Knowing Less is More: Observational Learning in Random Networks

Julian Lorenz *, Martin Marciniszyn and Angelika Steger

Institute of Theoretical Computer Science, ETH Zurich, 8092 Zurich, Switzerland

Abstract

In the standard model of observational learning, n agents sequentially decide between two alternatives a or b, one of which is objectively superior. Their choice is based on a stochastic private signal and the decisions of others. Assuming a rational behavior, it is known that informational cascades arise, which cause an overwhelming fraction of the population to make the same choice, either correct or false. Assuming that each agent is able to observe the actions of *all predecessors*, it was shown by Bikhchandani, Hirshleifer, and Welch (1992, 1998) that, independently of the population size, false informational cascades are quite likely.

In a more realistic setting, agents observe just a *subset of their predecessors*, modeled by a random network of acquaintanceships. We show that the probability of false informational cascades depends on the edge probability p of the underlying network. As in the standard model, if p is a constant, the emergence of false cascades is quite likely. This extends the result of Bikhchandani et al. (1992, 1998) as their model corresponds to $p = 1$. In contrast to that, false cascades are very unlikely if $p = p(n)$ is a sequence that decreases with the size n of the population. Provided the decay of p is not too fast, correct cascades emerge almost surely, benefiting the entire population.

Key words: Networks, Social Learning, Informational Cascade, Random Graphs.
JEL classification: D82, D83

An extended abstract of this paper appeared in the Proceedings of the 20th Annual Conference on Learning Theory (Lorenz, Marciniszyn, and Steger, 2007).

* Corresponding author.
Email addresses: `jlorenz@inf.ethz.ch` (Julian Lorenz), `mmarcini@inf.ethz.ch` (Martin Marciniszyn), `steger@inf.ethz.ch` (Angelika Steger).

1. Introduction

In recent years, there has been growing interest in modeling and analyzing processes of observational learning, first introduced by Banerjee (1992) and Bikhchandani et al. (1992, 1998). In the model of Bikhchandani et al. (1992, 1998), individuals make a once-in-a-lifetime choice between two alternatives sequentially. Each individual has access to private information, which is hidden to other individuals, and also observes the choices made by his predecessors. Since each action taken provides an information externality, individuals may start to imitate their predecessors so as to maximize their objective. Although such *herding behavior* is a locally optimal strategy for each individual, it might not be beneficial for the population as a whole. In the models of Banerjee (1992) and Bikhchandani et al. (1992, 1998), imitation may cause an informational cascade such that all subsequent individuals make the same decision, regardless of their private information. One of the main results in Banerjee (1992) and Bikhchandani et al. (1992, 1998) states that the probability of a cascade that leads most members of the population into the false decision is constant, independently of the population size.

The work of Bikhchandani, Hirshleifer and Welch has been widely recognized, and set of interest in the importance and ubiquity of informational cascades. Besides being investigated in several scientific papers in the following years (see Sect. 1.3 for a literature review), it also attracted attention in the popular press (The Economist, 1994; Investor's Business Daily, 1994; Business Week, 1995; Fortune, 1996).

The result of Bikhchandani et al. (1992, 1998) seems counterintuitive to our every day experience since at many occasions taking the choice of others into account is wise and beneficial for the entire society. In fact, imitation has been recognized as an important manifestation of intelligence and social learning. For instance, in his popular bestseller "The Wisdom of Crowds", Surowiecki (2005) praises the superior judgment of large groups of people over an elite few. This became evident, for example, when Google launched their web search engine, at that time offering a superior service quality. Encouraged by their acquaintances, more and more users adopted Google as their primary index to the web. Moreover, the Google search engine itself leverages the wisdom of crowds by ranking their search results with the PageRank algorithm (Brin and Page, 1998).

The reason that herding could be rather harmful in the model studied by Bikhchandani et al. (1992, 1998) is that each individual has unlimited observational power over the actions taken by *all predecessors*. In a more realistic model, information disseminates not perfectly so that individuals typically observe merely a *small subset of their predecessors*.

In this paper, we propose a generalization of the sequential learning model considered by Bikhchandani et al. (1992, 1998). Suppose the population has size n. For each individual $i \in \{1, \ldots, n\}$, a set of acquaintances $\Gamma(i)$ among all predecessors $j < i$ is selected, where each predecessor of individual i is included in $\Gamma(i)$ with probability $p = p(n)$, $0 \leq p \leq 1$, independently of all other members. Only the actions taken by members of $\Gamma(i)$ are revealed to the individual i, all other actions remain unknown to i. Thus, the underlying social network is a random graph according to the model of Erdős and Rényi (1959). Setting $p = 1$ resembles the model of Bikhchandani et al. (1992, 1998).

Extending the result of Bikhchandani et al. (1992, 1998), we show that if p is a constant, the probability that a false informational cascade occurs during the decision process is constant, i.e., independent of the population size n. On the other hand, if $p = p(n)$ is a function that decays with n arbitrarily slowly, the probability of a false informational cascade tends to 0 as n tends to infinity. Informally speaking, almost all members of fairly large, moderately linked social networks make the correct choice with probability very close to 1, which is in accordance with our every day experience.

1.1. *Model of Sequential Observational Learning in Networks*

We consider the following framework of sequential learning in social networks that naturally generalizes the setting of Bikhchandani et al. (1992, 1998). There are n individuals (or equivalently, *agents* or *decision-makers* in the following), $V = \{v_1, \ldots, v_n\}$, facing a once-in-a-lifetime decision between two alternatives a and b. Decisions are made sequentially in the order of the labeling of V. One of the two choices is objectively superior, but which one that is remains unknown to all individuals throughout. Let $\theta \in \{a, b\}$ denote that superior choice. The a-priori probabilities of being the superior choice are

$$\mathbb{P}[\theta = a] = \mathbb{P}[\theta = b] = \frac{1}{2} \ .$$

Each agent $v_i \in V$ makes his choice $\mathrm{ch}(v_i) \in \{a, b\}$ based on two sources of information: a private signal $s(v_i) \in \{a, b\}$ and public information. The private signal $s(v_i)$ is only observed by the individual v_i. All private signals are identically and independently distributed, satisfying $\mathbb{P}[s(v_i) = \theta] = \alpha$. That is, α is the probability that a private signal correctly recommends the superior choice. The value of α remains unchanged throughout the entire process and is known to all agents. We assume that $1/2 < \alpha < 1$, excluding the trivial case $\alpha = 1$.

The actions $\{\mathrm{ch}(v_i) \mid 1 \leq i \leq n\}$ are public information, but an individual v_i can only observe the actions of a subset $\Gamma_i \subseteq V_{i-1} = \{v_1, \ldots, v_{i-1}\}$ of acquaintances. For all agents v_i, $2 \leq i \leq n$, each of the possible acquaintances $v_j \in V_{i-1}$ is included with probability $0 \leq p = p(n) \leq 1$ into Γ_i, independently of all other elements in V_{i-1}. Equivalently, the underlying social network can be represented as a labeled, undirected random graph $G = G_{n,p}$ on the vertex set V, where each possible edge is included with probability p, independently of all other edges. Then the set of acquaintances Γ_i of agent v_i that already made a decision is given by $\Gamma_G(v_i) \cap V_{i-1}$, where $\Gamma_G(v_i)$ denotes the neighborhood of v_i in G. It is easily seen that both representations are equivalent (Bollobás, 2001; Janson, Łuczak, and Rucinski, 2000) and yield a random graph in the classical model of Erdős and Rényi (1959). We shall assume throughout this paper that the social network is exogenously determined before all decisions take place and represented in form of a random graph $G = G_{n,p}$.

Various models of social networks were proposed in the literature, for instance by Barabási and Albert (1999). The classical random graph model of Erdős and Rényi is analytically well understood and, despite its idealistic assumptions, powerful enough to explain essential features of sequential social learning well. Moreover, it naturally generalizes the model proposed by Bikhchandani et al. (1992, 1998), which is captured in the case $p = 1$.

1.2. *Main Result*

All agents employ the following deterministic rule for making decisions, which is a slight variation of the decision rule in Bikhchandani et al. (1992, 1998).

Definition 1 (Decision rule) *Suppose individual v_i has received the private signal $s(v_i)$, and, among his acquaintances $\Gamma(i)$, m_a chose option a and m_b chose option b. Then the decision* $\mathrm{ch}(v_i)$ *of agent v_i is given by*

$$\mathrm{ch}(v_i) = \begin{cases} a & \text{if } m_a - m_b \geq 2 \ , \\ b & \text{if } m_b - m_a \geq 2 \ , \\ s(v_i) & \text{otherwise } . \end{cases}$$

In Sect. 2 we show that on a complete graph this strategy is locally optimal for each individual assuming that the actions of acquaintances are given in an aggregated form, that is, agent v_i merely observes how many times either of the options a and b was chosen before (see Lemma 1). We will also discuss how this decision rule relates to the decision rule used by Bikhchandani et al. (1992, 1998).

