is the total edge weight of the subgraph, i.e.,  $w(S) = \sum_{i \in S} \sum_{j \in \mathcal{N}} w_{ij}$ .

**Proof.** Consider an undirected graph  $(\mathcal{N}, \mathcal{A})$  with modified edge weights  $\tilde{w}_{ij}$  given by

$$\tilde{w}_{ij} = \begin{cases} w_{ij}, & i \neq j \in S, \text{ or } i \neq j \in S^c, \\ 0, & i \in S, j \in S^c, \\ w_{ii} + \sum_{k \in S^c} w_{ik}, & i = j. \end{cases}$$

Hence,  $\tilde{w}_{ij} \leq w_{ij}$  for all  $i \neq j$ , but the total edge weight  $w$  remains unchanged. By the Monotonicity law (cf. Lemma 13), the mean commute time in the original graph is bounded by that of the modified graph, i.e.,

$$m_{ab} + m_{ba} \leq \bar{m}_{ab} + \bar{m}_{ba}. \quad (31)$$

The mean commute time in the modified graph can be characterized using Lemma 12 in terms of the Dirichlet form defined in Definition 6. In particular,

$$\begin{aligned} (\bar{m}_{ab} + \bar{m}_{ba})^{-1} &= \inf_{0 \leq g \leq 1} \left\{ \frac{1}{w} \sum_{i, j \in \mathcal{N}} \tilde{w}_{ij} (\tilde{g}(i) - \tilde{g}(j))^2 : g(a) = 0, g(b) = 1 \right\} \\ &= \inf_{0 \leq g \leq 1} \left\{ \frac{1}{w} \sum_{i, j \in S} w_{ij} (\tilde{g}(i) - \tilde{g}(j))^2 : g(a) = 0, g(b) = 1 \right\} \\ &\quad + \inf_{0 \leq g \leq 1} \left\{ \frac{1}{w} \sum_{i, j \in S^c} w_{ij} (\tilde{g}(i) - \tilde{g}(j))^2 \right\} \\ &= \frac{w(S)}{w} \inf_{0 \leq g \leq 1} \left\{ \sum_{i, j \in S} \frac{\bar{w}_{ij}}{w(S)} (\tilde{g}(i) - \tilde{g}(j))^2 : g(a) = 0, g(b) = 1 \right\} \\ &= \frac{w(S)}{w} (\bar{m}_{ab} + \bar{m}_{ba})^{-1}, \end{aligned}$$

where the second equality holds by definition of  $\tilde{w}$ , and the last equality is given by definition of  $\bar{w}$ , and the extremal characterization of the mean commute time in the subgraph. The desired result is established by combining the above relation with (31).  $\square$

**Theorem 11.** Let  $\mathcal{G} = (\mathcal{N}, \mathcal{A})$  be the social network graph induced by the social network matrix  $T$  and consider a Markov chain with transition matrix  $T$ . For any  $a, b \in \mathcal{N}$ , and any  $S \subseteq \mathcal{N}$  containing  $a$  and  $b$ , we have

$$m_{ab} + m_{ba} \leq \frac{3n \log |S|}{\rho_{ab}(S)},$$

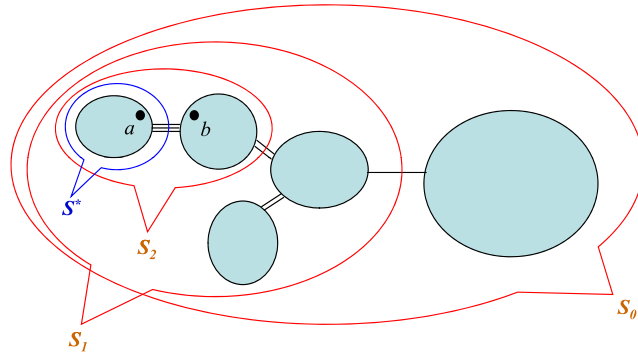


Fig. 3. A highly clustered graph with a hierarchy of clusters,  $S_0, S_1, S_2$ , containing the test nodes  $a$  and  $b$ .

where  $\rho_{ab}(S)$  is the minimum normalized cut value between  $a$  and  $b$  on the subgraph of  $(\mathcal{N}, \mathcal{A})$  induced by  $S$ , i.e.,

$$\rho_{ab}(S) = \min_{S' \subset S} |S| \frac{\sum_{i \in S', j \in S \setminus S'} T_{ij}}{|S'| \cdot |S \setminus S'|}. \quad (32)$$

**Proof.** By Lemma 4, we have

$$\begin{aligned} m_{ab} + m_{ba} &\leq \frac{w}{w(S)} (\bar{m}_{ab} + \bar{m}_{ba}) \\ &= \frac{n}{|S|} (\bar{m}_{ab} + \bar{m}_{ba}), \end{aligned} \quad (33)$$

where  $\bar{m}_{ab}$  is the mean first passage time on the subgraph  $(S, \mathcal{A}_S)$ .

On the other hand, Definition 5 implies that for the subgraph  $(S, \mathcal{A}_S)$ , we have for every  $i \in S$

$$\sum_{k \in S} \bar{w}_{ik} = \sum_{k \in S \setminus \{i\}} \bar{w}_{ik} + \bar{w}_{ii} = \sum_{k \in \mathcal{N}} w_{ik} = \sum_{k \in \mathcal{N}} T_{ik} = 1.$$

Hence, the stationary distribution of the Markov chain on the subgraph is uniform. Therefore, we can apply Lemma 10 to relate the mean commute time within the subgraph  $(S, \mathcal{A}_S)$  to its normalized relative cuts, i.e.,

$$\bar{m}_{ab} + \bar{m}_{ba} \leq \frac{3|S| \log |S|}{\rho_{ab}(S)},$$

where  $\rho_{ab}(S)$  is the minimum normalized cut between  $a$  and  $b$  given by Definition 4 on the subgraph. Since the stationary distribution of the random walk on the subgraph is uniform, we can rewrite  $\rho_{ab}(S)$  as in (32). Combining the above inequality with Eq. (33) establishes the theorem.  $\square$

Theorem 11 states that if the local neighborhood around the forceful links is highly connected, the mean commute times between the forceful and the influenced agents will be small, implying a smaller excess influence for all agents, hence limited spread of misinformation in the society. The economic intuition for this result is similar to that for our main characterization theorems: forceful agents get (their limited) information intake from their local neighborhoods. When these local neighborhoods are also connected to the rest of the network, forceful agents will be indirectly influenced by the rest of the society and this will limit the spread of their (potentially extreme) opinions. In contrast, when their local neighborhoods obtain most of their information from the forceful agents, the opinions of these forceful agents will be reinforced (rather than moderated) and this can significantly increase their excess influence and the potential spread of misinformation.

