Social learning by chit-chat

Edoardo Gallo*

January 2013

Abstract

Individuals learn by chit-chatting with others as a by-product of their online and offline activities. Social plugins are an example in the online context: they embed information from a friend, acquaintance or even a stranger on a webpage and the information is usually independent of the content of the webpage. We formulate a novel framework to investigate how the speed of learning by chit-chat depends on the structure of the environment. A network represents the environment that individuals navigate to interact with each other. We derive an exact formula to compute how the expected time between meetings depends on the underlying network structure and we use this quantity to investigate the speed of learning in the society. Comparative statics show that the speed of learning is sensitive to a mean-preserving spread of the degree distribution (MPS). Specifically, if the number of individuals is low (high), then a MPS of the network increases (decreases) the speed of learning. The speed of learning is the same for all regular networks independent of network connectivity. An extension explores the effectiveness of one agent, the influencer, at influencing the learning process.

Keywords: social learning, network, speed of learning, mean preserving spread, influencer.

JEL: D83, D85.

---

1 Introduction

The advent of social plugins has significantly changed the way we learn from others online. A social plugin embeds information from a friend, acquaintance or even a stranger on a webpage that you are visiting. This information is usually independent of the content of the webpage and it can be about what this other individual likes, recommends or a comment that she has left. For instance, you may open an article on the US election on the New York Times and a social plugin will inform you that an acquaintance likes a certain product or recommends a review article about a trendy travel destination. This may in turn influence your beliefs on the best product to buy or where to go on vacation next year.

Social plugins have increased the prominence of a learning process that has always been present in the offline world. As we go about our daily activities we sometimes engage in chance conversations, or chit-chat, with friends, acquaintances or even strangers whom we happen to meet. More than 20 years ago, Huckfeldt [1986] had already pointed out the importance of chit-chat for political choices: "[t]he less intimate interactions that we have ignored - discussions over backyard fences, casual encounters while taking walks, or standing in line at the grocery store, and so on - may be politically influential even though they do not occur between intimate associates." Likewise, companies have realized the potential of chit-chat to influence consumers, and attempt to exploit it through a variety of marketing gimmicks.[1]

The objective of this paper is to formulate a model of the learning that happens through social plugins or chance conversations, which we dub *social learning by chit-chat*, and to investigate the speed of learning in this framework. This learning process is distinctly different from the learning processes captured by standard models in the economic literature, in which agents purposely learn how to play an underlying game. In social learning by chit-chat the agents are strategic as they go about their daily activities, but these activities are independent of the learning process. For instance, an individual may be strategic in the choice of news articles she reads about the US election, but the information she learns about the trendy travel hotspot is independent of these strategic considerations. The presumption of the social learning literature is that individuals *actively* learn from the decisions and experiences of their neighbors and/or social relations. In social learning by chit-chat an individual can *passively* learn from another individual independently of the presence of a social relation, as Huckfeldt [1986] points out in the context of political choices.

The main determinants of the speed of learning in the social learning by chit-chat framework are *the size of the population and the environment* these individuals are living in. For instance, if we think of the environment as a physical environment, the speed of learning may be very different in a scarcely populated village where there is one central

---

[1]For instance, see "The Body as Billboard: Your Ad Here," New York Times, February 17th, 2009.

square vis-à-vis a metropolis buzzing with people with multiple meeting points. Moreover, in this framework *individuals are learning as a by-product of their other activities* and *they do not rely exclusively on their social relations to learn*: they discuss politics with their work colleagues during coffee break, they chat about fashion trends with strangers at the gym, or they exchange opinions on the latest movie with acquaintances at the golf course.

We use environment as a broad umbrella term that encompasses a variety of meanings. A prominent meaning is *online environments*: webpages with social plugins allow individuals to exchange information while they navigate the web using hyperlinks that connect these webpages. Another meaning is *physical environments* such as the various locations in an urban area or the meeting places within an organization. Finally, the term environment may also be interpreted as an *abstract space of circumstances* that prompt individuals to exchange views on a topic. An example could be the set of topics that may be salient in an individual's mind: if the same topic is salient in two individuals' minds then they are in the same location in this abstract space and they may exchange information.

Technically, we represent the structure of these different types of environments using a *network*. The elements of the network represent the webpages, locations or circumstances where individuals can meet, and throughout the paper we will refer to these elements as *sites* as a reminder of the online environment. Individuals move across sites using links, which represent the hyperlinks, roads or mental associations that connect webpages, urban locations and circumstances respectively. The network is fixed and exogenously given throughout the paper. This is not because we believe that the environment never changes, but simply because it changes much more slowly that the learning process that we are interested in so it can be considered fixed for our purposes. For instance, the layout of an urban area changes over at least several decades, while social learning by chit-chat happens over days, weeks or at most months.

Individuals travel in the environment by randomly moving from one site to one of the neighbouring sites. This is a simplifying assumption for tractability purposes, but it is a reasonable representation of the way individuals navigate websites or move across physical locations. For instance, Google's PageRank algorithm, which determines the importance that Google's search engine gives to a webpage, models websurfers as randomly moving across sites along hyperlinks. Similarly, in the Operations Research literature, mobility models have been used extensively to model individuals' movements across physical locations. The benchmark version is the *Random Walk Mobility Model* where agents are simply random walkers.[2]

Specifically, we consider a population of n agents who are random walkers on a given network **g**. At each point in time a randomly chosen individual moves from her current site to one of the neighboring sites, with equal probability for each neighboring site. If

---

[2]See Camp et al. [2002] for a review. For instance, Zonoozi and Dassanayake [1997] use it to model the movements of mobile phone users.

there are other individuals on the destination site then one of them is randomly chosen and a meeting occurs. Otherwise, no meeting takes place. Within this framework we model learning by chit-chat as follows. At time T each agent receives a noisy, independent and identically distributed signal about the true state of the world. Every time two agents meet they truthfully reveal their current beliefs and they update their own belief by taking a weighted average of their previous belief and the belief of the agent that they have met.

In the long term all agents' beliefs converge (almost surely) to the same value, and the distribution of the limit belief has a mean equal to the average of the initial signals. Theorem 1 presents the first main result of the paper: the derivation of an explicit expression for the *expected time τ an individual waits between two meetings* as a function of the number of individuals and the network structure. The formula for τ is an exact result and the comparative statics in the paper are formally stated in terms of how τ varies with changes in the network. We interpret results on the variations in τ as variations in the speed of learning by using a mean-field approximation.

The second main result, in Theorem 2, relates the expected time between meetings, and therefore the speed of learning, to the underlying network structure. *If the number of agents is low relative to the number of sites, then a mean preserving spread of the degree distribution decreases the expected time between meetings*, i.e. the speed of learning increases with the level of variability in number of connections across sites. The intuition is that in a heterogenous environment most of the paths in the network lead to the well-connected sites, and the presence of these sites increases the probability of meetings among the small number of agents. On the other hand, *if the number of agents is high relative to the number of sites, then a mean preserving spread of the degree distribution increases the expected time between meetings*. The intuition is that in a homogenous environment the sites have a similar number of connections so the agents will be distributed uniformly across them, and this minimizes the probability that one agent moves to an unoccupied site.

One may conjecture that the density of connections in the network would also have an effect on the speed of learning. However, proposition 1 shows that *the expected time between meetings is invariant to changes in the connectivity of the environment*. Specifically, it shows that the speed of learning is the same for any environment that is represented by a regular network independent of network connectivity. The intuition is that there are two effects at play that cancel each other. On the one hand an increase in connectivity allows agents access to a larger number of sites from their current location, increasing in this way the probability of access to an occupied site where a meeting can occur. On the other hand this increase in accessibility means that if two agents are at neighboring sites then it is less likely that one of them will randomly move to the other's location, so this effect decreases the frequency of meetings in better connected environments. These two effects exactly cancel each other in regular networks, so the expected time between meetings, and therefore the speed of learning, is invariant to changes in the connectivity

4

of the network.

The last part of the paper extends the basic framework to analyze the impact that one agent, the *influencer*, has on the learning by chit-chat process. Specifically, the goal is to highlight how the structure of the environment makes a society more or less susceptible to the presence of the influencer. We model the influencer as an agent who does not change her belief and who is positioned at a site in the network, instead of traveling around for purposes independent of the learning process.

The first result is that the site that maximizes the effectiveness of the influencer is the most connected site in the network. Assuming that the influencer is at the most connected site, *a society is minimally susceptible to the presence of the influencer if and only if the underlying environment is a regular network*. Moreover, if the number of individuals in the society is above a threshold, then *a mean preserving spread of the degree distribution of the underlying network makes a society more susceptible to the presence of the influencer*. Two effects drive this result. The first one is that a direct application of Theorem 2 shows that in a more heterogeneous environment the frequency of meetings involving non-influencers decreases. The second one is that a shift to a more heterogeneous environment leads to a (weak) increase in the connectivity of the most connected site where the influencer resides, leading to an increase in the frequency of meetings involving the influencer. Thus, the overall impact of the influencer on the learning process increases.

The rest of this section presents a literature review. Section 2 describes the general framework and the learning process. Section 3 considers two stylized networks to illustrate the intuition behind the main results, which are presented in section 4. Section 5 extends the model to investigate the effect of an influencer. Appendix A contains all the proofs.

1.1 Literature review

The literature on learning in economics is extensive and the purpose here is not to give a complete overview of this large body of work. This section will focus on social learning models with a network to highlight how this paper differs from previous contributions.

Bala and Goyal [1998] examine an observational social learning model in which agents take actions at regular intervals and these actions generate noisy payoffs. Agents learn from the payoffs they receive and from observing the payoffs of their neighbors in the social network. This type of social learning is called observational in the sense that agents observe the actions, not the beliefs, of the other agents. The main result is that everyone converges to the optimal action as long as there is no subset of agents that are too influential in terms of their position in the network. Recent contributions by Acemoglu et al. [2012] and Lamberson [2010] present further results in this framework.

DeMarzo et al. [2003] investigate a different set-up in which agents receive a noisy signal of the underlying state of the world at time 0 and then they update this signal by

taking a weighted average of the beliefs of their neighbors in the social network.[3] The weights are fixed by the exogenously given network. Agents report their beliefs truthfully, and they are boundedly rational because they do not take into account repetitions of information in the social network. They show that if there is sufficient communication then the beliefs of all agents converge in the long-run to a weighted average of initial beliefs, but the information is not aggregated optimally unless an agent's prominence in the social network coincides with the accuracy of his initial signal. They also show that the speed of convergence is governed by the size of the second eigenvalue of the network matrix. Golub and Jackson [2010] generalize some of these results.