For any two sequences a_n and b_n, $n \in \mathbb{N}$, we write $a_n \ll b_n$ if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0 \ .$$

Then our result reads as follows.

Theorem 1 *Suppose a social network with n agents $V = \{v_1, \ldots, v_n\}$ is given as a random graph $G = G_{n,p}$ with vertex set V and edge probability $p = p(n)$. Assume that private signals are correct with probability $1/2 < \alpha < 1$ and each agent applies the decision rule in Definition 1. Let $c_{\alpha,p}(n)$ be a random variable counting the number of agents that make the correct choice.*

(i) If $n^{-1} \ll p \ll 1$, for all $\gamma > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left[c_{\alpha,p}(n) \geq (1-\gamma)n\right] = 1 \ . \tag{1}$$

(ii) If $0 \leq p \leq 1$ is a constant, then there exist constant $\varrho = \varrho(\alpha, p) > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}\left[c_{\alpha,p}(n) = 0\right] \geq \varrho \ . \tag{2}$$

In moderately linked social networks as in (i), the entire society benefits from learning. On the other hand, if each individual has very many acquaintances on average as in (ii), incorrect informational cascades that lead the entire population into the false decision are quite likely. Note that if agents ignored the actions of others completely, typically a $(1-\alpha)$-fraction of the population would make the false decision.

In very sparse random networks with $p = c/n$ for some constant $c > 0$, no significant herding will arise since those networks typically contain γn isolated vertices for some constant $\gamma = \gamma(c) > 0$ (Bollobás, 2001; Janson et al., 2000). These agents make their decision independently of all other agents and, hence, we expect that both, the group of agents choosing a as well as the group of agents choosing b, contain a linear fraction of the whole population.

The crucial difference between the model of Bikhchandani et al. (1992, 1998), which assumes that the underlying graph of the social network is complete, and our model is that in the former the probability of a false informational cascade primarily depends on the decision of very few agents at the beginning of the process. For instance, with constant probability the first three agents make the false decision, no matter which decision rule they apply. Since in a complete graph each subsequent agent observes these actions, the entire population will be tricked into the false decision. In contrast to that, information accumulates locally in the beginning if the underlying network is sparse as in (i). During a relatively long phase of the process, individuals make an independent decision because none of their acquaintances has decided yet. Hence, after that phase typically a fraction very close to $\alpha > 1/2$ of these agents made the correct choice. In later phases of the process, agents observe this bias among their acquaintances and, trusting the majority, make the correct decision, thereby increasing the bias even more. In the end, almost all agents are on the correct side.

1.3. *Related Results*

As already mentioned, Bikhchandani et al. (1992, 1998) consider the case when the social network is a complete graph; here informational cascades arise quickly, and it is quite likely that they are false. They consider a decision rule that is slightly different from the one in Definition 1. As we will show in Sect. 2, although both rules are locally optimal, false informational cascades are more likely with the rule in Bikhchandani et al. (1992, 1998).

Models of observational learning processes were investigated in several papers. Banerjee (1992) analyzes a model of sequential decision making that provokes herding behavior; as before, each decision-maker can observe the actions taken by *all* of his predecessors. In the model of Çelen and Kariv (2004), decision-makers can only observe the action of their *immediate* predecessor. Banerjee and Fudenberg (2004) consider the model in which each agent can observe the actions of a sample of his predecessors. This is comparable to our model with an underlying random network $G_{n,p}$. However, their model of making decisions is different; at each point in time, a proportion of the entire population leaves and is replaced by newcomers, who simultaneously make their decision. Similarly to our result, Banerjee and Fudenberg (2004) show that, under certain assumptions, informational cascade are correct in the long run. In the learning process studied by Gale and Kariv (2003), agents make decisions simultaneously rather than in a sequential order, but they may repeatedly revise their choice. Watts (2002) studies random social networks, in which agents can either adopt or not. Starting with no adopters, in each round all agents update their state according to some rule depending on the state of their neighbors. In this model, the emergence of global informational cascades also depends on the density of the underlying random network.

1.4. *Organization of the Paper*

The paper is organized as follows. In Sect. 2, we discuss the agents' local decision rule. In Sect. 3 we present the proof of Theorem 1(i). An outline of this proof is contained in Sect. 3.1, where we also state a series of technical lemmas, which are proved in Sect. 3.2. In Sect. 4 we give the proof of Theorem 1(ii). We conclude with experimental results in Sect. 5.

2. Agents' Local Decision Rule

Before we begin with the proof of the Main Theorem, we shall first show that the decision rule stated in Definition 1 is indeed the locally optimal decision strategy for each agent v_i under the assumption that all agents act Bayes rational.

Lemma 1 *Let the social network be given as the complete graph on n vertices. Suppose that previous actions are observable in an aggregated form, and all individuals behave Bayes rational. Then by acting according to the rule in Definition 1 each agent maximizes the a-posteriori probability of making the correct decision.*

Proof. We consider a Markov chain with state variable

$$\Delta_j = \left|\left\{v \;\middle|\; \mathrm{ch}(v) = \theta, v \in V_j\right\}\right| - \left|\left\{v \;\middle|\; \mathrm{ch}(v) \neq \theta, v \in V_j\right\}\right| \ , \tag{3}$$

the difference after j individuals between correct and incorrect decision-makers, assuming that all individuals follow the decision rule in Definition 1. From the decision rule, for all $j \geq 0$ we have the transition probabilities

$$\mathbb{P}\Big[\Delta_{j+1} = \Delta_j + 1 \;\Big|\; \Delta_j \geq 2\Big] = \mathbb{P}\Big[\Delta_{j+1} = \Delta_j - 1 \;\Big|\; \Delta_j \leq -2\Big] = 1 \ , \tag{4}$$

and

$$\mathbb{P}\Big[\Delta_{j+1} = \Delta_j + 1 \;\Big|\; |\Delta_j| \leq 1\Big] = \alpha \ , \tag{5}$$

$$\mathbb{P}\Big[\Delta_{j+1} = \Delta_j - 1 \;\Big|\; |\Delta_j| \leq 1\Big] = 1 - \alpha \ . \tag{6}$$

For all $j \geq 1$ and $-j \leq i \leq j$, define

$$f_{j,i} = \mathbb{P}\Big[\Delta_j = i\Big] \ .$$

The probabilities $f_{j,i}$ will be useful later to prove the local optimality of the decision rule. From the transition probabilities (4)-(6), we will first compute $f_{j,i}$ explicitly. Since the first two individuals always decide independently, we have

$$f_{1,1} = \alpha,\ f_{1,-1} = 1 - \alpha,\ f_{2,2} = \alpha^2,\ f_{2,0} = 2\alpha(1-\alpha) \text{ and } f_{2,-2} = (1-\alpha)^2 \ . \tag{7}$$

In order to have $\Delta_{2j+2} = 0$, we must have $\Delta_{2j} = 0$ and that the actions of the two individuals in $V_{j+2} \setminus V_j$ are a and b in any order. Thus, we have

$$f_{2j+2,0} = 2\alpha(1-\alpha)\, f_{2j,0} \ ,$$

and because of $f_{2,0} = 2\alpha(1-\alpha)$ we conclude

$$f_{2j,0} = 2^j \alpha^j (1-\alpha)^j \ . \tag{8}$$

By similar inductive reasoning with (7) as the base case and using (4)-(6) for the inductive step, we obtain

$$f_{2j-1,1} = 2^{j-1}\alpha^j(1-\alpha)^{j-1} \ , \qquad f_{2j-1,-1} = 2^{j-1}\alpha^{j-1}(1-\alpha)^j \ ,$$

$$f_{2j,2} = 2^{j-1}\alpha^{j+1}(1-\alpha)^{j-1} \ , \qquad f_{2j,-2} = 2^{j-1}\alpha^{j-1}(1-\alpha)^{j+1} \ .$$

Therefore, we have

$$\frac{f_{2j-1,1}}{f_{2j-1,1} + f_{2j-1,-1}} = \alpha \quad \forall\, j \geq 1 \ , \tag{9}$$

and

$$\frac{f_{2j,2}}{f_{2j,2}+f_{2j,-2}} = \frac{\alpha^2}{\alpha^2+(1-\alpha)^2} \quad \forall\, j \geq 1 \ . \tag{10}$$

Because of (4), we have for $j \geq 2$ and $3 \leq i \leq j+1$

$$\frac{f_{j+1,i}}{f_{j+1,i}+f_{j+1,-i}} = \frac{f_{j,i-1}}{f_{j,i-1}+f_{j,-(i-1)}} \ , \tag{11}$$

and thus, inductively by (10) and (11), for all $j \geq 2$ and $2 \leq i \leq j$ we have

$$\frac{f_{j,i}}{f_{j,i}+f_{j,-i}} = \begin{cases} \dfrac{\alpha^2}{\alpha^2+(1-\alpha)^2} & \text{if } j \equiv i \mod 2 \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{12}$$