Let us revisit Example 3, and apply the result of Theorem 11 where the selected subgraph is the left cluster of nodes. The left bell is approximately a complete graph. We observe that the minimum normalized cut in the subgraph would be of the form of  $C_3$  in Fig. 2, and hence the upper bound on the mean commute time between  $i$  and  $j$  is  $O(n \log n)$ , which is close to the mean commute time on a complete graph of size  $n$ .

Note that it is possible to obtain the tightest upper bound on mean commute time between two nodes by minimizing the bound in Theorem 11 over all subgraphs  $S$  of the social network graph. Intuitively, for any two particular nodes, the goal is to identify whether such nodes are highly connected by identifying a cluster of nodes containing them, or a (local) bottleneck that separates them.

Fig. 3 illustrates a highly clustered graph. Each of the regions in Fig. 3 demonstrates a highly connected subgraph. We desire to obtain a tight bound on the mean commute time between nodes  $a$  and  $b$  of the graph. We observe that for such highly clustered graphs, the best candidate for subgraph  $S$  is the smallest cluster of nodes that contain both nodes  $a$  and  $b$ . In this example,  $S_2$  would be the subgraph to which we apply Theorem 11.



Clusters of nodes in a highly clustered graph can be identified in a systematic way using hierarchical graph clustering methods such as the approach by Shi and Malik (2000) applied to image segmentation problems. Interested readers could refer to Acemoglu et al. (2009) for a hierarchical clustering algorithm designed to obtain a tight bound on the mean commute time between two particular nodes.

## 6. Conclusions

This paper analyzes the spread of misinformation in large societies. Our analysis is motivated by the widespread differences in beliefs across societies and more explicitly, the presence of many societies in which beliefs that appear to contradict the truth can be widely held. We argued that the possibility that such misinformation can arise and spread is the manifestation of the natural tension between information aggregation and misinformation spreading in the society.

We modeled a society as a social network of agents communicating (meeting) with each other. Each individual holds a belief represented by a scalar. Individuals meet pairwise and exchange information, which is modeled as both individuals adopting the average of their pre-meeting beliefs. When all individuals engage in this type of information exchange, the society will be able to aggregate the initial information held by all individuals. This effective information aggregation forms the benchmark against which we compared the possible spread of misinformation.

Misinformation is introduced by allowing some agents to be “forceful,” meaning that they influence the beliefs of (some of) the other individuals they meet, but do not change their own opinions. When the influence of forceful agents is taken into account, this defines a stochastic process for belief evolution, and our analysis exploited the fact that this stochastic process (Markov chain) can be decomposed into a part induced by the social network matrix and a part corresponding to the influence matrix.

Under the assumption that even forceful agents obtain some information (however infrequently) from some others, we first show that beliefs in this class of societies converge to a consensus among all individuals (under some additional weak regularity conditions). This consensus value is a random variable, and the bulk of our analysis characterizes its behavior, in particular, providing bounds on how much this consensus can differ from the efficient information aggregation benchmark.

We presented three sets of results. Our first set of results quantify the extent of misinformation in the society as a function of the number and properties of forceful agents and the mixing properties of the Markov chain induced by the social network matrix. In particular, we showed that social network matrices with large second eigenvalues, or that correspond to fast-mixing graphs, will place tight bounds on the extent of misinformation. The intuition for this result is that in such societies individuals that ultimately have some influence on the beliefs of forceful agents rapidly inherit the beliefs of the rest of the society and thus the beliefs of forceful agents ultimately approach those of the rest of the society and cannot have a large impact on the consensus beliefs. The extreme example is provided by expander graphs, where, when the number and the impact of forceful agents is finite, the extent of misinformation becomes arbitrarily small as the society becomes large. In contrast, the worst outcomes are obtained when there are several forceful agents and forceful agents themselves update their beliefs only on the basis of information they obtain from individuals most likely to have received their own information previously (i.e., when the graph is slow-mixing).

Our second set of results exploit more explicitly the location of forceful agents within a social network. A given social network will lead to very different types of limiting behavior depending on the context in which the forceful agents are located. We provided a tight characterization for graphs with forceful bridges, that is, graphs representing societies in which a forceful agent links two disconnected clusters. Such graphs approximate situations in which forceful agents, such as media outlets or political leaders, themselves obtain all of their information from a small group of other individuals. The interesting and striking result in this case is that the excess influence of all of the members of the small group are the same, even if some of them are not directly linked to forceful agents. We then extended these findings to more general societies using the notion of information bottlenecks.

We view our paper as a first attempt in quantifying misinformation in society. As such, we made several simplifying assumptions and derived characterization results that apply to general societies. Many areas of future investigation stem from this endeavor. First, it is important to consider scenarios in which learning and information updating are, at least partly, Bayesian. Our non-Bayesian framework is a natural starting point, both because it is simpler to analyze and because the notion of misinformation is more difficult to introduce in Bayesian models. Nevertheless, game theoretic models of communication can be used for analyzing situations in which a sender may explicitly try to mislead one or several receivers. Second, one can combine a model of communication along the lines of our setup with individuals taking actions with immediate payoff consequences and also updating on the basis of their payoffs. Misinformation will then have short-run payoff consequences, but whether it will persist or not will depend on how informative payoffs are and on the severity of its short-run payoff consequences. Third, it would be useful to characterize what types of social networks are more robust to the introduction of misinformation and how agents might use simple rules in order to avoid misinformation.