The main conceptual difference between learning by chit-chat and these previous contributions is the function played by the network. In social learning models the network constrains whom the agents can learn from, either because of physical or social constraints. The interpretation of the physical constraints is that agents have restrictions on their mobility that do not allow them to observe all other agents in the society. A plausible interpretation of the social constraints is that agents rely only upon agents they trust in order to learn about payoff-relevant information. On the other hand, in learning by chit-chat the function of the network is to model the structure of the environment where agents live and that determines the frequency of their meetings to exchange information. The presence of a social tie is not a prerequisite for exchanging information because agents are learning while interacting with a variety of individuals as they are going about their daily activities. Moreover, the type of information being exchanged is not particularly sensitive so agents learn from friends as well as from strangers. Examples include learning about the best restaurant in town, fashion trends or the latest movie release.

This major conceptual difference motivates the use of a novel methodology to investigate the speed of convergence in the learning by chit-chat context. We rely on a result by Kac [1947] in the theory of stochastic processes to explicitly compute how the expected time an agent waits between two interactions depends on the underlying network structure. This explicit solution permits a very clear comparative statics analysis that shows which type of changes in network structure have an impact on the speed of convergence, and how that depends on the number of agents in the society. This is in sharp contrast to the notorious difficulty in deriving results relating the speed of convergence to the social network structure in the social learning literature.

An important question in the social learning literature is whether the society aggregates information efficiently and converges to the payoff-maximizing outcome. In observational social learning models the spirit of the results is that there is convergence to the payoff-maximizing outcome as long as some very weak conditions on the network structure are satisfied. The usual form of these conditions is that there should not be one or a few agents that have a disproportionate influence on the rest of society. On the other hand, in the DeMarzo et al. [2003] framework information is usually not aggregated

---

[3]This learning rule was first proposed in DeGroot [1974].

optimally because agents' influence differ depending on their position in the network and their influence does not usually correlate with the informativeness of their initial signal. DeMarzo et al. [2003] show how to compute the influence of each agent on the common limiting belief.

The learning by chit-chat framework is similar to the DeMarzo et al. [2003] model because learning is not observational and agents truthfully reveal their beliefs about the underlying state of the world. However, in the basic learning by chit-chat framework the agents are interchangeable, so in the long-term each agent has the same influence on the common belief the society converges to. Our extension involving an influencer can be viewed as complementary to the approach in DeMarzo et al. [2003] because we show how the effectiveness of the influencer depends on the environment.

Finally, it is important to notice that the main results of the paper are applicable beyond the context of the learning model analyzed in this paper. The general framework is a stochastic dynamic system of agents travelling on a network and the main results relate the structure of the network with the frequency of encounters. This process is potentially relevant to other economic phenomena such as the impact of epidemics and the spread of innovations.

2 The Model

This section presents the main elements of the model: the network concepts and terminology, the way agents move on the network and the learning process.

Network. Consider a finite set of nodes $S = \{1, ..., s\}$, which will be called *sites*. The links among the sites are described by an adjacency matrix $\mathbf{g} = [g_{ij}]$, where $g_{ij} \in \{0, 1\}$ and $g_{ij} = g_{ji}$ for all $i, j \in S$ $(j \neq i)$ and $g_{ii} = 0$ and for all i. Let $N_a(\mathbf{g}) = \{b \in S | g_{ab} = 1\}$ be the *neighborhood* of site a in network $\mathbf{g}$. The *degree* $d_a(\mathbf{g}) = |N_a(\mathbf{g})|$ is the size of the neighborhood of site a in network $\mathbf{g}$, i.e. the number of sites directly connected to site a. Let $D(\mathbf{g}) = \sum_i d_i(\mathbf{g})$ be the sum of the degrees of all the sites in the network, which is equal to twice the number of links in the network.

Denote by $P(d)$ the *degree distribution* of sites in a network, and let $\mu[P(d)]$ denote the mean of the distribution. The degree distribution is a description of the relative frequencies of sites that have different degrees. The comparative statics analysis will investigate changes in the network structure that are captured by a mean-preserving spread of this distribution. The following is a more formal definition of this notion.

Definition 1. A distribution $P'(d)$ is a mean-preserving spread (MPS) of another distribution $P(d)$ if $\mu[P(d)] = \mu[P'(d)]$ and if $\sum_{Y=0}^{Z} \left[\sum_{d=0}^{Y} P'(d) - P(d) \right] \geq 0$ for all $Z \in [1, s]$.

Consider two networks $\mathbf{g}$ and $\mathbf{g}'$ with degree distributions $P(d)$ and $P'(d)$ respectively. If $P'(d)$ MPS $P(d)$ then the $\mathbf{g}'$ network is *more heterogeneous* than the $\mathbf{g}$ network, where

more heterogeneous means that there is more variability across sites in $\mathbf{g}'$ than across sites in $\mathbf{g}$ in terms of the number of their connections. In the rest of the paper we will also use the shorthand terminology that a network $\mathbf{g}'$ is a mean-preserving spread of $\mathbf{g}$ to mean that the degree distribution of the network $\mathbf{g}'$ is a mean-preserving spread of the degree distribution of the network $\mathbf{g}$.

Process. Consider a discrete-time process $t \in \{0, 1, 2, ...\}$. Let $N \in \{1, ..., n\}$ be a finite set of agents that travel on the network $\mathbf{g}$.[4] At time $t = 0$ agents are randomly allocated to sites. At each t one agent $i \in N$ is randomly picked and he moves from his current site $a \in S$ to a new site $b \in N_a(\mathbf{g})$. Note that all agents are equally likely to be picked regardless of their current location and the previous history. The agents are random walkers so there is an equal probability that an agent moves to any of the neighboring sites $b \in N_a(\mathbf{g})$. Once agent i arrives at site b a meeting may occur. If there are $m \in (0, n)$ agents at site b then one agent j is randomly picked to meet i. If there are no agents on site b then no meeting occurs.

The state of the system is a vector $\mathbf{x}_{ij}^t \in \mathbf{X} \in \mathbb{R}^{n+2}$ that captures the position of each agent in the network at time t *after* agent i has moved, and the subscripts i and j denote the agents that have been selected to interact at t, with $j = 0$ if i has moved to an empty site. Note that the definition of a state explicitly includes the information on whether an interaction occurred and between which agents.

Learning. At time T, where T is large, each agent receives a signal $\theta_i = \theta + \epsilon_i$ about an unknown state of the world whose true value is θ. Assume that $\epsilon_i \sim N(0, \sigma)$ is an error term independently drawn for each i from a common normal distribution with mean 0. The learning process is as follows. Suppose that at time $t > T$ agent i moves to site b and agent j, who was already at b, is picked to meet with i. During the meeting i and j truthfully reveal their current view on the underlying state of the world, and they update their own view accordingly. The initial signal is never revealed.

More formally, let x_i^t denote agent i's beliefs at time t. At $t = T$ we assume that $x_i^T = \theta_i = \theta + \epsilon_i$. Suppose that at time $t > T$ agent i meets with agent j, then after the meeting agent i's revised beliefs are $x_i^t = x_i^{t-1} + \alpha(x_j^{t-1} - x_i^{t-1})$ where $\alpha \in \left(0, \frac{1}{2}\right)$ is an exogenous parameter which is the same for all agents.[5] The updating rule is the same one used in DeMarzo et al. [2003] and it is very simple: the agent's belief after the meeting is a weighted average of his previous belief and the previous belief of the agent he has met, where the factor α captures the weight given to the other agent's belief.

---

[4]Notice that the set N does not have to include every agent on $\mathbf{g}$. N may simply indicate the (possibly small) subset of agents that receive a signal and are involved in the learning process described below.

[5]The restriction that $\alpha < \frac{1}{2}$ is a sufficient, but not necessary, condition to ensure convergence. Given that our main focus is not the learning rule we chose a more restrictive bound to avoid dealing with convergence issues as α approaches 1.

The model makes the following two assumptions on the interaction process: *(a)* interactions can occur only as a result of movement by one of the agents; *(b)* one and only one interaction is possible per time period. The primary purpose of these two assumptions is to realistically capture the social learning by chit-chat process described in the introduction. Let's imagine a prototypical story of learning by chit-chat: an individual i comes to a grocery store and starts a conversation with j, as they are busy with their one-to-one conversation other individuals may move around and start further one-to-one conversations, until at a certain point the conversation between i and j stops and i (stays idle for a while until he) meets another individual who just moved to the location, etc. Assumption *(a)* captures the reality that learning happens because the social mix at any given location is changing as a result of individuals moving in the environment. Assumption *(b)* captures the fact that learning is a result of personal interaction, not public broadcasting to everyone that happens to be at that location.

The secondary purpose of these assumptions is simplicity and analytical tractability. An extension to *(a)* would be to allow agents to stay put with some probability when they are picked to move. Similarly, one could relax assumption *(b)* to allow interactions between more than two agents. Both these extension would fit the prototypical story and the learning by chit-chat framework. There are different potential formulations of each extension, which would introduce additional complexity to the derivation of the formula in Theorem 1. The two effects driving the comparative statics in Theorem 2 would stay unchanged, but each extension would introduce one additional effect, whose strength would depend on the exact formulation of the extension. Given the novelty of the framework in this paper, we prioritize simplicity and avoid the risk of obfuscating the intuition behind the main results. The exploration of these extensions is therefore left to future work.

3   Motivating examples

The general framework allows for any network to represent the environment that agents navigate. This generality may sometimes obfuscate the intuition behind the results so here we consider two simple stylized networks for illustrative purposes. The two networks are represented in figure 1 for the case when there are only 4 sites. The first network is the star network $\mathbf{g}^*$: all sites are connected to one central site and there is no other link in the network. The second network is the line network $\mathbf{g}^{\bullet-\bullet}$: each site $i \in \{1, s\}$ is connected to the "predecessor" site $i - 1$ and the "successor" site $i + 1$, except for site 1 which is only connected to the successor 2 and site s which is only connected to the predecessor $s - 1$. Note that $\mathbf{g}^*$ and $\mathbf{g}^{\bullet-\bullet}$ have the same number of links and $\mathbf{g}^*$ is a mean-preserving spread of $\mathbf{g}^{\bullet-\bullet}$.