We will now prove that the decision rule in Definition 1 yields the locally optimal decision for each individual v_j, $1 \leq j \leq n$. For v_1, having no observations, the optimal decision is to follow his private signal, since $\alpha > 1/2$. Suppose now that the individual v_{j+1} has to make his decision, observing m_a individuals that made the choice a and m_b that made the choice b. By the induction hypothesis, we can assume that all his j predecessors followed the rule in Definition 1. We distinguish the cases j even and j odd.

j *even:* Note that $|m_a - m_b|$ is always even. If $m_a = m_b$, then v_{j+1} cannot learn anything about the correct decision by the observation of his predecessors. Since his signal $s(v_{j+1})$ is correct with $\alpha > 1/2$, his optimal choice is $s(v_{j+1})$ in that case. On the other hand, if $|m_a - m_b| \geq 2$, we have

$$\mathbb{P}\left[\theta = a \,\middle|\, m_a - m_b = 2i\right] = \frac{\mathbb{P}\left[\Delta_j = 2i\right]}{\mathbb{P}\left[\Delta_j = 2i\right] + \mathbb{P}\left[\Delta_j = -2i\right]} = \frac{f_{j,2i}}{f_{j,2i}+f_{j,-2i}} \ ,$$

and also

$$\mathbb{P}\left[\theta = b \,\middle|\, m_b - m_a = 2i\right] = \frac{\mathbb{P}\left[\Delta_j = 2i\right]}{\mathbb{P}\left[\Delta_j = 2i\right] + \mathbb{P}\left[\Delta_j = -2i\right]} = \frac{f_{j,2i}}{f_{j,2i}+f_{j,-2i}} \ .$$

Because of (12), decision rule in Definition 1 gives individual v_{j+1} a probability of making the correct choice of

$$\mathbb{P}\left[\mathrm{ch}(v_{j+1}) = \theta \,\middle|\, |m_a - m_b| \geq 2\right] = \frac{\alpha^2}{\alpha^2+(1-\alpha)^2} > \alpha \quad \text{for all } \frac{1}{2} < \alpha < 1 \ . \tag{13}$$

Since this yields a confidence strictly larger than the confidence α of his private signal, the decision rule in Definition 1 is indeed locally optimal.

j *odd:* $|m_a - m_b|$ is always odd, and $m_a = m_b$ can never occur. If $|m_a - m_b| = 1$, by analogous reasoning as in the case of j even above, we obtain from (9)

$$\mathbb{P}\left[\mathrm{ch}(v_{j+1}) = \theta \,\middle|\, |m_a - m_b| = 1\right] = \alpha \qquad \text{for all } \frac{1}{2} < \alpha < 1 \ .$$

Thus, following his private signal as proposed by Definition 1 is locally optimal, since following the thin majority would give agent v_{j+1} only the same probability of making the correct decision. For $|m_a - m_b| \geq 3$, (12) yields

$$\mathbb{P}\left[\mathrm{ch}(v_{j+1}) = \theta \,\middle|\, |m_a - m_b| \geq 3\right] = \frac{\alpha^2}{\alpha^2+(1-\alpha)^2} > \alpha \qquad \text{for all } \frac{1}{2} < \alpha < 1 \ ,$$

and indeed it is optimal for agent v_{j+1} to follow the majority as proposed by Definition 1 regardless of his own signal. Thus, we obtain that the decision rule in Definition 1 is also locally optimal in the case of j odd, which completes the proof of the inductive step.

□

As mentioned above, Bikhchandani et al. (1992, 1998) use a slightly different version of the individuals' local decision rule than in Definition 1. Instead, they consider the following version:

Definition 2 (Local decision rule in Bikhchandani et al. (1992, 1998)) *Suppose individual v_i has received the private signal $s(v_i)$, and, among his acquaintances $\Gamma(i)$, m_a chose option a and m_b chose option b. Let z be drawn uniformly at random from $\{a,b\}$. Then*

$$\mathrm{ch}(v_i) = \begin{cases} a & \text{if } m_a - m_b \geq 2 \quad \text{or} \quad (m_a - m_b = 1) \wedge (s(v_i) = a) \ , \\ b & \text{if } m_a - m_b \leq -2 \quad \text{or} \quad (m_a - m_b = -1) \wedge (s(v_i) = b) \ , \\ z & \text{if } (m_a - m_b = -1) \wedge (s(v_i) = a) \\ & \text{or } (m_a - m_b = 1) \wedge (s(v_i) = b) \ , \\ s(v_i) & \text{if } m_b - m_a = 0 \ . \end{cases}$$

As the following lemma shows, this decision rule yields inferior global behavior on the complete network $G = K_n$ compared to the decision rule given in Definition 1.

Lemma 2 *Suppose the network of acquaintanceships is $G = K_n$. Then the probabilities of ending up in a correct cascade are given by*

$$f_0 = \frac{\alpha^2}{1 - 2\alpha + 2\alpha^2} \quad \text{and} \quad g_0 = \frac{\alpha(1+\alpha)}{2(1-\alpha+\alpha^2)} \ ,$$

if all individuals employ the decision rule in Definition 1 and Definition 2, respectively. For all $1/2 < \alpha < 1$, we have $f_0 > g_0$.

Proof. We consider a Markov chain where the state variable Δ is the difference between correct and incorrect decision-makers. The decision rules in Definitions 1 and 2 yield transition probabilities as given in Fig. 1.

The probability of entering a correct informational cascade is the probability of being absorbed in state $\Delta \geq 2$ when starting in state $\Delta = 0$. Under the decision rule in Definition 1, for $-1 \leq i \leq 1$ let f_i be the probability of eventually being absorbed in state $\Delta \geq 2$ when starting in state $\Delta = i$. Analogously, let g_i be the corresponding probabilities under the decision rule in Definition 2. We obtain the systems of linear equations

$$f_1 = \alpha + (1-\alpha) f_0 \ , \quad f_0 = \alpha f_1 + (1-\alpha) f_{-1} \ , \quad f_{-1} = \alpha f_0 \ ,$$

and

$$g_1 = (1+\alpha)/2 + (1-\alpha) g_0/2 \ , \quad g_0 = \alpha g_1 + (1-\alpha) g_{-1} \ , \quad g_{-1} = \alpha g_0/2 \ ,$$

which yield

$$f_0 = \frac{\alpha^2}{1 - 2\alpha + 2\alpha^2} \quad \text{and} \quad g_0 = \frac{\alpha(1+\alpha)}{2(1-\alpha+\alpha^2)} \ ,$$

and it is straightforward to check that $f_0 > g_0$ for all $1/2 < \alpha < 1$.

□

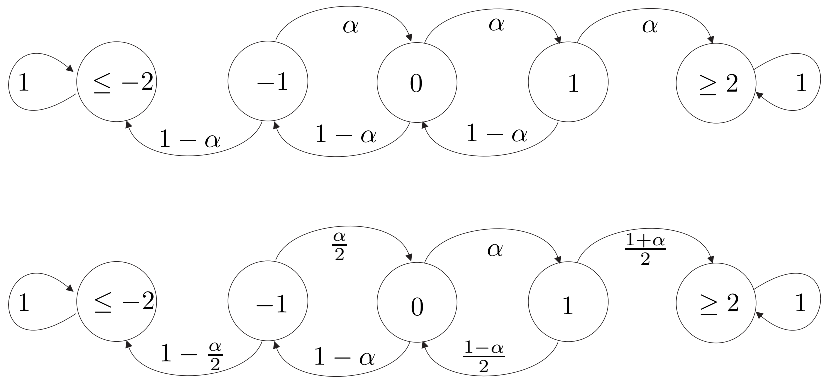

Fig. 1. Markov chains for the difference Δ of correct and incorrect decision-makers under the decision rules in Definition 1 (upper part) and in Definition 2 (lower part).

The alternative decision rule in Definition 2 is a locally optimal strategy for the individuals as well. The only difference between Definition 1 and Definition 2 is the coin flipping if an individual v_i observes a thin majority $|m_a - m_b| = 1$ in conjunction with $s(v_i)$ contradicting this majority vote. From (9) we see that such a thin majority vote has the same significance α to be correct, exactly the same as v_i's private signal. If the majority vote and $s(v_i)$ do not coincide, both choices a and b have therefore the same a-posteriori probability to be correct. Thus, flipping a coin is a locally optimal decision, as well as following the private signal as suggested by the rule in Definition 1. The reason that Definition 1 yields better global behavior on $G = K_n$ (as shown in Lemma 2) is that if individuals follow their private signals when $|m_a - m_b| = 1$ instead of flipping a coin, a greater information externality is provided, benefiting subsequent agents.