Finally, our approach implies that the society (social network) will ultimately reach a consensus, even though this consensus opinion is a random variable. In practice, there are widespread differences in beliefs in almost all societies. There is little systematic analysis of such differences in beliefs in the literature at the moment, and this is clearly an important and challenging area for future research. Our framework suggests two fruitful lines of research. First, although a stochastic consensus is eventually reached in our model, convergence can be very slow. Thus characterizing the rate of convergence to consensus in this class of models might provide insights about what types of societies and which sets of issues should

lead to such belief differences. Second, if we relax the assumption that even forceful agents necessarily obtain some (albeit limited) information from others, thus removing the “no man is an island” feature, then it can be shown that the society will generally not reach a consensus. Nevertheless, characterizing differences in opinions in this case is difficult and requires a different mathematical approach. We plan to investigate this issue in future work.

#### Appendix A. Preliminary lemmas, Sections 3 and 4

This appendix presents two lemmas that will be used in proving the convergence of agent beliefs (i.e., Theorem 1) and in establishing properties of the social network matrix  $T$  in Appendix C.

The first lemma provides conditions under which a nonnegative  $n \times n$  matrix  $M$  is *primitive*, i.e., there exists a positive integer  $k$  such that all entries of the  $k$ th power of  $M$ ,  $M^k$ , are positive (see Seneta, 1985). The lemma also provides a positive uniform lower bound on the entries of the matrix  $M^k$  as a function of the entries of  $M$  and the properties of the graph induced by the positive entries of matrix  $M$ . A version of this lemma was established in Nedić and Ozdaglar (forthcoming). We omit the proof here since it is not directly relevant to the rest of the analysis.

**Lemma 5.** *Let  $H$  be a nonnegative  $n \times n$  matrix that satisfies the following conditions:*

- (a) *The diagonal entries of  $H$  are positive, i.e.,  $H_{ii} > 0$  for all  $i$ .*
- (b) *Let  $\mathcal{E}$  denote a set of edges such that the graph  $(\mathcal{N}, \mathcal{E})$  is connected. For all  $(i, j) \in \mathcal{E}$ , the entry  $H_{ij}$  is positive, i.e.,  $\mathcal{E} \subset \{(i, j) \mid H_{ij} > 0\}$ .*

*Let  $d$  denote the maximum shortest path length between any  $i, j$  in the induced graph  $(\mathcal{N}, \mathcal{E})$ , and  $\eta > 0$  be a scalar given by*

$$\eta = \min \left\{ \min_{i \in \mathcal{N}} H_{ii}, \min_{(i,j) \in \mathcal{E}} H_{ij} \right\}.$$

*Then, we have*

$$[H^d]_{ij} \geq \eta^d \quad \text{for all } i, j.$$

The second lemma considers a sequence  $z(k)$  generated by a linear time-varying update rule, i.e., given some  $z(0)$ , the sequence  $\{z(k)\}$  is generated by

$$z(k) = H(k)z(k-1) \quad \text{for all } k \geq 0,$$

where  $H(k)$  is a stochastic matrix for all  $k \geq 0$ . We introduce the matrices  $\tilde{\Phi}(k, s) = H(k)H(k-1) \cdots H(s)$  to relate  $z(k+1)$  to  $z(s)$  for  $s \leq k$ , i.e.,

$$z(k+1) = \tilde{\Phi}(k, s)z(s).$$

The lemma shows that, under some assumptions on the entries of the matrix  $\tilde{\Phi}(k, s)$ , the *disagreement in the components* of  $z(k)$ , defined as the difference between the maximum and minimum components of  $z(k)$ , decreases with  $k$  and provides a bound on the amount of decrease.

**Lemma 6.** *Let  $\{H(k)\}$  be a sequence of  $n \times n$  stochastic matrices. Given any  $z(0) \in \mathbb{R}^n$ , let  $\{z(k)\}$  be a sequence generated by the linear update rule*

$$z(k) = H(k)z(k-1) \quad \text{for all } k \geq 0. \tag{34}$$

*Assume that there exist some integer  $B > 0$  and scalar  $\mu > 0$  such that*

$$[\tilde{\Phi}(s+B-1, s)]_{ij} \geq \mu \quad \text{for all } i, j, \text{ and } s \geq 0.$$

*For all  $k \geq 0$ , define  $M(k) \in \mathbb{R}$  and  $m(k) \in \mathbb{R}$  as follows:*

$$M(k) = \max_{i \in \mathcal{N}} z_i(k), \quad m(k) = \min_{i \in \mathcal{N}} z_i(k). \tag{35}$$

*Then, for all  $s \geq 0$ , we have  $n\mu \leq 1$  and*

$$M(s+B) - m(s+B) \leq (1 - n\mu)(M(s) - m(s)).$$

**Proof.** In view of the linear update rule (34), we have for all  $i$ ,

$$z_i(s+B) = \sum_{j=1}^n [\tilde{\Phi}(s+B-1, s)]_{ij} z_j(s) \quad \text{for all } s \geq 0.$$

We rewrite the preceding relation as

$$z_i(s+B) = \sum_{j=1}^n \mu z_j(s) + \sum_{j=1}^n [\hat{\Phi}(s+B-1, s)]_{ij} z_j(s), \quad (36)$$

where  $[\hat{\Phi}(s+B-1, s)]_{ij} = [\tilde{\Phi}(s+B-1, s)]_{ij} - \mu$  for all  $i, j$ . Since by assumption  $[\tilde{\Phi}(s+B-1, s)]_{ij} \geq \mu$  for all  $i, j$ , we have

$$[\hat{\Phi}(s+B-1, s)]_{ij} \geq 0 \quad \text{for all } i, j.$$

Moreover, since the matrices  $H(k)$  are stochastic, the product matrix  $\Phi(s+B-1, s)$  is also stochastic, and therefore we have

$$\sum_{j=1}^n [\hat{\Phi}(s+B-1, s)]_{ij} = 1 - n\mu \quad \text{for all } i.$$

From the preceding two relations, we obtain  $1 - n\mu \geq 0$  and

$$(1 - n\mu)m(s) \leq \sum_{j=1}^n [\hat{\Phi}(s+B-1, s)]_{ij} z_j(k) \leq (1 - n\mu)M(s),$$

where  $m(s)$  and  $M(s)$  are defined in Eq. (35). Combining this relation with Eq. (36), we obtain for all  $i$

$$(1 - n\mu)m(s) \leq z_i(s+B) - \sum_{j=1}^n \mu z_j(s) \leq (1 - n\mu)M(s).$$

Since this relation holds for all  $i$ , we have

$$(1 - n\mu)m(s) \leq m(s+B) - \sum_{j=1}^n \mu z_j(s),$$

$$M(s+B) - \sum_{j=1}^n \mu z_j(s) \leq (1 - n\mu)M(s),$$

from which we obtain

$$M(s+B) - m(s+B) \leq (1 - n\mu)(M(s) - m(s)) \quad \text{for all } s \geq 0. \quad \square$$