The star network can be a stylized representation of many online environments, which are characterized by the presence of very well-connected websites or hubs. In the example
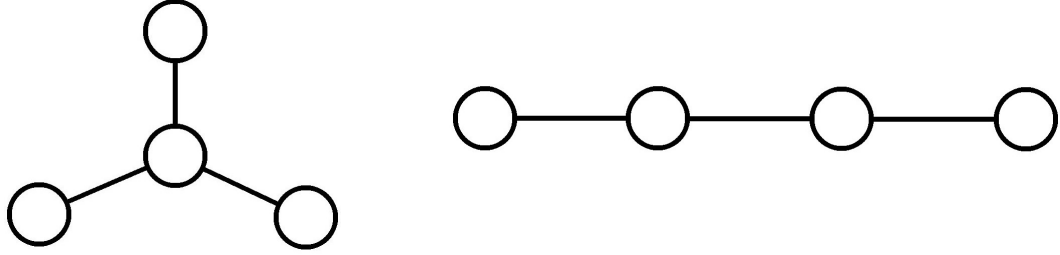
Figure 1: Left: star network $\mathbf{g}^*$ with $s = 4$. Right: line network $\mathbf{g}^{\bullet-\bullet}$ with $s = 4$.

in the introduction the New York Times website would be the center of the star while the peripheral sites could be websites of local newspapers. There are hyperlinks from the NYT's to local newspapers' websites and vice versa, but no hyperlink from a local newspaper to another one. Similarly, the line network can be a stylized representation of an urban area with multiple meeting locations that individuals visit during their day, e.g. the office building, the cafeteria, the gym, the grocery store, the park, etc. The link between two sites indicates that they are directly connected with each other, e.g. the cafeteria and the office building, while the absence of a link between two sites indicates that individuals have to pass through other locations to move between them, e.g. they have to go through the cafeteria to move from the office building to the park.

Denote by $q(\mathbf{g})$ the long-run probability that two agents meet in network $\mathbf{g}$. First, let us consider the case when there are a few agents, and for simplicity let $n = 2$. The following remark compares the probabilities that the two agents meet in $\mathbf{g}^*$ and $\mathbf{g}^{\bullet-\bullet}$.

Remark 1. *Assume that there are only two agents and s sites. Consider environments represented by a line network $\boldsymbol{g}^{\bullet-\bullet}$ and by a star network $\boldsymbol{g}_k^*$. For $s \geq 3$, we have that:*

$$q(\boldsymbol{g}^*) = \frac{1}{4} + \frac{1}{4(s-1)} \geq \frac{1}{(s-1)^2} \left(s - \frac{3}{2} \right) = q(\boldsymbol{g}^{\bullet-\bullet}) \tag{1}$$

Note that $q(\boldsymbol{g}^*) = q(\boldsymbol{g}^{\bullet-\bullet})$ holds only for $s = 3$. Also, note that $\frac{2}{n} > q(\boldsymbol{g}^*)$.

The probability $q(\mathbf{g}^*)$ that the two agents meet in the star network is higher than the probability that they meet in the line network $q(\mathbf{g}^{\bullet-\bullet})$. Note that the expression for $q(\mathbf{g}^*)$ is bounded below by $\frac{1}{4}$. This is because the probability that one agent is at the center of the star is $\frac{1}{2}$ so the probability that both agents are at the center of the star, and therefore a meeting occurs there, is equal to $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. The second term captures the probability that the two agents meet at a peripheral site, which is decreasing in the number of peripheral sites $s - 1$ because the more peripheral sites there are the more unlikely it is that both agents will be at the same site.

The intuition for the statement of Remark 1 is that if there are only two agents then most of the time an agent moves there is no meeting because the destination site is empty.

10

Thus the presence of the central site in the star network makes it more likely that the two agents will meet at this trafficked site. On the other hand, in the line network the two agents may spend a considerable amount of time in different parts of the line where any one-step movement will result in no meeting. Note that the probability of meeting is always lower than $\frac{2}{n}$ because it is possible that no meeting occurs after one agent has moved.

It is helpful to consider some numerical examples to further illustrate the intuition. If $s = 3$ we have that $q(\mathbf{g}^*) = \frac{1}{4} + \frac{1}{16} + \frac{1}{16} = q(\mathbf{g}^{\bullet-\bullet})$, which is what we expect because the line and star networks are the same network if there are only 3 sites. The $\frac{1}{4}$ term is the probability that they meet at the center of the star and $\frac{1}{16}$ is the probability that they meet at each of the peripheral sites. Now consider the $s = 4$ case, which are the networks in figure 1. For the star network we have that $q(\mathbf{g}^*) = \frac{1}{4} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36}$, where the $\frac{1}{4}$ term is the probability that they meet at the center of the star and $\frac{1}{36}$ is the probability that they meet at each of the 3 peripheral sites. For the line network we have that $q(\mathbf{g}^{\bullet-\bullet}) = \frac{1}{36} + \frac{1}{36} + \frac{1}{18} + \frac{1}{18}$, where the first two terms capture the probability that they meet at one of the peripheral sites and the last two terms capture the probability that they meet at one of the non-peripheral sites. Clearly, $q(\mathbf{g}^*) = \frac{1}{3} > \frac{5}{18} = q(\mathbf{g}^{\bullet-\bullet})$. The gap between $q(\mathbf{g}^*)$ and $q(\mathbf{g}^{\bullet-\bullet})$ keeps on widening as s increases, in the $s \to \infty$ limit we have that $\lim_{s\to\infty}[q(\mathbf{g}^*)] = \frac{1}{4} > 0 = \lim_{s\to\infty}[q(\mathbf{g}^{\bullet-\bullet})]$.

Second, let us consider the alternative case when there are many agents, and for illustrative purposes assume that $n \gg s$ in this example.

Remark 2. *Assume that there are n agents and $s \geq 3$ sites with $n \gg s$. Consider two environments represented by a line network $\mathbf{g}^{\bullet-\bullet}$ and a star network $\mathbf{g}^*$. For $s \geq 3$, we have that:*

$$q(\boldsymbol{g}^*) = \frac{1 - \frac{1}{2n} - \frac{1}{2}W^{n-1}}{n} \leq \frac{1 - W^{n-1} + (s-2)\left[1 - \left(W + \frac{1}{s-1}\right)^{n-1}\right]}{n(s-1)} = q(\boldsymbol{g}^{\bullet-\bullet})$$

where $W \equiv \frac{2s-3}{2s-2}$. Note that $q(\boldsymbol{g}^*) = q(\boldsymbol{g}^{\bullet-\bullet})$ holds only for $s = 3$. Moreover, $\frac{2}{n} > q(\boldsymbol{g}^{\bullet-\bullet})$.

The derivation of the analytical expressions for $q(\mathbf{g}^*)$ and $q(\mathbf{g}^{\bullet-\bullet})$ is available in Appendix A. The result is the opposite of what we found for the case with only two agents: the probability that they meet in the line network $q(\mathbf{g}^{\bullet-\bullet})$ is higher than the probability $q(\mathbf{g}^*)$ that they meet in the star network.

The intuition is that if there are many agents, then most of the time the traveling agent moves to a site where there are other agents and therefore a meeting occurs almost always. Thus, the key to maximizing the probability of a meeting is to minimize the probability that there is no meeting when the agent moves to one of the less trafficked sites. The crucial structural feature to minimize the probability that there is no meeting is the absence of poorly connected sites. In the line network, all sites, except for the sites at the end of the line, have two connections so the agents are spread uniformly across the

line and this minimizes the probability that there is no meeting. On the other hand, in the star network all the sites, except for the central one, are poorly connected with only one link and therefore there is a non-negligible probability that an agent may end up on an unoccupied peripheral site. Thus, the probability that there is a meeting is higher in the line network than in the star network. Note that the probability of meeting is always lower than $2/n$ because it is possible that no meeting occurs after one agent has moved.

Recall that the star network is a mean-preserving spread of the line network. Remark 1 states that if there are only 2 agents then a mean-preserving spread of the network increases the probability of a meeting. On the other hand, remark 2 states that if there are many agents then a mean-preserving spread of the network decreases the probability of a meeting. This suggests a conjecture that this statement may extend to *any* network. The next section proves this conjecture.

4 Speed of learning and network structure

The first part of this section shows that in the long-term the system converges to an equilibrium where all agents hold the same belief. The second part contains the main results of the paper that show how the speed of learning depends on the underlying network structure.

4.1 Convergence

The main focus of the paper is an analysis of the speed of learning. However, this type of analysis matters only if one can show that there is some meaningful learning at the societal level. The following lemma shows that this is the case.

Lemma 1. *Assume that θ is the mean of the initial signals received by the agents. There exists a a random variable x^0 with distribution $F(x^0)$ such that:*

(i) *$Pr\left(\lim_{t\to\infty} x_i^t = x^0\right) = 1$ for all $i \in N$*

(ii) *$\mu[F(x^0)] = \theta$*

The first part of the lemma says that in the long-term everyone in the society converges (almost surely) to the same limit belief about the underlying state of the world. The second part states that on average this limit belief will be the mean of the initial signals. Thus, if the initial signals are unbiased, in the sense that their mean is equal to the true value θ of the underlying state of the world, then on average there is meaningful learning at the societal level.

The intuition is that all the agents are interchangeable so over time each one of them will have the same impact in determining the belief everyone converges to, and this leads the society to aggregate the initial signals optimally. This result is independent of the
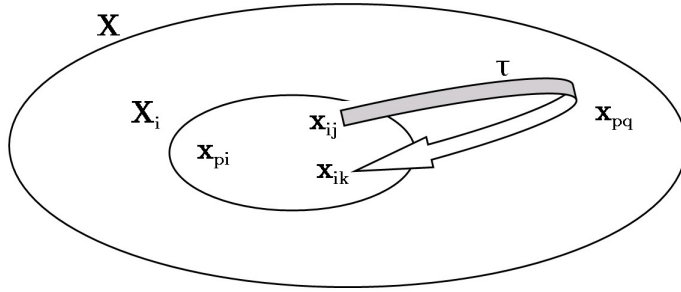
Figure 2: Graphical illustration of the mean recurrence time τ to a subset of states $\mathbf{X}_i$ of the Markov process.

network structure because the underlying network is the same for every agent so it does not introduce any asymmetry across the agents. However, the network matters in determining the speed of convergence.

The quantity that we will compute to analyze the speed of learning is the *mean time* τ an agent has to wait between one meeting and the next. Formally, the quantity that we are interested in is the *mean recurrence time* to a subset of states of the Markov process describing this system. Consider the finite set $\mathbf{X}$ whose elements are all the possible states $\mathbf{x}_{ij}$ of the system. Note that the definition of a state $\mathbf{x}_{ij}$ explicitly specifies the traveling agent i and the agent j, if there is any, that i has met. As illustrated in figure 2 below, we can consider the subset of states $\mathbf{X}_i = \{\mathbf{x}_{ij}, \mathbf{x}_{ji} | j \neq 0\}$ in which agent i met with another agent j in the last time period. Suppose that at time t the system is in state $\mathbf{x}_{ij}$ where i has just moved and met j (or in state $\mathbf{x}_{ji}$ where j has just moved and met i), and suppose that at time $t + 1$ the system is in a state $\mathbf{x}_{pq} \notin \mathbf{X}_i$. The mean recurrence time τ to the subset $\mathbf{X}_i$ is then the mean time it takes the system to return to a state $\mathbf{x}_{ik}$ (or $\mathbf{x}_{ki}$) inside the subset $\mathbf{X}_i$.