3. Proof of Theorem 1(i)

We shall now proceed with the proof of the main Theorem. Suppose $n^{-1} \ll p \leq 1$ is given as in Theorem 1, and consider a random graph $G = G_{n,p}$ on the vertex set V with edge set E. For any set $V' \subseteq V$, let $E(V')$ denote the set of edges induced by V' in G.

For any j, $1 \leq j \leq n$, let $V_j = \{v_1, \ldots, v_j\}$ denote the set of the first j agents. Recall that $\theta \in \{a, b\}$ denotes the objectively superior choice between a and b. For any set of agents $V' \subseteq V$, let

$$C(V') = \{v \in V' : \mathrm{ch}(v) = \theta\}$$

be the subset of agents in V' who made the correct decision. We denote the cardinality of $C(V')$ by $c(V')$. Let $\bar{C}(V') = V' \setminus C(V')$, and $\bar{c}(V') = |\bar{C}(V')|$.

The binomial distribution with parameters n and p is denoted by $\mathrm{Bin}(n, p)$. In the following, if we refer to the "$h(x)$-th agent" for some real valued function $h(\cdot)$, we shall always assume an implicit use of floor functions; for our asymptotic considerations the effect of rounding to the next integer is not vital.

3.1. *Outline of the Proof*

The proof of Theorem 1(i) is based on a series of lemmas that we state here. The proofs are deferred to Sect. 3.2. We will distinguish three phases: Phase I comprising agents $V_I = \{v_1, \dots, v_{k_0}\}$, Phase II comprising agents $V_{II} = \{v_{k_0+1}, \dots, v_{k_1}\}$, and finally Phase III comprising agents $V_{III} = \{v_{k_1+1}, \dots, v_n\}$; we will specify $1 \leq k_0 < k_1 \leq n$ as functions of n below.

In *Phase I*, the phase of the *early adopters*, most decision-makers don't observe more than the decision of one neighbor, and will follow their private signal according to the decision rule in Definition 1. Therefore, almost all agents in V_I make their decisions based solely on their private signal, which yields approximately an α-fraction of individuals in V_I who opted for θ. More specifically, we can establish the following lemma.

Lemma 3 *Let $\omega = \omega(n)$ be a sequence satisfying $1 \ll \omega \ll n$. Let $1/2 < \alpha < 1$, $\omega/n < p \leq 1/\omega$ and $k_0 = p^{-3/4}$ be given. Then we have*

$$\mathbb{P}\left[c(V_{k_0}) \geq \left(1 - p^{1/4}\right)\alpha k_0\right] = 1 - o(1) \quad .$$

Note that if $0 < p \leq 1$ is a constant independent of n, this statement does not hold; there is no $k_0 \geq 1$ such that the number of correctly decided agents in V_{k_0} is roughly k_0 with probability $1-o(1)$. That is exactly what makes the situation in part (ii) of Theorem 1 different.

In *Phase II*, more and more agents face decisions of their acquaintances. An important observation is that the subsequent agent v_{j+1} makes the correct choice with probability at least α if v_{j+1} obeys the decision rule in Definition 1 and "almost" an α-fraction of his predecessors are correct.

Lemma 4 *For every $1/2 < \alpha < 1$ there exists an $\varepsilon > 0$ such that for all $i > k \geq 1$ we have*

$$\mathbb{P}\left[\mathrm{ch}(v_i) = \theta \,\middle|\, c(V_k) \geq (1-\varepsilon)\alpha k \ \wedge\ \Gamma_i \subseteq V_k\right] \geq \alpha \ .$$

That is, following the majority and using the private signal only to break ties does not decrease the chances of any agent even if his acquaintances are randomly selected, provided that there is a bias among all predecessors towards the right direction. This enables us to show that, throughout the first stage, a bias of $\bar{\alpha} > 1/2$ remains stable in the group of decided agents, and also holds at the end of the second phase.

Lemma 5 *Let $\omega = \omega(n)$ be a sequence satisfying $1 \ll \omega \ll n$. Let $1/2 < \alpha < 1$, $0 < p \leq 1/\omega$ and $k_0 = p^{-3/4}$ and $k_1 = p^{-1}\omega^{1/8}$ be given. Then we have*

$$\mathbb{P}\left[c(V_{k_1}) \geq (1 - p^{1/10})\alpha k_1 \,\middle|\, c(V_{k_0}) \geq (1 - p^{1/4})\alpha k_0\right] = 1 - o(1) \quad .$$

At the beginning of *Phase III* (i.e. after agent $k_1 = p^{-1}\omega^{1/8}$), on average every agent v_i has $\mathbb{E}\left[|\Gamma_i|\right] \geq pk_1 \gg 1$ neighbors that already decided. With high probability v_i disregards his private signal and follows the majority vote among its acquaintances, thereby making the correct choice. This means that almost all subsequent agents are correct, if the majority before them was correct. Due to many dependencies between agents, we have to consider increasing groups of agents, and prove the following Lemma.

Lemma 6 *Let $\omega = \omega(n)$ be a sequence satisfying $1 \ll \omega \ll n$, and let $\frac{1}{2} < \bar{\alpha} \leq 1$, $\omega/n \leq p \leq 1/\omega$ and $k \geq k_1 = p^{-1}\omega^{1/8}$. Then we have*

$$\mathbb{P}\left[c(V_{2k} \setminus V_k) \geq \left(1 - \omega^{-1/5}\right)k \,\middle|\, c(V_k) \geq \bar{\alpha} k\right] \geq 1 - \mathrm{e}^{-kp} \, .$$

This Lemma will allow us to prove that with high probability almost all agents make the correct choice in Phase III.

Lemma 7 *Let $\omega = \omega(n)$ be a sequence satisfying $1 \ll \omega \ll n$. Let $1/2 < \alpha < 1$, $\omega/n \leq p \leq 1/\omega$ (i.e. $1/n \ll p \ll 1$) and $k_1 = p^{-1}\omega^{1/8}$ be given. Then we have*

$$\mathbb{P}\left[c(V_n) \geq \left(1-\omega^{-1/6}\right)n \;\middle|\; c(V_{k_1}) \geq \left(1-p^{1/10}\right)\alpha k_1\right] = 1 - o(1) \quad .$$

Combining Lemmas 3, 5 and 7, Theorem 1 follows immediately.

Proof of Theorem 1 (i). We consider the following three events

$$\begin{aligned} E_1: &\quad c(V_{k_0}) \geq \left(1-p^{1/4}\right)\alpha k_0 \quad ,\\ E_2: &\quad c(V_{k_1}) \geq \left(1-p^{1/10}\right)\alpha k_1 \quad ,\\ E_3: &\quad c(V_n) \geq \left(1-\omega^{-1/6}\right)n \quad . \end{aligned}$$

As $\mathbb{P}\left[\overline{E}_3\right] \leq \mathbb{P}\left[\overline{E}_3 \mid E_2\right] + \mathbb{P}\left[\overline{E}_2 \mid E_1\right] + \mathbb{P}\left[\overline{E}_1\right]$, we conclude $\mathbb{P}\left[\overline{E}_3\right] = o(1)$ by Lemmas 3, 5 and 7. Since for n sufficiently large, $\omega^{-1/6} < \gamma$ for every given constant $\gamma > 0$, the theorem follows.

□

What makes the crucial difference between parts (i) and (ii) of Theorem 1 is that if p is a constant, for all $j \geq 1$ the condition $c(V_j) \geq \bar{\alpha} j$ in Lemmas 4 and 6 only holds with probability bounded away from 1. Then it is quite likely that agents experience a bias towards the false direction among their acquaintances, and the same herding behavior as before evokes a false informational cascade.

3.2. *Proofs of Auxiliary Lemmas*

Here we present the proofs of the Lemmas that were stated in the previous section. We will frequently make use of the following Chernoff tail bounds. The reader is referred to standard textbooks (Bollobás, 2001; Janson et al., 2000) for proofs.

Lemma 8 *Let $X_1, \ldots, X_n$ be independent Bernoulli trials with $\mathbb{P}[X_i = 1] = p_i$. Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$. Then we have*

(a) $\mathbb{P}[X \geq (1+\delta)\mu] \leq \mathrm{e}^{-\mu\delta^2/3}$ *for all* $0 < \delta \leq 1$,

(b) $\mathbb{P}[X \leq (1-\delta)\mu] \leq \mathrm{e}^{-\mu\delta^2/2}$ *for all* $0 < \delta \leq 1$,

(c) $\mathbb{P}[X \geq t] \leq \mathrm{e}^{-t}$ *for all* $t \geq 7\mu$ *, and*

(d) $\mathbb{P}[X \geq \mu + t] \leq \mathrm{e}^{-t^2/(2\mu+2t/3)}$ *for all* $t \geq 0$.

We first give the proof of Lemma 3, which makes an assertion on the number of correct decision-makers in Phase I.