## Appendix B. Properties of the mean interaction and transition matrices, Sections 3 and 4

We establish some properties of the mean interaction matrix  $\tilde{W}$  and the transition matrices  $\Phi(k, s)$  under the assumptions discussed in Section 2.2. Recall that transition matrices are given by

$$\Phi(k, s) = W(k)W(k-1) \cdots W(s+1)W(s) \quad \text{for all } k \text{ and } s \text{ with } k \geq s, \quad (37)$$

with  $\Phi(k, k) = W(k)$  for all  $k$ . Also note that the mean interaction matrix is given by  $\tilde{W} = E[W(k)]$  for all  $k$ . In view of the belief update model (4)–(5), the entries of the matrix  $\tilde{W}$  can be written as follows. For all  $i \in \mathcal{N}$ , the diagonal entries are given by

$$[\tilde{W}]_{ii} = 1 - \frac{\sum_{j \neq i} (p_{ij} + p_{ji})}{n} + \frac{1}{n} \left[ \sum_{j \neq i} p_{ij} \left( \frac{\beta_{ij}}{2} + \alpha_{ij}\epsilon + \gamma_{ij} \right) + \sum_{j \neq i} p_{ji} \left( \frac{\beta_{ji}}{2} + \alpha_{ji} + \gamma_{ji} \right) \right], \quad (38)$$

and for all  $i \neq j \in \mathcal{N}$ , the off-diagonal entries are given by

$$[\tilde{W}]_{ij} = \frac{1}{n} \left[ p_{ij} \left( \frac{\beta_{ij}}{2} + \alpha_{ij}(1 - \epsilon) \right) + p_{ji} \frac{\beta_{ji}}{2} \right]. \quad (39)$$

Using the assumptions of Section 2.2, Lemma 5, and the explicit expressions for the entries of the matrix  $\tilde{W}$ , we have the following result.

**Lemma 7.** Let  $d$  be the maximum shortest path length between any  $i, j$  in the graph  $(\mathcal{N}, \mathcal{E})$  [cf. Eq. (3)], and  $\eta$  be a scalar given by

$$\eta = \min \left\{ \min_{i \in \mathcal{N}} [\tilde{W}]_{ii}, \min_{(i,j) \in \mathcal{E}} [\tilde{W}]_{ij} \right\} \quad (40)$$

[cf. Eqs. (38) and (39)].

(a) The scalar  $\eta$  is positive and we have

$$[\tilde{W}^d]_{ij} \geq \eta^d \text{ for all } i, j.$$

(b) We have

$$P\left\{[\Phi(s+d-1, s)]_{ij} \geq \frac{\eta^d}{2}\right\} \geq \frac{\eta^d}{2} \text{ for all } s \geq 0, i, \text{ and } j.$$

**Proof.** (a) We show that under Assumptions 1 and 3, the mean interaction matrix  $\tilde{W}$  has positive diagonal entries and the set  $\mathcal{E}$  [cf. Eq. (2)] is a subset of the link set induced by the positive elements of  $\tilde{W}$ . Together with the connectivity assumption, part (a) then follows from Lemma 5.

By Assumption 1, we have for all  $i$ ,  $\sum_{j \neq i} p_{ij} = 1$  and  $p_{ij} \geq 0$  for all  $j$ . This implies that  $\sum_{j \neq i} p_{ji} \leq n-1$  and therefore

$$1 - \frac{\sum_{j \neq i} (p_{ij} + p_{ji})}{n} \geq 0 \text{ for all } i.$$

Since  $\sum_{j \neq i} p_{ij} = 1$  for all  $i$ , there exists some  $j$  such that  $p_{ij} > 0$ , i.e.,  $(i, j) \in \mathcal{E}$ . In view of the information exchange model, we have  $\beta_{ij} > 0$  or  $\alpha_{ij} > 0$  or  $\gamma_{ij} > 0$ , implying that

$$p_{ij} \left( \frac{\beta_{ij}}{2} + \alpha_{ij}\epsilon + \gamma_{ij} \right) > 0.$$

Combining the preceding two relations with Eq. (38), we obtain

$$[\tilde{W}]_{ii} > 0 \text{ for all } i. \quad (41)$$

We next show that for any link  $(i, j)$  in the set  $\mathcal{E}$ , the entry  $[\tilde{W}]_{ij}$  is positive, i.e.,

$$\mathcal{E} \subset \{(i, j) \mid [\tilde{W}]_{ij} > 0\}.$$

For any  $(i, j) \in \mathcal{E}$ , we have  $p_{ij} > 0$ , and therefore  $\beta_{ij} + \alpha_{ij} > 0$  (cf. Assumption 3). This implies that

$$p_{ij} \left( \frac{\beta_{ij}}{2} + \alpha_{ij}(1 - \epsilon) \right) > 0,$$

which by Eq. (39) yields  $[\tilde{W}]_{ij} > 0$ . Together with Eq. (41), this shows that the scalar  $\eta$  defined in (40) is positive. By Assumption 2, the graph  $(\mathcal{N}, E)$  is connected. Using the identification  $H = \tilde{W}$  in Lemma 5, we see that the conditions of this lemma are satisfied, establishing part (a).