The following theorem states the first main result of the paper: an explicit expression that captures how the expected time an agent has to wait between two meetings depends on the underlying environment and the number of other agents. For notational convenience we will drop the expectation operator hereafter, so $\tau(n, s, \mathbf{g})$ denotes the expected time $E[\tau(n, s, \mathbf{g})]$ between two meetings.

Theorem 1. *The expected time $\tau(n, s, \boldsymbol{g})$ agent i has to wait between two meetings is equal to:*

$$\tau(n, s, \boldsymbol{g}) = \frac{nD}{2s \sum_{d=1}^{s-1} d \left[1 - \left(1 - \frac{d}{D} \right)^{n-1} \right] P(d)} \tag{2}$$

where $D = \sum_{k=1}^{s} d_k$.

The proof consists of three parts. First, by Kac's Recurrence Theorem, the mean time $\tau(n, s, \mathbf{g})$ an agent i has to wait between two meetings is equal to the inverse of the probability $q^i(n, s, \mathbf{g})$ that the agent meets another agent. Second, the probability that an agent is at a given site depends only on the degree of the site by a known result on the

theory of random walks on a graph. Third, the bulk of the proof consists in using this result to compute $q^i(n, s, \mathbf{g})$ explicitly.

The dependence of τ on the network structure is fully captured by the degree distribution of the network. This is a consequence of a well-known result in the theory of random walks on a graph, which states that in the stationary state the probability that an agent is at a given site is proportional to the degree of the site. This result holds exactly for all networks that are undirected and unweighted.[6] The requirement that the network is undirected ensures that a detailed balance condition holds. Using this condition it is straightforward to show that in the stationary distribution the probability that an agent is at a node depends only on the degree of that node. The requirement that system should be in the stationary state is satisfied because of the assumption that T is large: the agents have been moving in the environment for a long time before they receive the signals about the underlying state of the world at time T, which is a natural assumption in our framework.

4.2 Comparative statics

We use the result in Theorem 1 to investigate how the speed of learning depends on the network structure $\mathbf{g}$: the shorter is the expected time between meetings, the higher is the frequency of meetings and therefore the faster is the learning process.

The result in Theorem 1 is an exact result, but there is one approximation involved in the interpretation that variations in τ correspond to variations in the speed of learning. The expected time τ between meetings is the first moment of the distribution of the recurrence times, but the variance (and higher moments) of this distribution are also likely to vary depending on the network structure.[7] In interpreting variations in τ as variations in the speed of learning we make the assumption that in the long-term the first moment is the main determinant of the speed of learning, and that the impact of variations in the higher moments are negligible. In other words, we focus on a mean-field approximation to study the learning process.[8] In order to provide a clear separation between the exact results and their interpretation, we formally state the comparative statics results in terms of changes in τ and we discuss their implications for the speed of learning in the discussion of the results.

Recall from the formal definition in section 2 that a mean-preserving spread of the degree distribution of the network captures a specific type of change in network struc-

---

[6]See most textbooks on the subject, e.g. Aldous and Fill [2002]. A concise proof is in Noh and Rieger [2004].

[7]An equivalent result to Kac's Recurrence Theorem for the variance (and higher moments) of the distribution of recurrence times is an unsolved problem in the theory of stochastic processes.

[8]Mean-field approximations are a standard tool used to investigate the evolution of stochastic systems. Examples of the application of this technique in different contexts within the economics of networks literature include Jackson and Rogers [2007a], Jackson and Rogers [2007b] and Lopez-Pintado [2008].

ture. Intuitively, a mean-preserving spread makes the environment more heterogeneous by increasing the variation in connectivity across sites. The following theorem captures how the expected time τ between meetings, and therefore the speed of learning, changes with the level of heterogeneity of the environment.

Theorem 2. *Consider two environments represented by networks $\boldsymbol{g}$ and $\boldsymbol{g'}$ with degree distributions $P(d)$ and $P'(d)$ respectively and such that $P'(d)$ is a mean-preserving spread of $P(d)$, then:*

(i) If $n > \bar{n}$ then the expected time τ between meetings is lower in $\boldsymbol{g}$ than in $\boldsymbol{g'}$

(ii) If $n < \underline{n}$ then the expected time τ between meetings is lower in $\boldsymbol{g'}$ than in $\boldsymbol{g}$

where $\bar{n} \approx \frac{2D}{d_{min}(\boldsymbol{g'})}$ and $\underline{n} \approx \frac{D}{d_{max}(\boldsymbol{g'})}$ and $d_{min}(\boldsymbol{g'})$ and $d_{max}(\boldsymbol{g'})$ are the minimum and maximum degree in $\boldsymbol{g'}$ respectively.

If there is a low number of agents then a heterogeneous environment leads to a shorter expected time between meetings and therefore faster learning. When there are a few agents it is likely that when an agent moves to a new site there is nobody at that site and therefore no meeting occurs. In a heterogeneous environment the presence of highly connected sites increases the probability that the few agents will meet because most of the paths in the network end up at one of these sites.

On the contrary, if the number of agents is high then a homogeneous environment leads to a shorter expected time between meetings and therefore faster learning. Unlike in the case with a few agents, when there are many agents the probability of a meeting is usually high. Thus, the crucial point is to minimize the probability that an agent moves to an unoccupied site. In a heterogeneous environment most of the paths in the network lead away from peripheral sites, which makes it more likely that an agent moving to one of the peripheral sites will find it unoccupied. On the other hand, in a homogenous environment the agents will be distributed more uniformly across the different sites minimizing the probability that the traveling agent moves to an unoccupied site.

The intuition behind this result is similar to the one in Remarks 1 and 2. The star is a prototypical heterogeneous environment, while the line is the prototypical homogeneous environment with the same number of links as the star. Remark 1 is a special case of Theorem 2*(i)*: it shows that if the number of agents is 2 then the probability that i meets j in the star network is higher than in the line network because the star has a central site which increases the probability that the two agents meet. Remark 2 is a special case of Theorem 2*(ii)*: it shows that the opposite is true if the number of agents is large because in the line network the agents are uniformly distributed across the different sites minimizing in this way the probability that a site is unoccupied.

Another type of structural change of the network is a change in the density of connections. Naively one may think that a more densely connected network decreases the

15

expected time between meetings for an agent, but the proposition below shows that this is not the case.

Proposition 1. *Consider the class of environments represented by the regular networks $\boldsymbol{g}^d$ with connectivity d. The expected time τ between meetings is invariant to changes in the connectivity of the network.*

The intuition is that in a regular network all the sites have the same degree so they are all equivalent to each other. This implies that the network plays no role in differentiating across sites, so the expected time between meetings, and therefore the speed of learning, turns out to be independent of the network.

Even if the result in Proposition 1 is restricted to regular networks, it tells us that the level of connectivity of the environment is not a primary determinant of the speed of learning. On the one hand, a more connected environment allows agents to have access to a large number of sites from their current location. But, on the other hand, a more connected environment means that each agent has a large number of sites to move to, so, if two agents are at neighboring sites and one of them is picked to move, it is less likely that he will move to the site occupied by the other agent. These two effects cancel each other out in regular networks making the expected time between meetings independent from the connectivity of the environment.

The result in Theorem 1 also allows us to investigate how the expected time between meetings depends on the number of agents n in the society, as the following corollary shows.

Corollary 1. *Assume that $n \geq 3$, the expected time τ between meetings is*

(i) decreasing in the number of agents in the society if $\frac{d_s}{D} < \frac{1}{n(n-1)}$ for all sites $s \in S$

(ii) increasing in the number of agents in the society if $\frac{d_s}{D} > \frac{1}{n(n-1)}$ for all sites $s \in S$

The dependence on the number of agents is more nuanced because there are two effects at play. If the number of agents increases then there are more agents to meet so the probability of moving to an empty site decreases and therefore learning is faster. On the other hand, if the number of agents increases then any agent is less likely to be picked to move or to be picked to meet the agent that has moved because the pool of agents has increased and only one meeting is possible in each time period.

The first effect dominates if the number of agents is low and if there is no site that is much better connected than all the others. When there are few agents and the environment is such that there is no prominent site on which the few agents can randomly coordinate to meet then the probability of moving to an empty site is significant. Thus, an increase in the number of agents leads to a significant decrease in the probability to have unoccupied sites, and therefore a decrease in the expected time between meetings which leads to faster learning.

On the other hand, the first effect is negligible if there is a large number of agents and if there is no site that is very poorly connected. In a society where there are many agents and with an environment where each site has a non-negligible probability of being visited by agents, the probability of moving to an unoccupied site is negligible. Thus, the main effect of an increase in the number of agents is to decrease the probability that any given agent is picked, which leads to a decrease in the speed of learning.

However, it is worthwhile to notice that the second effect is dependant on our definition of time in the model. An alternative definition would be to normalize a time step by the number of agents so that each agent is picked to move once in each time period. Adopting this alternative definition we would have that the expected time between meetings is always decreasing in the number of agents because only the first effect would survive.

Up to now we have glossed over an important feature of social learning by chit-chat: learning can be unilateral, rather than bilateral. For instance, in the example of social learning by chit-chat using social plugins, only the individual that goes to the webpage updates her beliefs after seeing the recommendation or comment left by another user. The user who left the recommendation or comment is not learning from the interaction. Only a minimal change to the set-up of the model is required to investigate the case of unilateral learning. Specifically, we can investigate two alternative set-ups. In the first one we assume that only the agent who is picked to move in that time period updates her beliefs in case a meeting occurs at the destination site. In the second one we assume that only the agent who is picked to meet the moving agent updates his beliefs. The following corollary illustrates how the results change in these two alternative set-ups.

Corollary 2. *Consider the alternative set-up where only one agent (either always the moving agent or always the agent picked to meet the moving agent) updates beliefs in a meeting. The expected time $\tau_1(n, s, \boldsymbol{g})$ agent i has to wait between two learning opportunities is equal to:*

$$\tau_1(n, s, \boldsymbol{g}) = \frac{nD}{s \sum_{d=1}^{s-1} d \left[1 - \left(1 - \frac{d}{D}\right)^{n-1} \right] P(d)} = 2\tau(n, s, \boldsymbol{g}) \tag{3}$$

where $D = \sum_{k=1}^{s} d_k$. All the comparative statics results in section 4.2 apply to this alternative set-up as well.