Proof of Lemma 3. For all $2 \leq i < k_0$, we have

$$\mathbb{P}[|\Gamma_{i+1}| \geq 2] = \sum_{j=2}^{i}\binom{i}{j}p^j(1-p)^j \leq \sum_{j=2}^{i}(ip)^j \leq k_0^2p^2\sum_{j=0}^{\infty}(k_0p)^j \leq \frac{k_0^2p^2}{1-k_0p} \quad . \tag{14}$$

Let $A = \{v_i : |\Gamma_i| \le 1,\ 1 \le i \le k_0\}$, and $B = V_{k_0} \setminus A$ its complement. Note that all individuals in the set A make their decision solely based on their private signals. On the other hand, individuals in B might have observed an imbalance $|\Delta| \ge 2$ in the actions of their neighbors and chosen to follow the majority, disregarding their private signals. But because of (14) and the definition of k_0 we have

$$\mathbb{E}\left[|B|\right] = \sum_{i=1}^{k_0-1} \mathbb{P}\left[|\Gamma_{i+1}| \ge 2\right] \le \frac{k_0^3 p^2}{1-k_0 p} = {k_0}^3 p^2 \cdot (1+o(1)) = p^{-1/4} \cdot (1+o(1)) \ .$$

Let $\mathcal{E}$ denote the event that $|B| < p^{-3/8}$. As $p^{-1} \to \infty$ we can apply Lemma 8 (c) and deduce that

$$\mathbb{P}\left[\ \overline{\mathcal{E}}\ \right] = \mathbb{P}\left[|B| \ge p^{-3/8}\right] \le \mathrm{e}^{-p^{-3/8}} = o(1) \ .$$

Since by the decision rule in Definition 1 all individuals $v_i \in A$ follow their private signals, we have

$$\mathbb{E}\left[c(A) \,|\, \mathcal{E}\right] = \mathbb{E}\left[\alpha|A| \,\middle|\, \mathcal{E}\right] = \mathbb{E}\left[\alpha(k_0 - |B|) \,\middle|\, \mathcal{E}\right] \ge \alpha\big(k_0 - p^{-3/8}\big) = \alpha k_0\big(1 - p^{3/8}\big) \ .$$

For n sufficiently large, we have

$$\frac{1-p^{1/4}}{1-p^{3/8}} \le 1 - \frac{1}{2}p^{1/4} \ ,$$

and hence

$$\big(1-p^{1/4}\big)\,\alpha k_0 = \frac{1-p^{1/4}}{1-p^{3/8}}\,\alpha k_0 (1-p^{3/8}) \le \left(1-\frac{1}{2}p^{1/4}\right)\mathbb{E}\left[c(A)\,|\,\mathcal{E}\right] \ .$$

Therefore, using the Chernoff bound from Lemma 8 (b), we obtain

$$\mathbb{P}\left[c(A) \le \big(1-p^{1/4}\big)\,\alpha k_0 \,\middle|\, \mathcal{E}\right] \le \mathbb{P}\left[c(A) \le \left(1-\frac{1}{2}p^{1/4}\right)\mathbb{E}\left[c(A)\,|\,\mathcal{E}\right] \,\middle|\, \mathcal{E}\right]$$
$$\le \mathrm{e}^{-\mathbb{E}[c(A)\,|\,\mathcal{E}]p^{1/2}/8} \le \mathrm{e}^{-\alpha k_0\left(1-p^{3/8}\right)p^{1/2}/8} = o(1) \ .$$

As $A \subseteq V_{k_0}$, we conclude

$$\mathbb{P}\left[c(V_{k_0}) \ge \big(1-p^{1/4}\big)\alpha k_0\right] \ge \mathbb{P}\left[c(A) \ge \big(1-p^{1/4}\big)\alpha k_0 \,\middle|\, \mathcal{E}\right] \cdot \mathbb{P}\left[\mathcal{E}\right] \ge 1 - o(1) \ .$$

□

We continue with the proof of Lemma 4.

Proof of Lemma 4. Let $c(V_k) = \bar{\alpha}k$ for some constant $\bar{\alpha} > 1/2$. Furthermore, let

$$\Delta \ = \ c(V_k \cap \Gamma_i) \ - \ \bar{c}(V_k \cap \Gamma_i)$$

be the difference in the number of neighbors of agent i in $C(V_k)$ and in $\overline{C}(V_k)$, and let $p_j = \mathbb{P}\left[\Delta = j\right]$ denote the probability that this difference equals exactly j.

Let $\ell_1 = \min\{\bar{\alpha}k, (1-\bar{\alpha})k + j\}$ and $\ell_2 = (1-\bar{\alpha})k$. Note that $\ell_2 \le \ell_1$ since $\bar{\alpha} > 1/2$. Since $\Delta = j$ if and only if $c(V_k \cap \Gamma_i) = s$ and $\bar{c}(V_k \cap \Gamma_i) = s - j$ for some $s \ge j$, for all $j \ge 2$ we have

$$p_j = \sum_{s=j}^{\ell_1} \binom{\bar{\alpha}k}{s}\binom{(1-\bar{\alpha})k}{s-j} p^{2s-j}(1-p)^{k-(2s-j)}$$

and similarly

$$p_{-j} = \sum_{s=j}^{\ell_2} \binom{(1-\bar{\alpha})k}{s} \binom{\bar{\alpha}k}{s-j} p^{2s-j} (1-p)^{k-(2s-j)} \ .$$

For $r \geq s \geq 1$, let $r^{\underline{s}} = r\,(r-1)\ldots(r-s+1)$ be the falling factorial. For all $j \geq 1$ and $j \leq s \leq \ell_2$, we have

$$\begin{aligned}
\binom{\bar{\alpha}k}{s}\binom{(1-\bar{\alpha})k}{s-j} &= \frac{(\bar{\alpha}k)^{\underline{s}}((1-\bar{\alpha})k)^{\underline{s-j}}}{s!(s-j)!} \\
&= \frac{(\bar{\alpha}k)^{\underline{s-j}}((1-\bar{\alpha})k)^{\underline{s}}}{s!(s-j)!} \cdot \prod_{t=0}^{j-1} \frac{\bar{\alpha}k - s + j - t}{(1-\bar{\alpha})k - s + j - t} \\
&\geq \binom{(1-\bar{\alpha})k}{s}\binom{\bar{\alpha}k}{s-j} \cdot \left(\frac{\bar{\alpha}}{1-\bar{\alpha}}\right)^j \ .
\end{aligned}$$

Therefore we have

$$p_j \geq \left(\frac{\bar{\alpha}}{1-\bar{\alpha}}\right)^2 p_{-j} \qquad \forall\, j \geq 2 \ ,$$

and

$$\begin{aligned}
\mathbb{P}\left[\Delta \geq 2\right] &\geq \left(\frac{\bar{\alpha}}{1-\bar{\alpha}}\right)^2 \mathbb{P}\left[\Delta \leq -2\right] \\
&= \left(\frac{\bar{\alpha}}{1-\bar{\alpha}}\right)^2 \Big(1 - \mathbb{P}\left[-1 \leq \Delta \leq 1\right] - \mathbb{P}\left[\Delta \geq 2\right]\Big) \ .
\end{aligned}$$

Thus, solving for $\mathbb{P}\left[\Delta \geq 2\right]$, we have

$$\mathbb{P}\left[\Delta \geq 2\right] \geq \frac{1}{1 + \left(\frac{1-\bar{\alpha}}{\bar{\alpha}}\right)^2} \Big(1 - \mathbb{P}\left[-1 \leq \Delta \leq 1\right]\Big) \ . \tag{15}$$

Since $\alpha > 1/2$ and $f(\alpha) = \frac{1-\alpha}{\alpha}$ is monotonically decreasing with $0 < f(\alpha) < 1$ in $1/2 < \alpha < 1$, there exists $\varepsilon > 0$ such that $(1-\varepsilon)\alpha > 1/2$ and $f(\bar{\alpha})^2 \leq f(\alpha)$ for all $1 \geq \bar{\alpha} \geq (1-\varepsilon)\alpha$. Hence,

$$\frac{1}{1 + \left(\frac{1-\bar{\alpha}}{\bar{\alpha}}\right)^2} \geq \frac{1}{1 + \left(\frac{1-\alpha}{\alpha}\right)} = \alpha \qquad \forall\, 1 \geq \bar{\alpha} \geq (1-\varepsilon)\alpha > 1/2 \ . \tag{16}$$

Because of the decision rule given in Definition 1, using (15) and (16) we conclude

$$\mathbb{P}\left[\mathrm{ch}(v_i) = \theta\right] = \alpha \mathbb{P}\left[-1 \leq \Delta \leq 1\right] + \mathbb{P}\left[\Delta \geq 2\right] \geq \alpha$$

for all $\bar{\alpha} \geq (1-\varepsilon)\alpha$.

□

Using Lemma 4, we now present the proof of Lemma 5, which asserts that roughly an α-fraction of correct decision-makers is maintained throughout Phase II.