(b) For all  $i, j$  and  $s \geq 0$ , we have

$$\begin{aligned} P\left\{[\Phi(s+d-1, s)]_{ij} \geq \frac{\eta^d}{2}\right\} &= P\left\{1 - [\Phi(s+d-1, s)]_{ij} \leq 1 - \frac{\eta^d}{2}\right\} \\ &= 1 - P\left\{1 - [\Phi(s+d-1, s)]_{ij} \geq 1 - \frac{\eta^d}{2}\right\}. \end{aligned} \quad (42)$$

The Markov inequality states that for any nonnegative random variable  $Y$  with a finite mean  $E[Y]$ , the probability that the outcome of the random variable  $Y$  exceeds any given scalar  $\delta > 0$  satisfies

$$P\{Y \geq \delta\} \leq \frac{E[Y]}{\delta}.$$

By applying the Markov inequality to the random variable  $1 - [\Phi(s+d-1, s)]_{ij}$  [which is nonnegative and has a finite expectation in view of the stochasticity of the matrix  $\Phi(s+d-1, s)$  for all  $s \geq 0$ ], we obtain

$$P\left\{1 - [\Phi(s+d-1, s)]_{ij} \geq 1 - \frac{\eta^d}{2}\right\} \leq \frac{E[1 - [\Phi(s+d-1, s)]_{ij}]}{1 - \eta^d/2}.$$

Combining with Eq. (42), this yields

$$P\left\{[\Phi(s+d-1, s)]_{ij} \geq \frac{\eta^d}{2}\right\} \geq 1 - \frac{E[1 - [\Phi(s+d-1, s)]_{ij}]}{1 - \eta^d/2}. \quad (43)$$

By the definition of the transition matrices [cf. Eq. (37)], we have

$$E[\Phi(s+d-1, s)] = E[W(s+d-1)W(s+d-2) \cdots W(s)] = \tilde{W}^d,$$

where the second equality follows from the assumption that  $W(k)$  is independent and identically distributed over  $k$ . By part (a), this implies that

$$[E[\Phi(s+d-1, s)]]_{ij} \geq \eta^d \quad \text{for all } i, j,$$

which combined with Eq. (43) yields

$$P\left\{[\Phi(s+d-1, s)]_{ij} \geq \frac{\eta^d}{2}\right\} \geq 1 - \frac{1 - \eta^d}{1 - \eta^d/2} = \frac{\eta^d/2}{1 - \eta^d/2} \geq \frac{\eta^d}{2},$$

establishing the desired result.  $\square$

The next two lemmas establish properties of transition matrices.

**Lemma 8.**

- (a)  $[\Phi(k, s)]_{ii} \geq \epsilon^{k-s+1}$  for all  $k$  and  $s$  with  $k \geq s$ , and all  $i \in \mathcal{N}$ .  
 (b) Assume that there exist integers  $K, B \geq 1$  and a scalar  $\xi > 0$  such that for some  $s \geq 0$  and  $k \in \{0, \dots, K\}$ , we have

$$[\Phi(s + (k+1)B - 1, s + kB)]_{ij} \geq \xi \quad \text{for some } i, j.$$

Then,

$$[\Phi(s + KB - 1, s)]_{ij} \geq \xi \epsilon^{K-1}.$$

**Proof.** (a) We let  $s$  be arbitrary and prove the relation by induction on  $k$ . By the definition of the transition matrices [cf. Eq. (37)], we have  $\Phi(s, s) = W(s)$ . Thus, the relation  $[\Phi(k, s)]_{ii} \geq \epsilon^{k-s+1}$  holds for  $k = s$  from the definition of the update matrix  $W(k)$  [cf. Eq. (5)]. Suppose now that the relation holds for some  $k > s$  and consider  $[\Phi(k+1, s)]_{ii}$ . We have

$$[\Phi(k+1, s)]_{ii} = \sum_{h=1}^n [W(k+1)]_{ih} [\Phi(k, s)]_{hi} \geq [W(k+1)]_{ii} [\Phi(k, s)]_{ii} \geq \epsilon^{k-s+2},$$

where the first inequality follows from the nonnegativity of the entries of  $\Phi(k, s)$ , and the second inequality follows from the inductive hypothesis.

- (b) For any  $s \geq 0$ , we have

$$\begin{aligned} [\Phi(s + KB - 1, s)]_{ij} &= \sum_{h=1}^n [\Phi(s + KB - 1, s + (k+1)B)]_{ih} [\Phi(s + (k+1)B - 1, s)]_{hj} \\ &\geq [\Phi(s + KB - 1, s + (k+1)B)]_{ii} [\Phi(s + (k+1)B - 1, s)]_{ij} \\ &\geq \epsilon^{(K-k-1)B} [\Phi(s + (k+1)B - 1, s)]_{ij}, \end{aligned}$$

where the last inequality follows from part (a). Similarly,

$$\begin{aligned} [\Phi(s + (k+1)B - 1, s)]_{ij} &= \sum_{h=1}^n [\Phi(s + (k+1)B - 1, s + kB)]_{ih} [\Phi(s + kB - 1, s)]_{hj} \\ &\geq [\Phi(s + (k+1)B - 1, s + kB)]_{ij} [\Phi(s + kB - 1, s)]_{jj} \\ &\geq \xi \epsilon^{kB}, \end{aligned}$$

where the second inequality follows from the assumption  $[\Phi(s + (k+1)B - 1, s + kB)]_{ij} \geq \xi$  and part (a). Combining the preceding two relations yields the desired result.  $\square$

**Lemma 9.** We have

$$P\left\{[\Phi(s + n^2d - 1, s)]_{ij} \geq \frac{\eta^d}{2} \epsilon^{n^2-1} \text{ for all } i, j\right\} \geq \left(\frac{\eta^d}{2}\right)^{n^2} \quad \text{for all } s \geq 0,$$

where the scalar  $\eta > 0$  and the integer  $d$  are the constants defined in Lemma 7.

**Proof.** Consider a particular ordering of the elements of an  $n \times n$  matrix and let  $k_{ij} \in \{0, \dots, n^2 - 1\}$  denote the unique index for element  $(i, j)$ . From Lemma 8(b), we have

$$\begin{aligned} & P \left\{ [\Phi(s + n^2d - 1, s)]_{ij} \geq \frac{\eta^d}{2} \epsilon^{n^2-1} \text{ for all } i, j \right\} \\ & \geq P \left\{ [\Phi(s + (k_{ij} + 1)d - 1, s + k_{ij}d)]_{ij} \geq \frac{\eta^d}{2} \text{ for all } i, j \right\} \\ & = \prod_{(i,j)} P \left\{ [\Phi(s + (k_{ij} + 1)d - 1, s + k_{ij}d)]_{ij} \geq \frac{\eta^d}{2} \right\} \\ & \geq \left( \frac{\eta^d}{2} \right)^{n^2}. \end{aligned}$$