The two cases of unilateral learning are equivalent to each other and the comparative statics is unchanged from the case of bilateral learning analyzed so far. The intuition is that the mean time between two "learning opportunities" in either of the unilateral learning settings is twice the mean time between meetings because agents will learn only in half of the meetings. This is because by symmetry each agent is involved in the same number of meetings as the moving agent and as the agent who is picked to meet the moving agent. Learning is slower in the unilateral setting, but all the comparative statics results carry over because the relative strength of all the effects is unchanged.

5 Influencing the society

The role that learning by chit-chat plays in shaping the views that a society converges to means that there are individuals and/or businesses that have an interest in influencing this process. In the learning by chit-chat framework we can model influencers as individuals who never update their belief and who choose to position themselves at a specific site in order to influence the views of others, instead of moving around for reasons that are orthogonal to the learning process. This allows us to investigate which position in the network maximizes the impact of influencers, and which environments are more or less susceptible to influencers' actions.

Specifically, consider the set-up in section 2 and assume that there is one and only one agent i that we will call the *influencer*. Unlike the other agents who travel on the network, the influencer i is positioned at one site s_i and never moves away from s_i. If the influencer i is picked to move at time t then she remains on s_i and she meets one randomly picked agent k (if there is any) who is at site s_i at time t. If another agent j moves at site s_i at time t then the influencer i is included in the pool of agents among which one agent is picked to meet the traveling agent j. Moreover, we assume that $x_i^t \equiv x_i$ for any t, i.e. the influencer's belief is fixed and never changes over time.

Given that the influencer's belief x_i never changes, in the long-term the belief of all the agents in the society will converge to x_i. However, in most contexts the influencer does not just care about convergence per se, but how *fast* the convergence is. Thus, the effectiveness of the influencer in the social learning by chit-chat context is tied to the speed of convergence to x_i. The following definition captures this concept of effectiveness.

Definition 2. The ratio $r_i(n, s, \mathbf{g}) = \frac{\tau(n, s, \mathbf{g})}{\tau_i(n, s, \mathbf{g})}$ is the *effectiveness* of the influencer i. An *influencer i* in an environment represented by $\mathbf{g}$ is *effective* if $r_i(n, s, \mathbf{g}) > 1$.

The $r_i(n, s, \mathbf{g})$ metric is the ratio of the mean time between meetings for a non-influencer to the mean time between meetings involving the influencer. Clearly, the higher this ratio is, the more effective the influencer is, because the frequency of meetings involving the influencer increases compared to the frequency of meetings not involving the influencer. We assume that the presence of the influencer does not affect the value of $\tau(n, s, \mathbf{g})$, which is a reasonable approximation given that there is only one influencer

A natural question to investigate is to determine which position in the network maximizes the influencer's effectiveness. The following proposition shows that the influencer achieves the maximum effectiveness by placing herself at the most connected site.

Proposition 2. *Consider any environment represented by a network $\boldsymbol{g}$. The position that maximizes the effectiveness of the influencer is at the site with the largest degree, and the effectiveness of the influencer at this site is $r_i(n, s, \boldsymbol{g}) \geq 1$.*

The intuition is rather straightforward. The site with the highest degree is the most trafficked one so the influencer based on this site will have the highest frequency of meet-

ings and therefore she will be able to exert the strongest influence on the other agents. Moreover, the influencer at the most connected site will always be (weakly) effective. Hereafter we will assume that the influencer optimally chooses to be on the most connected site in order to maximize her effectiveness. The following definitions will be useful to analyze how the effectiveness of an influencer positioned at the most connected site depends on the underlying environment.

Definition 3. Consider two environments represented by $\mathbf{g}$ and $\mathbf{g}'$ and a society with one influencer i positioned at the site with the largest degree. We say that environment:
- $\mathbf{g}$ is *minimally susceptible* to the action of the influencer i if $r_i(n, s, \mathbf{g}) = 1$
- $\mathbf{g}$ is *more susceptible* to the action of influencer i than $\mathbf{g}'$ if $r_i(n, s, \mathbf{g}) > r_i(n, s, \mathbf{g}')$

The first result is that there is a class of environments that are minimally susceptible to the action of the influencer.

Proposition 3. *An environment represented by a network $\boldsymbol{g}$ is minimally susceptible to the influencer if and only if $\boldsymbol{g}$ is a regular network.*

In a regular network all the sites are equivalent and therefore there is no site where the influencer can position herself to increase the frequency of meetings vis-à-vis the other agents. The other important point is that the influencer is effective in all environments, except for the special class of environments represented by regular networks. Thus, the learning by chit-chat process is rather susceptible to the action of an influencer who is not particularly knowledgeable or sophisticated: she behaves in the same way as all the other agents and the additional knowledge required to exert her influence is simply the knowledge of the location of the most connected site.

In general, different environments will have different levels of susceptibility to the influence of the influencer. The following proposition provides a partial ranking of environments according to their level of susceptibility to the influencer.

Proposition 4. *Consider two environments represented by networks $\boldsymbol{g}$ and $\boldsymbol{g}'$ with degree distributions $P(d)$ and $P'(d)$ respectively and such that $P'(d)$ is a mean-preserving spread of $P(d)$. If $n > \overline{n} \equiv 2D/d_{min}(\boldsymbol{g}')$ then $\boldsymbol{g}'$ is more susceptible to the influencer than $\boldsymbol{g}$.*

The proof is mainly a consequence of Theorem 2. Recall from section 4.2 that if the number of agents is above the threshold $\overline{n}$ then a mean preserving spread increases the mean time τ that an agent has to wait between two meetings. Moreover, by definition of a mean preserving spread, if $\mathbf{g}'$ is a MPS of $\mathbf{g}$ then the degree of the most connected site in $\mathbf{g}'$ is (weakly) larger than the degree of the most connected site in $\mathbf{g}$. The mean time τ_i the influencer has to wait between two meetings is inversely related to the degree of the most connected site, and therefore a mean preserving spread decreases the time an influencer has to wait between two meetings. Thus, the effectiveness $r_i \equiv \tau/\tau_i$ of the influencer increases with a mean preserving spread of the underlying network.

Intuitively, there are two effects that explain the increase in the influencer's effectiveness. The first one is the same effect that drives the result in Theorem 2*(ii)*. If there are many agents then the key to shorten the mean time between meetings is to minimize the probability that an agent moves to an unoccupied site. A shift to a more heterogeneous environment increases the number of poorly connected sites and therefore it increases the probability of moving to an unoccupied site. Note that this does not affect the influencer, who is always at the most connected site whose probability of being unoccupied is negligible. If the probability of moving to an unoccupied site increases then the frequency of meetings that do not involve the influencer decreases, which helps to increase the effectiveness of the influencer. The second effect is that a shift to a more heterogeneous environment (weakly) increases the connectivity of the most connected site where the influencer is. This means that the traffic on the site where the influencer is increases and therefore the frequency of meetings involving the influencer increases as well.

If the number of agents is low then it is not feasible to rank the environments according to their susceptibility to the influencer because the first effect above is reversed, and therefore the two effects go into opposite directions. When there are a few agents, the result of Theorem 2 says that a shift to a more heterogenous environment leads to an increase in the frequency of meetings for any traveling agent because of an increase in the probability of meeting on a trafficked site. But the direction of the second effect is unchanged: a more heterogeneous environment increases the connectivity of the most connected site leading to an increase in the frequency of the influencer who is located at that site. Thus, here the first effect decreases the effectiveness of the influencer while the second one increases it, and therefore the overall effect of a mean preserving spread of the network will depend on the details of the two distributions.

When the number of agents is low there are examples of MPS changes in the network in which the effectiveness of the influencer can go up, and examples in which it can go down. For an example of a MPS shift in which the effect of the influencer goes down, consider two networks $\mathbf{g}$ and $\mathbf{g}'$ such that $\mathbf{g}'$ MPS $\mathbf{g}$ and $d_{max}(\mathbf{g}) = d_{max}(\mathbf{g}')$. It is straightforward to show that in this case $\tau_i(n, s, \mathbf{g}) = \tau_i(n, s, \mathbf{g}')$, and by the result of Theorem 2 we have that $\tau(n, s, \mathbf{g}) > \tau(n, s, \mathbf{g}')$. Thus, substituting these inequalities into the definition of effectiveness we have that $r_i(n, s, \mathbf{g}) > r_i(n, s, \mathbf{g}')$.

An interesting example of a company that generates revenue by influencing the social learning by chit-chat process is *iwearyourshirt.com*. In 2008 the Founder, Jason Sadler, decided to become a "human billboard" by selling the space on his t-shirt for a day to other firms: he would go about his daily activities wearing the t-shirt and random encounters would provide marketing for the advertisers.[9] He made more than $80,000 in the first year; in the last fiscal year the company had revenues of about $500,000 and it currently has 4 employees, or "t-shirt wearers." This section suggests that he

---

[9] "Man makes living by selling the shirt on his back," *Reuters*, 05/11/2009. URL: *http://www.webcitation.org/5l7tvi45i.*

would maximize his effectiveness if he positioned himself in the most connected location (Proposition 2) and if the environment was dominated by a single/few very well-connected locations (Proposition 4). The development of the *iwearyourshirt.com* business suggests that over time they moved in this direction as they shifted most of their activity to a specific (online) location.

6   Conclusion

This paper formulated a novel framework to study a type of learning process that we dub social learning by chit-chat. The main features of this learning process are that individuals learn as a by-product of their daily activities, they do not rely exclusively on their social relations to learn, and the speed of learning depends on the number of individuals in the population and on the environment in which they are embedded. Individuals are located on different nodes of a network, which represents an underlying environment. At time 0, they receive a noisy signal about the underlying state of world. In each successive period an individual is randomly picked to move to a new location, if there is at least another individual at the destination then a meeting occurs and individuals update their beliefs after having learnt the belief of the individual they have met.

In the long-term everyone in the society holds the same belief. The distribution of this limit belief is symmetric around a mean which is equal to the average of the initial signals. We derive an exact formula that describes how the expected time between meetings depends on the number of individuals and the underlying environment. We interpret results on the variations in τ as variations in the speed of learning by using a mean-field approximation. If the number of individuals is below (above) a threshold then a mean preserving spread of the network increases (decreases) the speed of learning. Moreover, the speed of learning is the same in all regular networks, irrespective of their degree. An extension analyzes whether an agent, dubbed the influencer, who does not change her belief and stations at one location, is able to influence the learning process and how the effectiveness of the influencer depends on the underlying network structure.

A   Appendix: Proofs

This appendix contains the proofs of all the statements made in the main body.