Proof of Lemma 5. Let $\mathcal{C}$ be the event that $c(V_{k_0}) \geq \left(1 - p^{1/4}\right)\alpha k_0$ is satisfied. We consider groups W_i of $m = p^{-1/4}$ agents. We have $\ell = (k_1 - k_0)/m \leq k_1/m \leq p^{-3/4}\omega^{1/8}$

groups between individuals k_0 and k_1. Let $\mathcal{E}_i$ be the event that there is at most one individual in W_i that has a neighbor in W_i, i.e. $|E(W_i)| \leq 1$. Let $\mathcal{E} = \mathcal{E}_1 \wedge \cdots \wedge \mathcal{E}_\ell$. Since $m^2 p = p^{1/2} = o(1)$, we have for n sufficiently large

$$\mathbb{P}\left[\,\overline{\mathcal{E}}_i\,\right] \leq \sum_{j=2}^{\binom{m}{2}} \binom{\binom{m}{2}}{j} p^j \leq \sum_{j=2}^{\binom{m}{2}} m^{2j} p^j \leq m^4 p^2 \sum_{j=0}^{\infty} m^{2j} p^j \leq \frac{m^4 p^2}{1 - m^2 p} \leq 2p \ ,$$

and

$$\mathbb{P}\left[\,\overline{\mathcal{E}}\,\right] \leq \sum_{i=1}^{\ell} \mathbb{P}\left[\,\overline{\mathcal{E}}_i\,\right] \leq \ell \cdot 2p \leq 2\, p^{1/4} \omega^{1/8} \leq 2\, \omega^{-1/8} \ . \tag{17}$$

We have

$$\mathbb{P}\left[c(V_{k_1}) < \left(1 - p^{1/10}\right)\alpha k_1 \,\middle|\, \mathcal{C}\right] \leq \mathbb{P}\left[c(V_{k_1}) < \left(1 - p^{1/10}\right)\alpha k_1 \,\middle|\, \mathcal{E} \wedge \mathcal{C}\right] + \mathbb{P}\left[\,\overline{\mathcal{E}}\,\right] \ ,$$

and defining $\mathcal{A}_i$ as the event that $c(W_i) \geq \alpha\left(1 - p^{1/10}\right)m$,

$$\begin{aligned}
&\mathbb{P}\left[c(V_{k_1}) < \left(1 - p^{1/10}\right)\alpha k_1 \,\middle|\, \mathcal{E} \wedge \mathcal{C}\right] \\
&\quad \leq \mathbb{P}\left[c(V_{k_1}) < \left(1 - p^{1/10}\right)\alpha k_1 \,\middle|\, \bigwedge_{i=1}^{\ell} \mathcal{A}_i \wedge \mathcal{E} \wedge \mathcal{C}\right] + \sum_{j=0}^{\ell-1} \mathbb{P}\left[\overline{\mathcal{A}}_j \,\middle|\, \bigwedge_{i=1}^{j-1} \mathcal{A}_i \wedge \mathcal{E} \wedge \mathcal{C}\right] \ .
\end{aligned}$$

Since $\mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_\ell \wedge \mathcal{C}$ implies $c(V_{k_1}) \geq \left(1 - p^{1/10}\right)\alpha k_1$, we conclude

$$\mathbb{P}\left[c(V_{k_1}) < \left(1 - p^{1/10}\right)\alpha k_1 \,\middle|\, \mathcal{C}\right] \leq \sum_{j=0}^{\ell-1} \mathbb{P}\left[\,\overline{\mathcal{A}}_j \,|\, \mathcal{E} \wedge \mathcal{A}_1 \wedge \cdots \wedge A_j \wedge \mathcal{C}\right] + \mathbb{P}\left[\,\overline{\mathcal{E}}\,\right] \ . \tag{18}$$

Let $\bar{\alpha} = \left(1 - p^{1/10}\right)\alpha$. The event $\mathcal{E} \wedge \mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_{j-1} \wedge \mathcal{C}$ means that before the individuals in group W_j have to make a decision, we have

$$c(V_{k_0 + (j-1)m}) \geq \bar{\alpha}(k_0 + (j-1)m) \ ,$$

and there is at most one individual $w_j \in W_j$ with a neighbor in W_j that made his decision before w_j. Let $\widehat{W}_j = W_j \setminus w_j$ and $\hat{m} = m - 1$. Lemma 4 asserts, that there is an $\varepsilon > 0$ and $\bar{k} \geq 1$ (which both depend only on α), such that for all $k \geq \bar{k}$ we have $\mathbb{P}\left[\mathrm{ch}(v) = \theta\right] \geq \alpha$ for all $v \in \widehat{W}_j$, if $p^{1/10} < \varepsilon$. But since $p \leq 1/\omega$ and $k_0 \geq \omega^{3/4}$, for n sufficiently large we certainly have $k_0 \geq \bar{k}$ and $\bar{\alpha} \geq (1 - \varepsilon)\alpha$. Hence, we have

$$\mathbb{E}\left[c(\widehat{W}_j) \,\middle|\, \mathcal{E} \wedge \mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_{j-1} \wedge \mathcal{C}\right] \geq \alpha\, \hat{m} \ .$$

The Chernoff bound in Lemma 8(b) implies

$$\mathbb{P}\left[c(\widehat{W}_j) \leq \left(1 - \frac{1}{2}p^{1/10}\right)\alpha \hat{m} \,\middle|\, \mathcal{E} \wedge \mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_{j-1} \wedge \mathcal{C}\right] \leq \mathrm{e}^{-\alpha \hat{m} p^{1/5}/8} \leq \mathrm{e}^{-\alpha p^{-1/20}/8} \ .$$

and hence for n sufficiently large

$$\begin{aligned}
\mathbb{P}\left[\overline{\mathcal{A}}_j \,\middle|\, \mathcal{E} \wedge \mathcal{C} \wedge \bigwedge_{i=1}^{j-1} \mathcal{A}_i\right] &= \mathbb{P}\left[c(W_j) \leq \left(1 - p^{1/10}\right)\alpha m \,\middle|\, \mathcal{E} \wedge \mathcal{C} \wedge \bigwedge_{i=1}^{j-1} \mathcal{A}_i\right] \\
&\leq \mathbb{P}\left[c(\widehat{W}_j) \leq \left(1 - \frac{1}{2}p^{1/10}\right)\alpha \hat{m} \,\middle|\, \mathcal{E} \wedge \mathcal{C} \wedge \bigwedge_{i=1}^{j-1} \mathcal{A}_i\right] \leq \mathrm{e}^{-\alpha p^{-1/20}/8} \ .
\end{aligned}$$

Since $\ell \leq p^{-3/4}\omega^{1/8} \leq p^{-7/8}$, we have

$$\sum_{j=1}^{\ell} \mathbb{P}\left[\overline{\mathcal{A}}_j \,|\, \mathcal{C} \wedge \mathcal{E} \wedge \bigwedge_{i=1}^{j-1} \mathcal{A}_i\right] \leq \ell \, \mathrm{e}^{-\alpha p^{-1/20}/8} = o(1) \quad , \tag{19}$$

and because of (17), (18) and (19) we can conclude

$$\mathbb{P}\left[c(V_{k_1}) \geq \left(1 - p^{1/10}\right)\alpha k_1 \,\middle|\, \mathcal{C}\right] = 1 - o(1) \quad .$$

□

In the following, we need the following technical Lemma about binomial random variables.

Lemma 9 *Let $p > 0$, and let $x \geq 1$ and $y \geq 1$ with $px \geq 4c/(c-1)$ and $x/y \leq c$ for some $c > 1$. Let $X \sim Bin(x,p)$ and $Y \sim Bin(y,p)$ be independent. Then*

$$\mathbb{P}\left[X \geq Y + 2\right] \geq 1 - 2\exp(-Cpx) \tag{20}$$

with $C = \frac{(c-1)^2}{12c(c+1)^2}$.

Proof. Let $Y' \sim Bin(x/c,p)$ and $Y'' \sim Bin(\beta x/c, p)$ with $\beta < 1$. Then we have $\mathbb{P}\left[X \geq Y'' + 2\right] \geq \mathbb{P}\left[X \geq Y' + 2\right]$. From this observation we see that we may assume that $x/y = c$.