Here the second equality follows from the independence of the random events

$$\left\{ [\Phi(s + (k + 1)d - 1, s + kd)]_{ij} \geq \frac{\eta^d}{2} \right\}$$

over all  $k = 0, \dots, n^2 - 1$ , and the last inequality follows from Lemma 7(b).  $\square$

#### Appendix C. Properties of the social network matrix, Section 4

The next lemma studies the properties of the social network matrix  $T$ . Note that the entries of the matrix  $T$  can be written as follows: For all  $i \in \mathcal{N}$ , the diagonal entries are given by

$$[T]_{ii} = 1 - \frac{\sum_{j \neq i} (p_{ij} + p_{ji})}{n} + \frac{1}{n} \left[ \sum_{j \neq i} p_{ij} \left( \frac{1 - \gamma_{ij}}{2} + \gamma_{ij} \right) + \sum_{j \neq i} p_{ji} \left( \frac{1 - \gamma_{ji}}{2} + \gamma_{ji} \right) \right], \quad (44)$$

and for all  $i \neq j \in \mathcal{N}$ , the off-diagonal entries are given by

$$[T]_{ij} = \frac{1}{n} \left[ p_{ij} \frac{1 - \gamma_{ij}}{2} + p_{ji} \frac{1 - \gamma_{ji}}{2} \right]. \quad (45)$$

**Lemma 10.** Let  $T$  be the social network matrix [cf. Eq. (9)]. Then, we have:

(a) The matrix  $T^k$  converges to a stochastic matrix with identical rows  $\frac{1}{n}e$  as  $k$  goes to infinity, i.e.,

$$\lim_{k \rightarrow \infty} T^k = \frac{1}{n}ee'.$$

(b) For any  $z(0) \in \mathbb{R}^n$ , let the sequence  $z(k)$  be generated by the linear update rule

$$z(k) = Tz(k-1) \quad \text{for all } k \geq 0.$$

For all  $k \geq 0$ , define  $M(k) \in \mathbb{R}$  and  $m(k) \in \mathbb{R}$  as follows:

$$M(k) = \max_{i \in \mathcal{N}} z_i(k), \quad m(k) = \min_{i \in \mathcal{N}} z_i(k).$$

Then, for all  $k \geq 0$ , we have

$$M(k) - m(k) \leq \delta^k (M(0) - m(0)).$$

Here  $\delta > 0$  is a constant given by

$$\begin{aligned} \delta &= (1 - n\chi^d)^{\frac{1}{d}}, \\ \chi &= \min_{(i,j) \in \mathcal{E}} \left\{ \frac{1}{n} \left[ p_{ij} \frac{1 - \gamma_{ij}}{2} + p_{ji} \frac{1 - \gamma_{ji}}{2} \right] \right\}, \end{aligned}$$

and  $d$  is the maximum shortest path length in the graph  $(\mathcal{N}, \mathcal{E})$  [cf. Eq. (3)].

**Proof.** (a) By Assumption 1, we have for all  $i$ ,  $\sum_{j \neq i} p_{ij} = 1$  and  $p_{ij} \geq 0$  for all  $j$ . This implies that  $\sum_{j \neq i} p_{ji} \leq n - 1$  and therefore

$$1 - \frac{\sum_{j \neq i} (p_{ij} + p_{ji})}{n} \geq 0 \quad \text{for all } i. \quad (46)$$

Since  $\sum_{j \neq i} p_{ij} = 1$  for all  $i$ , there exists some  $j$  such that  $p_{ij} > 0$ , i.e.,  $(i, j) \in \mathcal{E}$ . By Assumption 3, this implies that  $\beta_{ij} + \alpha_{ij} = 1 - \gamma_{ij} > 0$ , showing that  $T_{ii} > 0$  for all  $i$ . Similarly, for any  $(i, j) \in \mathcal{E}$ , we have  $p_{ji} > 0$  and therefore  $1 - \gamma_{ij} > 0$ , showing that  $T_{ij} > 0$  for all  $(i, j) \in \mathcal{E}$ . Using Eq. (46) in Eqs. (44) and (45), it follows that for all  $i$

$$[T]_{ii} \geq T_{ij} \quad \text{for all } j.$$

Thus, we can use Lemma 5 with the identification

$$\chi = \min_{(i,j) \in \mathcal{E}} \left\{ \frac{1}{n} \left[ p_{ij} \frac{1 - \gamma_{ij}}{2} + p_{ji} \frac{1 - \gamma_{ji}}{2} \right] \right\}, \quad (47)$$

and obtain

$$[T^d]_{ij} \geq \chi^d \quad \text{for all } i, j, \quad (48)$$

i.e.,  $T$  is a primitive matrix and therefore the Markov chain with transition probability matrix  $T$  is regular. It follows from Theorem 3(a) that for any  $z(0) \in \mathbb{R}^n$ , we have

$$\lim_{k \rightarrow \infty} T^k z(0) = e \bar{z},$$

where  $\bar{z}$  is given by  $\bar{z} = \pi' z(0)$  for some probability vector  $\pi$ . Since  $T$  is a stochastic and symmetric matrix, it is doubly stochastic. Denoting  $z(k) = T^k z(0)$ , this implies that the average of the entries of the vector  $z(k)$  is the same for all  $k$ , i.e.,

$$\frac{1}{n} \sum_{i=1}^n z_i(k) = \frac{1}{n} \sum_{i=1}^n z_i(0) \quad \text{for all } k \geq 0.$$

Combining the preceding two relations, we obtain

$$\lim_{k \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n z_i(k) = \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i(0),$$

establishing the desired relation.

(b) In view of Eq. (48), we can use Lemma 6 with the identifications

$$H(k) = T, \quad B = d, \quad \mu = \chi^d,$$

where  $\chi$  is defined in Eq. (47), and obtain

$$M(k) - m(k) \leq (1 - n\chi^d)^{\frac{k}{d}} (M(0) - m(0)). \quad \square$$

**Lemma 11.** Let  $D^{ij} = \frac{p_{ij}\alpha_{ij}}{n} [J^{ij} - A^{ij}]$ , where  $J^{ij}$  and  $A^{ij}$  are defined in (5). Also, let  $Y$  be the fundamental matrix corresponding to a Markov chain with transition matrix  $T$  given by (9). Then

$$D^{ij} Y (I - D^{ij} Y)^{-1} = \left( 1 - \frac{\zeta_{ij}}{n^2} \right)^{-1} D^{ij} Y,$$

where

$$\zeta_{ij} = \frac{p_{ij}\alpha_{ij}}{2} [m_{ij} - (1 - 2\epsilon)m_{ji}],$$

in which  $m_{ij}$  is the mean first passage time of the Markov chain from state  $i$  to state  $j$ .