Proof of Remark 1. There are two ways for agent i to have a meeting with j: either i is picked to move and travels to the site where j is, or j is picked to move and she moves to the site where i is. By switching the labels i and j it is particularly easy to see here that the probabilities of these two events are the same, hence the probability $q(\mathbf{g})$ that i has a meeting is equal to twice the probability $q_1(\mathbf{g})$ that i is picked to move and travels to the site where j is.

For the line network we have that the probability $q_1(\mathbf{g}^{\bullet-\bullet})$ that i is picked to move and travels to the site where j resides is:

$$q_1(\mathbf{g}^{\bullet-\bullet}) = \frac{1}{2} \left\{ 2 \left[\frac{2}{2(s-1)} \frac{1}{2} \frac{1}{2(s-1)} \right] + 2 \left[\frac{2}{2(s-1)} \frac{1}{2} \frac{2}{2(s-1)} + \frac{1}{2(s-1)} 1 \frac{2}{2(s-1)} \right] \right\}$$
$$+ \frac{1}{2} \left\{ (s-4) \left[\frac{2}{2(s-1)} \frac{1}{2} \frac{2}{2(s-1)} + \frac{2}{2(s-1)} \frac{1}{2} \frac{2}{2(s-1)} \right] \right\}$$

where the $1/2$ factor upfront is the probability that i is selected; the three terms in square brackets are the probabilities that i meets j after moving to the node at the end of the line, to the neighbor of the node at the end of the line and to one of the remaining nodes in the middle respectively. Each term in the square bracket is composed by the probability that i is at the starting node, times the probability that i moves to the node where j is, times the probability that j is at the destination node. The number of terms in the square bracket indicates the number of links to reach the destination site where j is, and the factor in front of the square brackets is the number of the type of destination nodes (e.g. there are two nodes at each end of the line). Summing up we obtain that:

$$q(\mathbf{g}^{\bullet-\bullet}) = 2q_1(\mathbf{g}^{\bullet-\bullet}) = \frac{1}{(s-1)^2} \left(s - \frac{3}{2} \right) \tag{4}$$

Similarly, for the star network $\mathbf{g}^*$ we have that:

$$q_1(\mathbf{g}^*) = \frac{1}{2} \left\{ (s-1) \left[\frac{1}{2(s-1)} \cdot 1 \cdot \frac{s-1}{2(s-1)} \right] \right\} + \frac{1}{2} \left\{ (s-1) \left[\frac{s-1}{2(s-1)} \frac{1}{s-1} \frac{1}{2(s-1)} \right] \right\}$$

where the first term is the probability that i meets j after moving to the center site and the second term is the probability that i meets j after moving to a peripheral site. The explanation for the various terms in the brackets is the same as above. Summing up we obtain that:

$$q(\mathbf{g}^*) = 2q_1(\mathbf{g}^*) = \frac{1}{4} + \frac{1}{4(s-1)} \tag{5}$$

From equations (4) and (5) it is easy to see that $q(\mathbf{g}^*) > q(\mathbf{g}^{\bullet-\bullet})$ for $s > 3$. Note that if $s = 3$ then $q(\mathbf{g}^*) = q(\mathbf{g}^{\bullet-\bullet})$, as expected because if $s = 3$ then the star and the line network are the same network. $\qquad\square$

Proof of Remark 2. For the line network we have that the probability $q_1(\mathbf{g}^{\bullet-\bullet})$ that i has a meeting after he was picked to move is:

$$q_1(\mathbf{g}^{\bullet-\bullet}) = \frac{2}{n}\left\{1 \cdot \frac{2}{2(s-1)}\frac{1}{2}\left[1 - \left(1 - \frac{1}{2(s-1)}\right)^{n-1}\right]\right\} +$$

$$+ \frac{s-4}{n}\left\{2 \cdot \frac{2}{2(s-1)}\frac{1}{2}\left[1 - \left(1 - \frac{2}{2(s-1)}\right)^{n-1}\right]\right\} +$$

$$+ \frac{2}{n}\left\{\frac{1}{2(s-1)} \cdot 1 \cdot \left[1 - \left(1 - \frac{2}{2(s-1)}\right)^{n-1}\right] + \frac{2}{2(s-1)}\frac{1}{2}\left[1 - \left(1 - \frac{2}{2(s-1)}\right)^{n-1}\right]\right\}$$

where the three terms are the probability that i has a meeting after moving to a node at the end of the line, in the middle of the line and a node neighboring a node at the end of the line. The factor in front of the curly brackets is the probability that i is selected times the number of nodes of that type in the line network. The term in square bracket is the probability that there is at least another agent at i's destination node, multiplied by the following factors: the number of nodes from which i could move to the destination node, the probability that i is at one of the departure nodes, and the probability that i moves to the destination node from the departure node. Summing and simplifying we obtain:

$$q_1(\mathbf{g}^{\bullet-\bullet}) = \frac{1}{n(s-1)}\left\{\left[1 - \left(1 - \frac{1}{2(s-1)}\right)^{n-1}\right] + (s-2)\left[1 - \left(1 - \frac{1}{s-1}\right)^{n-1}\right]\right\} \quad (6)$$

Similarly, for the star network $\mathbf{g}^*$ we have that:

$$q_1(\mathbf{g}^*) = \frac{s-1}{2n(s-1)}\left[2 - \left(1 - \frac{s-1}{2(s-1)}\right)^{n-1} - \left(1 - \frac{1}{2(s-1)}\right)^{n-1}\right] \quad (7)$$

Note that for $s = 3$ we have that $q_1(\mathbf{g}^{\bullet-\bullet}) = q_1(\mathbf{g}^*)$, which is what we would expect because when $s = 3$ the star network and the line network are the same network.

By a symmetry argument, or by explicit computation as carried out in the proof of Theorem 1 below, we have that $q(\mathbf{g}^*) = 2q_1(\mathbf{g}^*)$ and $q(\mathbf{g}^{\bullet-\bullet}) = 2q_1(\mathbf{g}^{\bullet-\bullet})$. If we let $W \equiv 1 - \frac{1}{2(s-1)} = \frac{2s-3}{2s-2}$ and substitute it in (6) and (7), we obtain the expressions in Remark 2. Finally, a comparison of expressions (6) and (7) shows that if $n \gg s$ then $q(\mathbf{g}^*) > q(\mathbf{g}^{\bullet-\bullet})$ for $s > 3$. $\qquad\square$

Proof of Lemma 1. Define $x_L^t, x_H^t : \Omega \to \mathbb{R}$ to be two random variables that capture the beliefs of the agents with the lowest and highest beliefs at time t, and where Ω is the space of all sample paths of the system. Let $\Delta^t \equiv x_H^t - x_L^t$. Let us prove each statement separately.

23

(i) First, we prove that x_H^t is decreasing over time. Suppose that at time t agent i has the highest belief $x_i^t > x_j^t$ for all $j \neq i$ and that agent i's next meeting is at time $t' > t$. By definition at time t we have that $x_H^t = x_i^t$. At any time step t'' such that $t \leq t'' < t'$ we claim that $x_H^{t''} = x_i^t$. This follows because no agent $j \neq i$ can have beliefs $x_j^{t''} \geq x_i^{t''}$. There are two possible cases. If $x_j^{t''} = x_j^t$ then $x_j^{t''} = x_j^t < x_i^t = x_i^{t''}$. If $x_j^{t''} \neq x_j^t$ then it must be that j met one or more agents $k \neq i$, assume without loss of generality that j met one agent k then $x_j^{t''} = x_j^{t''-1} + \alpha(x_k^{t''-1} - x_j^{t''-1}) < x_i^t = x_i^{t''}$ because $x_k^{t''-1} < x_i^t$ and $x_j^{t''-1} < x_i^t$. Given that at time t'' no agent j has beliefs $x_j^{t''} \geq x_i^{t''}$ and that i has no meeting between t and t'' then it must be that $x_H^{t''} = x_i^t$. Now consider time t' when agent i meets agent j who has beliefs $x_j^{t'-1} < x_i^t = x_i^{t'-1}$, then we have that i's beliefs at time t' are equal to $x_i^{t'} = x_i^{t'-1} + \alpha(x_j^{t'-1} - x_i^{t'-1}) < x_i^{t'-1} = x_i^t$. Thus, x_H^t is decreasing over time. A similar argument shows that x_L^t is increasing over time.

Second, we show that $Pr\left(\lim_{t \to \infty} x_i^t = x^0\right) = 1$ for all $i \in N$. Pick any $\epsilon > 0$ and suppose that at time T' we have that $\Delta^{T'} = K\epsilon$ for $K > 1$. Define $p \equiv \frac{1}{2\alpha}\left(1 - \frac{1}{K}\right)$. With probability 1, there is a time $T'' > T'$ such that the agents with the highest and lowest belief have met at least p times since time T'.[10] It is then straightforward to see that at time T'' we have that:

$$\Delta^{T''} \leq k\epsilon[1 - 2p\alpha + O(\sum_{q>1} \alpha^q)] \approx k\epsilon[1 - 2p\alpha] = \epsilon \tag{8}$$

where $q \in \mathbb{N}$. This proves the first part of the statement.

(ii) Fix the allocation of agents to sites at time $t = T$. Without loss of generality, assume that the average of the n initial signals is equal to θ and allocate one signal to each agent. By part *(i)* above, as $t \to \infty$ we have that every agent's belief converges (almost surely) to x^0. Suppose that $x^0 = \theta + \delta$ for any $\delta \neq 0$. If this is the case it must be that a subset of agents were more influential in the learning process, because if all the agents had the same influence then the limit belief would be the average θ of the initial signals.

Now consider a different allocation of initial signals such that if agent j had received signal θ_j in the allocation considered above, then in the new allocation it receives a signal $-\theta_j$. By the symmetry of the distribution of initial signals, the two allocations have the same probability of occurrence. Given that the underlying network **g** is the same, there is also the same probability that the same meetings occur in the same order and therefore as $t \to \infty$ we have that every agent has the same belief $x^0 = \theta - \delta$. Thus, the distribution $F(x^0)$ of the limit belief x^0 is symmetric around a mean $\mu[F(x^0)] = \theta$. $\qquad \square$

---

[10]Note that if the agent with the highest (lowest) belief meets other agents other than the one with the lowest (highest) belief between T' and T'' then convergence will be even faster because, by the first part of the proof, this will decrease (increase) his belief. Similarly, if the agent with the highest (lowest) belief changes between T' and T'' then convergence will be even faster because, by the first part of the proof, this will decrease (increase) his belief.