Let $\delta = \frac{c-1}{2(c+1)}$. We have $0 < \delta < 1/2$. Since

$$(1-\delta)px - (1+\delta)py = \frac{px}{c}\left(c - 1 - (c+1)\delta\right) = \frac{px(c-1)}{2c} \geq 2 \quad ,$$

and X and Y are independent random variables, we conclude

$$\begin{aligned}
\mathbb{P}\left[X - Y \geq 2\right] &\geq \mathbb{P}\left[X \geq (1-\delta)px \;\wedge\; Y \leq (1+\delta)py\right] \\
&= \mathbb{P}\left[X \geq (1-\delta)px\right] \cdot \mathbb{P}\left[Y \leq (1+\delta)py\right] \\
&\geq 1 - \mathbb{P}\left[X \leq (1-\delta)px\right] - \mathbb{P}\left[Y \geq (1+\delta)py\right] \quad .
\end{aligned}$$

Using Chernoff bounds (a) and (b) in Lemma 8, we obtain

$$\mathbb{P}\left[X \leq (1-\delta)px\right] \leq \mathrm{e}^{-px\delta^2/2}$$

and

$$\mathbb{P}\left[Y \geq (1+\delta)py\right] \leq \mathrm{e}^{-py\delta^2/3} = \mathrm{e}^{-px\delta^2/(3c)} \quad .$$

Since $c > 1$, we conclude

$$\mathbb{P}\left[X - Y \geq 2\right] \geq 1 - 2\,\mathrm{e}^{-px\delta^2/(3c)} \quad .$$

□

We continue with the proof of Lemma 6.

Proof of Lemma 6. Let $m = p^{-1} \geq \omega$ and $\ell = k/m = kp \geq \omega^{1/8}$. For all $i = 0, \ldots, \ell-1$, let $W_i = \{v_j \mid k + im \leq j \leq k + (i+1)m\}$. That is, we consider ℓ groups until $2k$ agents have decided. Let A_i be the indicator variable defined by

$$A_i = \begin{cases} 1 & \text{if} \quad c(W_i) < \left(1 - \omega^{-1/4}\right) m \quad , \\ 0 & \text{otherwise} \quad . \end{cases}$$

We define $\mathcal{C}_0$ as the event that $c(V_k) \geq \bar{\alpha} k$, and for $i = 1 \ldots \ell$ let $\mathcal{C}_i$ be the event that $\sum_{j=1}^{i} A_j \leq \ell\omega^{-1/4} \wedge \mathcal{C}_0$. We will first show that

$$\mathbb{P}\left[A_i = 1 \;\middle|\; \mathcal{C}_{i-1}\right] \leq \mathrm{e}^{-\omega^{1/3}} \quad . \tag{21}$$

To prove (21), note that $\mathcal{C}_{i-1}$ implies that

$$c(V_{k+(i-1)m}) \geq c(V_k) + \sum_{j=1}^{i-1} c(W_j) \geq \bar{\alpha} k + \big((i-1) - \ell\omega^{-1/4}\big) m \big(1 - \omega^{-1/4}\big) \ .$$

Using $k = m\ell$, we thus obtain

$$c(V_{k+(i-1)m}) \geq \big(\bar{\alpha} - \omega^{-1/4}\big(1 - \omega^{-1/4}\big)\big) k + \big(1 - \omega^{-1/4}\big)(i-1)m \ ,$$

from which deduce that there exists a constant $\tilde{\alpha} > 1/2$ such that for n sufficiently large,

$$c(V_{k+(i-1)m}) \geq \tilde{\alpha}\big(k + (i-1)m\big) \ .$$

That is, before the first individual of group W_i decides, we have at least a fraction $\tilde{\alpha}$ of correct decision-makers. When $v \in W_i$ has to make his decision, we don't know how many of the individuals in W_i preceding v have made a correct or incorrect decision. Let $x = c(V_{k+(i-1)m}) \geq \tilde{\alpha}(k+(i-1)m)$ and $y = \bar{c}(V_{k+(i-1)m}) + |W_i|$. Since $|W_i| = p^{-1} = o(k)$, we have

$$y \leq (1 - \tilde{\alpha} + o(k))(k + (i-1)m) \ .$$

Hence, there exists a constant $c > 1$ sucht that $x/y \geq c$ for n sufficiently large, and we may apply Lemma 9 to obtain

$$\mathbb{P}\left[\mathrm{ch}(v) = \theta \;\middle|\; \mathcal{C}_{i-1}\right] \geq 1 - 2\,\mathrm{e}^{-pkC} \qquad \forall\, v \in W_i \ .$$

for some $C > 0$. Let

$$\mu = \mathbb{E}\left[c(W_i) \;\middle|\; \mathcal{C}_{i-1}\right] \geq \big(1 - 2\,\mathrm{e}^{-pkC}\big)\, m \geq \Big(1 - 2\,\mathrm{e}^{-C\omega^{1/8}}\Big)\, m \ . \tag{22}$$

Then, for n sufficiently large, we have $\mu \geq m/2 \geq \omega/2$. Lemma 8 (b) implies

$$\mathbb{P}\left[c(W_i) \leq \Big(1 - 2\omega^{-1/3}\Big)\, \mu \;\middle|\; \mathcal{C}_{i-1}\right] \leq \mathrm{e}^{-2\omega^{-2/3}\mu} \leq \mathrm{e}^{-\omega^{1/3}} \quad .$$

For n sufficiently large, (22) implies that $\big(1 - 2\omega^{-1/3}\big)\, \mu \geq \big(1 - \omega^{-1/4}\big)\, m$, and we obtain

$$\mathbb{P}\left[A_i = 1 \;\middle|\; \mathcal{C}_{i-1}\right] \leq \mathrm{e}^{-\omega^{1/3}} \quad ,$$

which completes the proof of (21).

Suppose $\sum_{j=1}^{\ell} A_j \geq \ell\omega^{-1/4}$. Consider the first $\ell\omega^{-1/4}$ groups W_i for which $A_i = 1$. For each of them, we clearly have $\sum_{j=1}^{i-1} A_j \leq \ell\omega^{-1/4}$ and hence (21) holds for each of those groups. Therefore, for n sufficiently large

$$\begin{aligned} \mathbb{P}\left[\sum_{j=1}^{\ell} A_j \geq \ell\omega^{-1/4}\right] \quad &\leq \quad \binom{\ell}{\ell\omega^{-1/4}}\, \mathrm{e}^{-\omega^{1/3}\ell\omega^{-1/4}} \\ &\overset{\binom{n}{k} \leq (\mathrm{e}\, n/k)^k}{\leq} \quad \Big(\mathrm{e}\ \omega^{1/4}\Big)^{\ell\omega^{-1/4}} \quad \mathrm{e}^{-\omega^{1/12}\ell} \leq \mathrm{e}^{-\ell} = \mathrm{e}^{-kp} \quad . \end{aligned}$$

Since $\sum_{j=1}^{\ell} A_j < \ell\omega^{-1/4}$ implies for n sufficiently large

$$c(V_{2k} \setminus V_k) \geq \big(1-\omega^{-1/4}\big)\ell \cdot \big(1-\omega^{-1/4}\big)m \geq \big(1-2\omega^{-1/4}\big)k \geq \big(1-\omega^{-1/5}\big)k \ ,$$

we have

$$\mathbb{P}\left[c(V_{2k} \setminus V_k) \geq \big(1-\omega^{-1/5}\big)k\right] \geq 1-\mathrm{e}^{-kp} \quad .$$

This concludes the proof of Lemma 6.

□

Now we are ready to prove Lemma 7.

Proof of Lemma 7. We consider subphases of increasing length. Subphase j consists of agents

$$W_j = \left\{v_s \,\middle|\, k_1 2^{j-1} < s \leq k_1 2^j\right\} \ .$$

We will have at most $\log(n-k_1) \leq \log n$ such subphases.

Let $\mathcal{C}_0$ be the event that the condition $c(V_{k_1}) \geq \big(1-p^{1/10}\big)\alpha k_1$ is satisfied. Furthermore, for $j \geq 1$, define $\mathcal{C}_j$ as the event that $\mathcal{C}_0, \ldots, \mathcal{C}_{j-1}$ hold and $c(W_j) \geq \big(1-\omega^{-1/5}\big)k_1 2^{j-1}$. Inductively, since for n sufficiently large there exists a constant $\bar{\alpha} > 1/2$ such that $\mathcal{C}_j$ implies that $c(V_{k_1 2^j}) \geq \bar{\alpha} k_1 2^j$, we can employ Lemma 6 for each subphase. We obtain

$$\begin{aligned}
\mathbb{P}\left[c(V_n \setminus V_{k_1}) < (1-\omega^{-\frac{1}{5}})\big(n-k_1\big) \,\middle|\, \mathcal{C}_0\right] &\leq \sum_{j=1}^{\log n} \mathbb{P}\left[c(W_j) < \big(1-\omega^{-\frac{1}{5}}\big)k_1 2^{j-1} \,\middle|\, \mathcal{C}_{j-1}\right] \\
&\leq \sum_{j=1}^{\log n} \mathrm{e}^{-pk_1 2^{j-1}} = \mathrm{e}^{-pk_1} \sum_{j=0}^{\log n - 1} \mathrm{e}^{-pk_1(2^j-1)} \\
&\leq \mathrm{e}^{-pk_1} \sum_{j=0}^{\infty} \mathrm{e}^{-pk_1 j} = \frac{\mathrm{e}^{-pk_1}}{1-\mathrm{e}^{-pk_1}} = o(1) \quad .
\end{aligned}$$