**Proof.** Let us first show the following equality:

$$(D^{ij} Y)^2 = \left( \frac{\zeta_{ij}}{n^2} \right) D^{ij} Y. \quad (49)$$

This can be confirmed by expanding the left-hand side as follows:

$$\begin{aligned}
(D^{ij}Y)^2 &= \left( \frac{p_{ij}\alpha_{ij}}{n} [J^{ij} - A^{ij}]Y \right)^2 \\
&= \left( \frac{p_{ij}\alpha_{ij}}{n} \left[ \left( \epsilon - \frac{1}{2} \right) e_i - \frac{1}{2} e_j \right] (e_i - e_j)' Y \right)^2 \\
&= \left( \frac{p_{ij}\alpha_{ij}}{n} \right)^2 \left[ \left( \epsilon - \frac{1}{2} \right) e_i - \frac{1}{2} e_j \right] (e_i - e_j)' Y \left[ \left( \epsilon - \frac{1}{2} \right) e_i - \frac{1}{2} e_j \right] (e_i - e_j)' Y \\
&= \left( \frac{p_{ij}\alpha_{ij}}{n} \right)^2 \left[ \left( \epsilon - \frac{1}{2} \right) e_i - \frac{1}{2} e_j \right] \left( \left( \epsilon - \frac{1}{2} \right) (Y_{ii} - Y_{ji}) - \frac{1}{2} (Y_{ij} - Y_{jj}) \right) (e_i - e_j)' Y \\
&= \left( \frac{p_{ij}\alpha_{ij}}{n} \right) \left( \left( \epsilon - \frac{1}{2} \right) (Y_{ii} - Y_{ji}) - \frac{1}{2} (Y_{ij} - Y_{jj}) \right) D^{ij}Y \\
&= \left( \frac{p_{ij}\alpha_{ij}}{2n^2} \right) (m_{ij} - (1 - 2\epsilon)m_{ji}) D^{ij}Y,
\end{aligned}$$

where the last equality is a direct result of Lemma 1(ii). Now, we can write

$$D^{ij}Y(I - D^{ij}Y) = D^{ij}Y - (D^{ij}Y)^2 = \left( 1 - \frac{\zeta_{ij}}{n^2} \right) D^{ij}Y.$$

The claimed result follows from multiplying both sides of the preceding equation by  $(1 - \zeta_{ij}/n^2)^{-1}(I - D^{ij}Y)^{-1}$  from the right. Note that  $(I - D^{ij}Y)$  is nonsingular by Theorem 4.  $\square$

#### Appendix D. Characterization of the mean commute time, Section 5

First, we characterize the mean commute time between two nodes for a random walk on an undirected graph using Dirichlet's principle and its dual, Thompson's principle.

**Definition 6.** Consider a random walk on a weighted undirected graph  $(\mathcal{N}, \mathcal{A})$  with weight  $w_{ij}$  associated to each edge  $\{i, j\}$ . Define the *Dirichlet form*  $\mathcal{E}$ , as follows. For functions  $g: \mathcal{N} \rightarrow \mathbb{R}$  write

$$\mathcal{E}(g, g) = \frac{1}{2} \sum_{i,j} \frac{w_{ij}}{w} (g(i) - g(j))^2,$$

where  $w = \sum_{i,j} w_{ij}$  is the total edge weight.

**Lemma 12.** Consider a random walk on a weighted undirected graph with weight  $w_{ij}$  associated to each edge  $\{i, j\}$ . For mean commute time between distinct nodes  $a$  and  $b$  we have

$$m_{ab} + m_{ba} = \sup \left\{ \frac{1}{\mathcal{E}(g, g)} : 0 \leq g \leq 1, g(a) = 0, g(b) = 1 \right\} \quad (50)$$

$$= w \inf \left\{ \frac{1}{2} \sum_{i,j} \frac{f_{ij}^2}{w_{ij}} : \mathbf{f} \text{ is a unit flow from } a \text{ to } b \right\}, \quad (51)$$

where  $m_{ab}$  is the mean first passage time from  $a$  to  $b$ ,  $w$  is the total edge weight, and a unit flow  $\mathbf{f}$  from  $a$  to  $b$  satisfies the following conditions:

$$\begin{aligned}
f_{ij} &= -f_{ji}, \\
\sum_{j \neq i} f_{ij} &= \begin{cases} 1 & \text{if } i = a, \\ -1 & \text{if } i = b, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

**Proof.** See Section 7.2 of Aldous and Fill (2002).  $\square$

It is worth mentioning that the two forms of the mean commute time characterization in Lemma 12 are dual of each other. The first form is a corollary of Dirichlet principle, while the second is immediate result of Thompson's principle. Using the electric circuit analogy, we can think of function  $g(i)$  as potential associated to node  $i$ , and flow  $f_{ij}$  as the current on edge  $\{i, j\}$  with resistance  $\frac{1}{w_{ij}}$ . The expressions in (51) are equivalent descriptions of minimum energy dissipation in such an electric network. Hence, we can interpret the mean commute time between two particular nodes as the effective



resistance between such nodes in a resistive network. This allows us to use the Monotonicity law to obtain simpler bounds for mean commute time.

**Lemma 13** (Monotonicity law). Let  $\tilde{w}_{ij} \leq w_{ij}$  be the edge weights for two undirected graphs. Then,

$$m_{av} + m_{va} \leq \left( \frac{w}{\tilde{w}} \right) (\tilde{m}_{av} + \tilde{m}_{va}) \quad \text{for all } a, v,$$

where  $w = \sum_{i,j} w_{ij}$  and  $\tilde{w} = \sum_{i,j} \tilde{w}_{ij}$  are the total edge weights.