Proof of Theorem 1. Let $\mathbf{X}$ be the set of states of the system, and let $\mathbf{x}_{ij}^t(s)$ be the state in which i moves to site s where i meets j at time t. It is clear that the stochastic process describing the evolution of the system is strongly stationary. Consider a completely additive probability measure q on $\mathbf{X}$, i.e. $q(\mathbf{X}) = 1$

Let A be a subset of states of the stochastic process, i.e. $\mathbf{x}_{ij}^t(s) \in A \subset \mathbf{X}$. It follows from strong stationarity that $P\{\mathbf{x}_{ij}^t(s) \in A\} = q\{\mathbf{x}_{ij}^t(s) \in A\}$ is the same for all t. Given that it is independent of t and for notational convenience, let $P\{\mathbf{x}_{ij}^t(s) \in A\} \equiv P(A)$. Similarly, let $P\{\mathbf{x}_{ij}^t(s) \notin A\} \equiv P(\overline{A})$. Furthermore, define the following quantities:

$$P^{(n)}(A) = P\{\mathbf{x}_{ij}^t(s) \in A, \mathbf{x}_{ij}^{t+1}(s) \notin A, ..., \mathbf{x}_{ij}^{t+n-1}(s) \notin A, \mathbf{x}_{ij}^{t+n}(s) \in A\}$$

$$P^{(n)}(\overline{A}) = P\{\mathbf{x}_{ij}^t(s) \notin A, \mathbf{x}_{ij}^{t+1}(s) \notin A, ..., \mathbf{x}_{ij}^{t+n}(s) \notin A\}$$

and

$$Q^{(n)}(A) \equiv P^{(n)}[\mathbf{x}_{ij}^{t+n}(s) \in A | \mathbf{x}_{ij}^t(s) \in A, \mathbf{x}_{ij}^{t+1}(s) \notin A] = \frac{P^{(n)}(A)}{P(A)}$$

Finally, define the mean recurrence time $\tau(A)$ to the subset A to be equal to:

$$\tau(A) = \sum_{n=2}^{\infty} Q^{(n)}(A) = \sum_{n=2}^{\infty} P^{(n)}(\mathbf{x}_{ij}^{t+n}(s) \in A | \mathbf{x}_{ij}^t(s) \in A, \mathbf{x}_{ij}^{t+1}(s) \notin A)$$

Kac's Recurrence Theorem (KRT) proves that if $\lim_{n \to \infty} P^{(n)}(\overline{A}) = 0$ then *(i)* $\sum_{n=2}^{\infty} Q^{(n)}(A) = 1$, and *(ii)* the mean recurrence time to the subset A is equal to:

$$\tau(A) = \frac{1}{P(A)} = \frac{1}{q(A)}$$

See Kac [1957] for a statement and a proof of this result. Informally, this statement says that if the probability of the system being in the subset A is nonzero, then the expected time for the system to return to a state in A is equal to the inverse of the probability of being in the subset of states A. Thus, in order to compute $\tau(n, s, \mathbf{g})$ we need to compute the probability that the system is in a state in which a given agent j has a meeting.

In computing this probability we make use of a result from the theory of random walks on a graph: an independent random walker on an unweighted, undirected network spends a proportion of time d_c/D at node c, i.e. a proportion of time proportional to the degree of the node. The fact that in the stationary state the probability that a random walker is at a node depends on a simple metric like the degree of the node is a result that holds exactly for unweighted and undirected networks.[11] Note that the system is in stationary state because T is large.

---

[11] The requirements that the network is undirected and unweighted ensure that a detailed balance condition holds. Specifically, $d_a p_{ab}(t) = d_b p_{ba}(t)$ where $p_{ab}(t)$ is the probability that an agent moves from a to b in t time steps. See Noh and Rieger [2004] for a concise proof, or most standard textbooks on the subject (e.g. Aldous and Fill [2002]).

There are two ways in which j can meet another agent. The first one is that there is a probability q_1^{jc} that he moves to a site c where there is at least one other agent. The second one is that there is a probability q_2^{jc} that another agent moves to site c where j currently is and that j is selected to meet the newcomer. Let us compute these two probabilities.

The probability that j is selected to move to a new site c is $1/n$ and, using the result from the theory of random walkers on a graph, the probability that c is the site that j ends up at is d_c/D. These two terms multiplied by the probability that there is at least one other agent at c give the following expression for q_1^{jc}:

$$q_1^{jc} = \frac{1}{n}\frac{d_c}{D}\left[1 - \left(1 - \frac{d_c}{D}\right)^{n-1}\right] \tag{9}$$

The expression for q_2^{jc} is similar to the one for q_1^{jc} above. The second and third term of (9) are unchanged because they capture the probability that the selected agent k ends up at site c where there is at least one other agent. However, now we have to multiply these two terms by the probability that the site where j currently resides is the one where k moved to and that j is picked to meet k. Consider the system *after* k has moved and let m be the number of agents on the site c where j is. The probability that j is among those agents is $m - 1/n$ and the probability that j is picked to meet k is $1/m - 1$. This gives the following expression for q_2^{jc}:

$$q_2^{jc} = \frac{1}{m-1}\frac{m-1}{n}\frac{d_c}{D}\left[1 - \left(1 - \frac{d_c}{D}\right)^{n-1}\right] = \frac{1}{n}\frac{d_c}{D}\left[1 - \left(1 - \frac{d_c}{D}\right)^{n-1}\right] \tag{10}$$

Summing up (9) and (10) we obtain the probability q^j that j is involved into a meeting at site c:

$$q^{jc} = \frac{2}{n}\frac{d_c}{D}\left[1 - \left(1 - \frac{d_c}{D}\right)^{n-1}\right] \tag{11}$$

The probability $q^j(d_c)$ that j is involved into a meeting at a site of degree d_c is then equal to:

$$q^j(d_c) = \frac{2}{n}\frac{d_c}{D}\left[1 - \left(1 - \frac{d_c}{D}\right)^{n-1}\right] s \cdot P(d_c) \tag{12}$$

Taking the expectation over the degree distribution of the network $\mathbf{g}$ we obtain the probability q^j that j is involved into a meeting:

$$q^j = \frac{2s}{nD}\sum_{d=1}^{s-1} d\left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right] P(d) \tag{13}$$

26

Finally, applying Kac's Recurrence Theorem we obtain the mean waiting time $\tau(n, s, \mathbf{g})$ that a given agent j has to wait before the next meeting:

$$\tau_j(d) = \frac{1}{q^j} = \frac{nD}{2s \sum_{d=1}^{s-1} d \left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right] P(d)} \tag{14}$$

$\square$

Proof of Theorem 2. Recall that from (24) the probability q^j that the system is in a state where j interacts with another agent is equal to:

$$q^j = \frac{2s}{nD} \sum_{d=1}^{s-1} d \left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right] P(d) = \frac{2s}{nD} \sum_{d=1}^{s-1} f(d) P(d) \tag{15}$$

Differentiating $f(d)$ we have:

$$f'(d) = \frac{\partial f(d)}{\partial d} = 1 + \frac{D}{(d-D)^2}(dn - D)\left(1 - \frac{d}{D}\right)^n \tag{16}$$

$$f''(d) = \frac{\partial^2 f(d)}{\partial d^2} = \frac{D(n-1)}{(d-D)^3}(dn - 2D)\left(1 - \frac{d}{D}\right)^n \tag{17}$$

Now consider each case separately.

(i) Assume that $n > \bar{n} = \frac{2D}{d_{min}}$. Now consider the various components of (16). By definition for any degree d we have that $0 < d < D$ so the term $\left(1 - \frac{d}{D}\right)^n > 0$ and the term $\frac{D}{(d-D)^2} > 0$. If $n > \frac{2D}{d_{min}}$ then for any value of d we have that the term $(dn - D) > 0$ and therefore we can conclude that if $n > \bar{n}$ then $f'(d) > 0$. Now consider the various components of (17). Again, by definition for any degree d we have that $0 < d < D$ so the term $\left(1 - \frac{d}{D}\right)^n > 0$ and the term $\frac{D(n-1)}{(d-D)^3} < 0$. If $n > \frac{2D}{d_{min}}$ then for any value of d we have that the term $(dn - 2D) > 0$ and therefore we can conclude that if $n > \bar{n}$ then $f''(d) < 0$. Thus, if $n > \bar{n}$ then $f'(d) > 0$ and $f''(d) < 0$, i.e. $f(d)$ is strictly increasing and concave in d. This and the fact that $P'(d)$ is a strict mean-preserving spread of $P(d)$ imply that:

$$q^j = \sum_{d=1}^{s-1} f(d) P(d) > \sum_{d=1}^{s-1} f(d) P'(d) = q'^j \tag{18}$$

and the result follows after application of Kac's Theorem.

(ii) Assume that $n < \underline{n} = \frac{D}{d_{max}}$. Similarly to part (i) above, by inspection of (16) and (17) we can see that if $n < \underline{n} = \frac{D}{d_{max}}$ then $f'(d) < 0$ and $f''(d) > 0$, i.e. $f(d)$ is strictly decreasing and convex in d. This and the fact that $P'(d)$ is a strict mean-preserving spread of $P(d)$ imply that:

$$q^j = \sum_{d=1}^{s-1} f(d) P(d) < \sum_{d=1}^{s-1} f(d) P'(d) = q'^j \tag{19}$$

27

and the result follows after application of Kac's Theorem.

□

Proof of Proposition 1. In the special case of a regular network $\mathbf{g}^d$ with degree d we have that $D = d \cdot s$ and $P(d)$ is a point mass at d, so by simple substitution in equation (2) we obtain that:

$$\tau(n, s, \mathbf{g}^d) = \frac{n}{2s \left[1 - \left(1 - \frac{1}{s}\right)^{n-1}\right]} \tag{20}$$

Clearly (20) is independent of d so in regular networks the mean time $\tau(n, s, \mathbf{g}^d)$, and therefore the speed of learning, is invariant to changes in the connectivity d of the network.