Since $k_1 \leq n\omega^{-7/8}$, $c(V_n \setminus V_{k_1}) \geq (n-k_1)\big(1-\omega^{-1/5}\big)$ implies

$$c(V_n) \geq \big(1-k_1/n\big)\big(1-\omega^{-1/5}\big)n \geq \big(1-\omega^{-1/6}\big)n$$

for n sufficiently large. Thus, we conclude

$$\mathbb{P}\left[c(V_n) \geq \big(1-\omega^{-1/6}\big)n \,\middle|\, \mathcal{C}_0\right] = 1-o(1) \quad .$$

□

4. Proof of Theorem 1(ii)

We proceed in two phases. The first phase lasts until agent j_1 for some $1 \leq j_1 \leq n$ to be specified later, and the second phase lasts from agent j_1+1 to agent n. We certainly have

$$\mathbb{P}\left[c(V_n) = 0\right] \geq \mathbb{P}\left[c(V_n \setminus V_{j_1}) = 0 \,\middle|\, c(V_{j_1}) = 0\right] \cdot \mathbb{P}\left[c(V_{j_1}) = 0\right] \quad . \tag{23}$$

The probability that *all agents* in V_{j_1} make the incorrect decision is at least

$$\mathbb{P}\left[c(V_{j_1}) = 0\right] \geq (1-\alpha)^{j_1} \quad . \tag{24}$$

Note that this value is a constant not depending on n if j_1 is a constant not depending on n.

Since the event $c(V_n \setminus V_{j_1}) \geq 1$, conditioned on $c(V_{j_1}) = 0$, can only hold if there exists a $j \geq j_1$ such that $\mathrm{ch}(v_{j+1}) = \theta \wedge c(V_j) = 0$, we have

$$\mathbb{P}\left[c(V_n \setminus V_{j_1}) \geq 1 \,\middle|\, c(V_{j_1}) = 0\right] \leq \sum_{j \geq j_1} \mathbb{P}\left[\mathrm{ch}(v_{j+1}) = \theta \,\middle|\, c(V_j) = 0\right] \; .$$

As $\mathbb{P}\left[\mathrm{ch}(v_{j+1}) = \theta \mid c(V_j) = 0\right]$ is the probability that v_{j+1} has at most one neighbor in V_j *and* has a correct private signal, we deduce

$$\begin{aligned}
\mathbb{P}\left[c(V_n \setminus V_{j_1}) \geq 1 \middle| c(V_{j_1}) = 0\right] &\leq \sum_{j=j_1}^{n} \alpha \cdot \left((1-p)^j + j(1-p)^{j-1}\right) \\
&\leq \sum_{j=j_1}^{n} \alpha \cdot \left((1-p)^{j-1} + j(1-p)^{j-1}\right) \\
&\leq \alpha(1-p)^{j_1-1} \sum_{j=0}^{\infty} \left((1-p)^j + (j+j_1)(1-p)^j\right) \\
&= \alpha(1-p)^{j_1-1} \left(\sum_{j=0}^{\infty} j_1(1-p)^j + \sum_{j=1}^{\infty} j(1-p)^{j-1} \right) \; .
\end{aligned}$$

As $\sum_{j \geq 0} q^j = 1/(1-q)$ and $\sum_{j \geq 1} j q^{j-1} = 1/(1-q)^2$ for any $0 < q < 1$, we have

$$\mathbb{P}\left[c(V_n \setminus V_{j_1}) \geq 1 \,\middle|\, c(V_{j_1}) = 0\right] \leq \alpha(1-p)^{j_1-1} \left(\frac{j_1}{p} + \frac{1}{p^2} \right) \; . \tag{25}$$

Since (25) becomes arbitrarily small for $j_1 \to \infty$, there exists a constant $j_1 = j_1(p, \alpha)$ such that $\mathbb{P}\left[c(V_n \setminus V_{j_1}) = 0 \,\middle|\, c(V_{j_1}) = 0\right] \geq 1/2$. Hence, with (23) and (24) we conclude that $\mathbb{P}\left[c(V_n) = 0\right] \geq (1-\alpha)^{j_1}/2$, completing the proof.

5. Numerical Experiments

The statements in Theorem 1 are asymptotic, asserting the emergence of informational cascades in the limit $n \to \infty$. As our numerical experiments show, these phenomena can be observed even with moderately small populations.

We conducted numerical simulations with varying population size n and edge probability $p = p(n)$. For each value of n and p, we sampled $N = 2000$ instances of random graphs $G = G_{n,p}$ and of private signals $s(v_i)$, $v_i \in V(G)$. The sequential decision process was evaluated on each of those instances following the decision rule in Definition 1. We identified an informational cascade in such an experiment if at least 95% of all agents opted for the same choice. We computed the relative frequency of informational cascades among the N samples for each value of n and p.

We ran the simulation for $\alpha = 0.75$, $n \in \{100 \cdot i : 1 \leq i \leq 20\}$, and three distinct sequences p. The results are plotted in Fig. 2. The solid and the dotted line represent the relative frequencies of correct and false cascades, respectively, for constant $p = 0.5$. In accordance with Theorem 1(ii), both events occur with constant frequency independent of the population size. The dashed and the dash-dotted line represent the relative

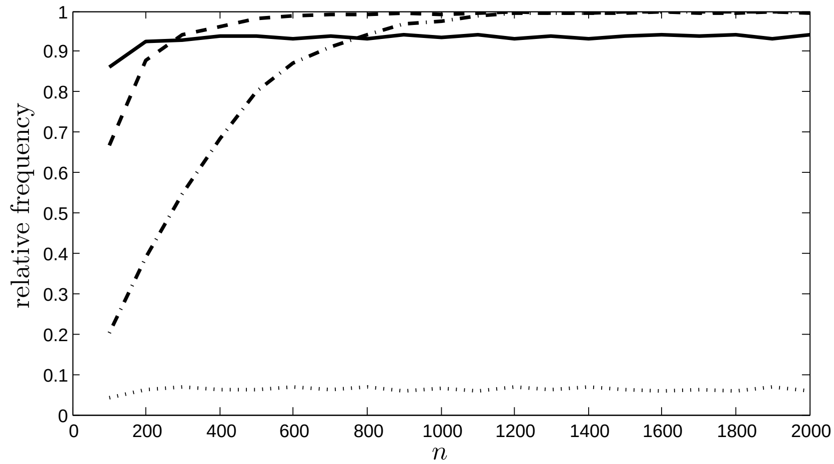

Fig. 2. Simulation results for $\alpha = 0.75$. The plot shows the relative frequencies of correct cascades for different values of the edge probability p as a function of n: $p = 0.5$ (solid line), $p = 1/\log n$ (dashed), and $p = n^{-1/2}$ (dash-dotted). The dotted line is the relative frequency of incorrect cascades for $p = 0.5$.

frequencies of correct cascades for $p = 1/\log n$ and $p = n^{-1/2}$, respectively. Confirming Theorem 1(i) those plots approach 1 as n grows.

Acknowledgment

Julian Lorenz was partially supported by UBS AG. We would like to thank Konstantinos Panagiotou and Florian Jug for helpful discussions.

References

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

Banerjee, A., Fudenberg, D., 2004. Word-of-mouth learning. Games Econom. Behav. 46 (1), 1–22.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.

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

Bikhchandani, S., Hirshleifer, D., Welch, I., Summer 1998. Learning from the behavior of others: Conformity, fads, and informational cascades. The Journal of Economic Perspectives 12 (3), 151–170.

Bollobás, B., 2001. Random graphs, 2nd Edition. Vol. 73 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30 (1–7), 107–117.

Business Week, 1995. Why investors stampede. In: Business Week, Feb 13. Business Week, p. 84.

Çelen, B., Kariv, S., 2004. Observational learning under imperfect information. Games Econom. Behav. 47 (1), 72–86.

Erdős, P., Rényi, A., 1959. On random graphs. I. Publ. Math. Debrecen 6, 290–297.

Fortune, 1996. The theory of fads. In: Fortune, Oct 14. Fortune, p. 49.

Gale, D., Kariv, S., 2003. Bayesian learning in social networks. Games Econom. Behav. 45 (2), 329–346, special issue in honor of Robert W. Rosenthal.

Investor's Business Daily, 1994. Why do shoppers often behave like lemmings? Investor's Business Daily, p. 4.

Janson, S., Łuczak, T., Rucinski, A., 2000. Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York.

Lorenz, J., Marciniszyn, M., Steger, A., 2007. Observational learning in random networks. In: Bshouty, N. H., Gentile, C. (Eds.), COLT. Vol. 4539 of Lecture Notes in Computer Science. Springer, pp. 574–588.

Surowiecki, J., 2005. The Wisdom of Crowds. Anchor.

The Economist, 1994. Yes, ten million people can be wrong. In: The Economist, Feb 19st. The Economist, p. 81.

Watts, D. J., 2002. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci. USA 99 (9), 5766–5771.