**Proof.** Let  $f^*$  and  $\tilde{f}^*$  be the optimal solutions of (51) for the original and modified graphs, respectively. We can write

$$\begin{aligned} m_{av} + m_{va} &= \frac{w}{2} \sum_{i,j} \frac{(f_{ij}^*)^2}{w_{ij}} \leq \frac{w}{2} \sum_{i,j} \frac{(\tilde{f}_{ij}^*)^2}{w_{ij}} \\ &\leq \frac{w}{2} \sum_{i,j} \frac{(\tilde{f}_{ij}^*)^2}{\tilde{w}_{ij}} = \left( \frac{w}{\tilde{w}} \right) \frac{\tilde{w}}{2} \sum_{i,j} \frac{(\tilde{f}_{ij}^*)^2}{\tilde{w}_{ij}} \\ &= \left( \frac{w}{\tilde{w}} \right) (\tilde{m}_{av} + \tilde{m}_{va}), \end{aligned}$$

where the first inequality follows from optimality of  $f^*$  and feasibility of  $\tilde{f}^*$ .  $\square$

By the electric network analogy, Lemma 13 states that increasing resistance in a circuit increases the effective resistance between any two nodes in the network. The Monotonicity law can be extremely useful in providing simple bounds for mean commute times.

## References

- Acemoglu, D., Dahleh, Munther, Lobel, Ilan, Ozdaglar, A., 2008. Bayesian learning in social networks. Preprint.
- Acemoglu, D., Ozdaglar, A., ParandehGheibi, A., 2009. Spread of (mis)information in social networks, LIDS report 2812.
- Aldous, D., Fill, J., 2002. Reversible Markov chains and random walks on graphs. Monograph, <http://www.stat.berkeley.edu/aldous/RWG/book.html>.
- Ambrus, A., Takahashi, S., 2008. Multi-sender cheap talk with restricted state spaces. Theoretical Econ. 3 (1), 1–27.
- Bala, V., Goyal, S., 1998. Learning from neighbours. Rev. Econ. Stud. 65 (3), 595–621.
- Bala, V., Goyal, S., 2001. Conformism and diversity under social learning. Econ. Theory 17, 101–120.
- Banerjee, A., 1992. A simple model of herd behavior. Quart. J. Econ. 107, 797–817.
- Banerjee, A., Fudenberg, D., 2004. Word-of-mouth learning. Games Econ. Behav. 46, 1–22.
- Bikchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as information cascades. J. Polit. Economy 100, 992–1026.
- Boyd, S., Ghosh, A., Prabhakar, B., Shah, D., 2005. Gossip algorithms: Design, analysis, and applications. In: Proceedings of IEEE INFOCOM.
- Bremaud, P., 1999. Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Springer, New York.
- Celen, B., Kariv, S., 2004a. Distinguishing informational cascades from herd behavior in the laboratory. Amer. Econ. Rev. 94 (3), 484–498.
- Celen, B., Kariv, S., 2004b. Observational learning under imperfect information. Games Econ. Behav. 47 (1), 72–86.
- Chung, F.R.K., 1997. Spectral Graph Theory. American Mathematical Society, Providence, RI.
- Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50 (6), 1431–1451.
- DeGroot, M.H., 1974. Reaching a consensus. J. Amer. Statistical Assoc. 69 (345), 118–121.
- DeMarzo, P.M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. Quart. J. Econ. 118 (3), 909–968.
- Fagnani, F., Zampieri, S., forthcoming. Randomized consensus algorithms over large scale networks. IEEE J. Sel. Areas Commun.
- Farrell, J., Gibbons, R., 1989. Cheap talk with two audiences. Amer. Econ. Rev. 79, 1214–1223.
- Gale, D., Kariv, S., 2003. Bayesian learning in social networks. Games Econ. Behav. 45 (2), 329–346.
- Galeotti, A., Ghiglino, C., Squintani, F., 2009. Strategic information transmission in networks. Preprint.
- Golub, B., Jackson, M.O., 2008. How homophily affects diffusion and learning in networks. Preprint.
- Golub, B., Jackson, M.O., forthcoming. Naive learning in social networks: Convergence, influence, and the wisdom of crowds. Amer. Econ. J. Microecon.
- Hagenbach, J., Koessler, F., 2009. Strategic communication networks. Preprint.
- Haviv, M., Van Der Heyden, L., 1984. Perturbation bounds for the stationary probabilities of a finite Markov chain. Advanc. Appl. Prob. 16 (4), 804–818.
- Jackson, M.O., 2008. Social and Economic Networks. Princeton University Press, Princeton, NJ.
- Jadbabaie, A., Lin, J., Morse, S., 2003. Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Trans. Automat. Control 48 (6), 988–1001.
- Kemeny, J.G., Snell, J.L., 1960. Finite Markov Chains. Van Nostrand, New York, NY.
- Nedić, A., Ozdaglar, A., forthcoming. Distributed subgradient methods for multi-agent optimization. IEEE Trans. Automat. Control.
- Olfati-Saber, R., Murray, R.M., 2004. Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans. Automat. Control 49 (9), 1520–1533.
- Olshchvsky, A., Tsitsiklis, J.N., forthcoming. Convergence speed in distributed consensus and averaging. SIAM J. Control Optim.
- Rudin, W., 1987. Real and Complex Analysis. McGraw-Hill, New York, NY.
- Schweitzer, P.J., 1968. Perturbation theory and finite Markov chains. J. Appl. Prob. 5 (2), 401–413.
- Seneta, E., 1985. Nonnegative Matrices and Markov Chains. Springer, Berlin. Translated from Russian by K.C. Kiwiell and A. Ruszczynski.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Machine Intell. (PAMI) 22 (8), 888–905.

- Smith, L., Sorensen, P., 1998. Rational social learning with random sampling. Unpublished manuscript.
- Smith, L., Sorensen, P., 2000. Pathological outcomes of observational learning. *Econometrica* 68 (2), 371–398.
- Sobel, J., 2009. Signaling games. In: *Encyclopedia of Complexity and System Science*. Springer.
- Tsitsiklis, J.N., 1984. Problems in decentralized decision making and computation. PhD thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
- Tsitsiklis, J.N., Bertsekas, D.P., Athans, M., 1986. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Automat. Control* 31 (9), 803–812.
- Watts, D., 2003. *Six Degrees: The Science of a Connected Age*. W.W. Norton and Company.