□

Proof of Corollary 1. Differentiating (24) with respect to n we have that:

$$\frac{\partial q^j(n, s, \mathbf{g})}{\partial n} = \frac{2s}{nD^2} \sum_{d=1}^{s-1} d \left\{ \left[\left(1 - \frac{d}{D}\right)^{n-1} - 1\right] - n\left(1 - \frac{d}{D}\right)^{n-1} \log\left(1 - \frac{d}{D}\right) \right\} P(d) \tag{21}$$

Note that the first term is always negative and the second term is always positive, so the sign depends on the relative magnitude of these two terms. If we let $x_d \equiv d/D$ then:

$$\operatorname{sgn}\left\{ \frac{\partial q^j(n, s, \mathbf{g})}{\partial n} \right\} = \operatorname{sgn}\left\{ (1 - x_d)^{n-1} - 1 - n(1 - x_d)^{n-1} \log(1 - x_d) \right\} \tag{22}$$

Notice that $x_d \in (0, 0.5]$. Applying Taylor we obtain that:

$$\operatorname{sgn}\left\{ (1 - x_d)^{n-1} - 1 - n(1 - x_d)^{n-1} \log(1 - x_d) \right\} \approx$$
$$\approx \operatorname{sgn}\left\{ (1 - x_d)^{n-1} - 1 + n(1 - x_d)^{n-1} x_d \right\} \approx$$
$$\approx \operatorname{sgn}\left\{ 1 + (1 - n)nx_d \right\}$$

where the first step uses the approximation $\log(1 - x_d) \approx -x_d$ and the second step uses the approximation $(1 - x_d)^{n-1} \approx 1 + (1 - n)x_d$ and simplifies the resulting expression. We now have that:

   (i) If $x_d < \frac{1}{n(n-1)}$ for all $s \in S$ then $\operatorname{sgn}\left\{ 1 + (1 - n)nx_d \right\} = +$ so the probability of a meeting is increasing with the number of agents and therefore, after application of KRT, the expected time between meetings decreases with the number of agents n.

   (ii) If $x_d > \frac{1}{n(n-1)}$ for all $s \in S$ then $\operatorname{sgn}\left\{ 1 + (1 - n)nx_d \right\} = -$ so the probability of a meeting is decreasing with the number of agents and therefore, after application of KRT, the expected time between meetings increases with the number of agents n. □

Proof of Corollary 2. Consider the case when only the moving agent revises his beliefs, i.e. has a learning opportunity. Expression (9) gives the probability q_1^{jc} that he moves to

a site c where there is at least one other agent and therefore the probability that he has a learning opportunity. The probability $q^j(d_c)$ that j has a learning opportunity at a site of degree d_c is then equal to:

$$q^j(d_c) = \frac{1}{n}\frac{d_c}{D}\left[1 - \left(1 - \frac{d_c}{D}\right)^{n-1}\right]s \cdot P(d_c) \tag{23}$$

Taking the expectation over the degree distribution of the network $\mathbf{g}$ we obtain the probability q^j that j has a learning opportunity:

$$q^j = \frac{s}{nD}\sum_{d=1}^{s-1}d\left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right]P(d) \tag{24}$$

Finally, applying Kac's Recurrence Theorem we obtain the mean waiting time $\tau_1(n, s, \mathbf{g})$ that a given agent j has to wait between two learning opportunities:

$$\tau_j(d) = \frac{1}{q^j} = \frac{nD}{s\sum_{d=1}^{s-1}d\left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right]P(d)} = 2\tau(n, s, \mathbf{g}) \tag{25}$$

By replacing $\tau(n, s, \mathbf{g})$ with $\tau_1(n, s, \mathbf{g})$ in the proofs of Theorem 2, Proposition 1 and Corollary 1 it is clear that the comparative statics results are unchanged.

The proof of the case where only the agent who meets the moving agent has a learning opportunity starts from expression (10), which gives the probability q_2^{jc} that he has a learning opportunity. Given that $q_1^{jc} = q_2^{jc}$ the proof is the same and it is therefore omitted. $\qquad\square$

Proof of Proposition 2. Suppose that the influencer i is at site z in the network. Similarly to a non-influencer, there are two ways in which i can meet another agent. The first one is that there is a probability q_1^{iz} that i is selected and she meets another agent if there is at least one other agent on site z. The second one is that there is a probability q_2^{jz} that another agent j moves to site z and i is the agent picked to meet the newcomer.

The probability that i is selected to move to a new site is $1/n$. By definition the influencer stays at site z so the probability of a meeting q_1^{iz} is given by the probability that the influencer is selected times the probability that there is at least one other agent at z:

$$q_1^{iz} = \frac{1}{n}\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right] \tag{26}$$

Now let us compute the probability q_{2j}^{iz} that an agent j moves to site z and the influencer i is picked to meet the newcomer. There is a probability $1/n$ that j gets picked and a probability d_z/D that it ends up at site z. By definition the probability that the influencer

i is at z is equal to 1 and the probability that the influencer i is picked to meet j is equal to $1/m$ where $m = (n-1)d_z/D$ is the number of agents at z. We have that:

$$q_{2j}^{iz} = \frac{1}{n}\frac{d_z}{D}1\frac{1}{m}\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right] = \frac{1}{n}\frac{d_z}{D}1\frac{D}{(n-1)d_z}\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right] \tag{27}$$

Summing up over the $n-1$ agents $j \neq i$ and simplifying (27) we have that:

$$q_2^{iz} = \sum_{j=1}^{n-1} q_{2j}^{iz} = \frac{1}{n}\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right] \tag{28}$$

Summing up (26) and (28) we get that the probability q^{iz} that the influencer i has a meeting if she is located on site z is equal to:

$$q^{iz} = q_1^{iz} + q_2^{iz} = \frac{2}{n}\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right] \tag{29}$$

Applying Kac's Recurrence Theorem we obtain that the mean time $\tau_i(n, s, \mathbf{g}; z)$ the influencer positioned at site z has to wait between two meetings is equal to:

$$\tau_i(n, s, \mathbf{g}; z) = \frac{1}{q^{iz}} = \frac{n}{2\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right]} \tag{30}$$

Using (2) and (30) we obtain that the effectiveness $r_i(n, s, \mathbf{g}; z)$ of influencer i at site z is equal to:

$$r_i(n, s, \mathbf{g}; z) = \frac{\tau(n, s, \mathbf{g})}{\tau_i(n, s, \mathbf{g}; z)} = \frac{D\left[1 - \left(1 - \frac{d_z}{D}\right)^{n-1}\right]}{s\sum_{d=1}^{s-1} d\left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right] P(d)} \tag{31}$$

Let $w \in S$ be the site with the maximum degree $d_{max}(\mathbf{g}) = d_w$. It is clear by simple inspection of (31) that $\max_{z \in S} r_i(n, s, \mathbf{g}; z) = r_i(n, s, \mathbf{g}; w)$ and therefore the position that maximizes the influencer's effectiveness is the site w with the largest degree. $\qquad\square$

Proof of Proposition 3. First, assume that $\mathbf{g}$ is a regular network $\mathbf{g}^d$ of degree d. By the definition of a regular network it follows that *(i)* $D = d \cdot s$, *(ii)* $P(d)$ is a point mass at d, and *(iii)* $\max_{z \in S}[d_z(\mathbf{g})] = d$. Replacing *(i)-(iii)* in (31) we have that:

$$r_i(n, s, \mathbf{g}^d; z) = \frac{\tau(n, s, \mathbf{g})}{\tau_i(n, s, \mathbf{g}; z)} = 1 \tag{32}$$

and therefore any regular network $\mathbf{g}^d$ is minimally susceptible to the influencer.

In order to prove the other direction, let us proceed by contradiction. Assume that $\mathbf{g}$ does not belong to the class of regular networks $\mathbf{g}^d$ and that $r_i(n, s, \mathbf{g}; w) = 1$ where w is the site with the maximum degree. By (31) it must be that:

$$D\left[1 - \left(1 - \frac{d_w}{D}\right)^{n-1}\right] = \sum_{d=1}^{s-1} d\left[1 - \left(1 - \frac{d}{D}\right)^{n-1}\right] P(d)$$

and note that by definition of w we have that $d_w \geq d_s$ for all $s \in S$. Thus, the only way for the equality to be satisfied is that $d_s = d_w$ for all $s \in S$, which contradicts the fact that $\mathbf{g}$ does not belong to the class of regular networks. $\qquad\square$

Proof of Proposition 4. Denote by $d_{max}(\mathbf{g})$ the degree of the site with the largest degree in $\mathbf{g}$. By the definition of MPS we have that if $\mathbf{g}'$ MPS $\mathbf{g}$ then $d_{max}(\mathbf{g}') \geq d_{max}(\mathbf{g})$ and $D(\mathbf{g}) = D(\mathbf{g}')$. Thus, from (30) we have that:

$$\tau_i(n, s, \mathbf{g}) = \frac{n}{2\left[1 - \left(1 - \frac{d_{max}(\mathbf{g})}{D(\mathbf{g})}\right)^{n-1}\right]} \geq \frac{n}{2\left[1 - \left(1 - \frac{d_{max}(\mathbf{g}')}{D(\mathbf{g}')}\right)^{n-1}\right]} = \tau_i(n, s, \mathbf{g}') \quad (33)$$

By the result in Theorem 2 we have that $\tau(n, s, \mathbf{g}) < \tau(n, s, \mathbf{g}')$. Thus, substituting this inequality and (33) into the definition of $r_i(n, s, \mathbf{g})$ we have that:

$$r_i(n, s, \mathbf{g}) = \frac{\tau(n, s, \mathbf{g})}{\tau_i(n, s, \mathbf{g})} < \frac{\tau(n, s, \mathbf{g}')}{\tau_i(n, s, \mathbf{g}')} = r_i(n, s, \mathbf{g}') \quad (34)$$

which is the desired result. $\qquad\square$

References

D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. *Review of Economic Studies* forthcoming, 2012.

D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2002. URL http://www.stat.berkeley.edu/users/aldous/RWG/book.html.

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

T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Commununications and Mobile Computing*, 2(7):483502, 2002.

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

P. M. DeMarzo, D. Vayanos, and J. Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968, 2003.

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

R. R. Huckfeldt. *Politics in context: Assimilation and conflict in urban neighborhoods*. Algora Publishing, 1986.

M. O. Jackson and B. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, 2007a.

M. O. Jackson and B. Rogers. Relating network structure to diffusion properties through stochastic dominance. *The B.E. Journal of Theoretical Economics (Advances)*, 7(1), 2007b.

M. Kac. On the notion of recurrence in discrete stochastic processes. *Bulletin of the American Mathematical Society*, 53:1002–1010, 1947.

M. Kac. *Lectures in Applied Mathematics: Probability and Related Topics in Physical Sciences*. American Mathematical Society, 1957.

P. Lamberson. Social learning in social networks. *The B.E. Journal of Theoretical Economics*, 10: Iss. 1 (Topics), Article 36, 2010.

D. Lopez-Pintado. Diffusion in complex social networks. *Games and Economic Behavior*, 62:573–590, 2008.

J. D. Noh and H. Rieger. Random walks on complex networks. *Physical Review Letters*, 118701, 2004.

M. M. Zonoozi and P. Dassanayake. User mobility modeling and characterization of mobility pattern. *IEEE Journal on Selected Areas in Communications*, 15(7):12391252, 1997.