Impact of social connectivity on herding behavior *

Deepanshu Vasal[1] [0000−0003−1089−8080]

[1] Department of Electrical and Computer Engineering, University of Texas, Austin, Texas, 78751, USA
{dvasal} at utexas.edu

Abstract. Information cascades have been studied in the literature where myopic selfish users sequentially appear and make a decision to buy a product based on their private observation about the value of the product and actions of their predecessors. Bikhchandani et. al (1992) and Banerjee (1992) introduced such a model and showed that after a finite time, almost surely, users discard their private information and herd on an action asymptotically. In this paper, we study a generalization of that model where we assume users are connected through a random tree, which locally acts as an approximation for Erdös-Rényi random graph when the degree distribution of each vertex of the tree is Binomial and as the number of nodes grows large. We show that informational cascades on such tree-structured networks may be analyzed by studying the extinction probability of a certain branching process. We use the theory of multi-type Galton-Watson branching process and calculate the probability of the tree network falling into a cascade. More specifically, we find conditions when this probability is strictly smaller than 1, that are in terms of the degree distributions of the vertices in the tree. Our results indicate that groups that are less tightly knit, i.e. have lesser connection probability (and as a result have lesser diversity of thought) tend to herd more than the groups that have more social connections.

Keywords: Information Cascades · Multi-type Galton-Watson process · Social Learning.

1 Introduction

People's decisions in the real world are influenced not only by their own opinions or predilections, but also by the decisions and opinions of their peers. For example, decisions such as buying a product, voting for a candidate, or choosing a restaurant among many others, are clearly a combined product of people's individual preferences and that of their peers. It is commonly observed that people tend to "herd"; i.e., follow the majority opinion. For instance, if there is a product on an online retail website with many positive reviews, it is highly probable that it will be chosen by a new buyer. Similarly, it is more likely that a person will choose to go to a restaurant that is recommended by many of their friends. With the ever increasing presence of internet platforms and social media apps, and the very real consequences of people making decisions based on

* Supported in part by Department of Defense grant W911NF1510225 and Simon's Foundation grant 26-7523-99.

dubious information that spreads via such apps, it is even more important to understand how people's actions are influenced by their peers and these actions further affect the spread of information in the network.

In the seminal papers [3, 4], the authors considered a simple model for decision making which demonstrated a herding behavior among users: users flocked to the same action regardless of their private opinions, and furthermore, with a non-zero probability this action may be wrong despite all the users being perfectly rational. The model considered a binary state of the world which represents value of a product, and infinite sequence of users that appear in some fixed order. Each user (or agent) makes an independent noisy observation of the state of the world. In addition to this observation, agents have access to the entire history of decisions by past agents. The current agent then makes a binary decision to either buy or not buy the product based on their private observation and the historical information about past actions. An agent's decision, once made, is fixed and does not change with time. In this model, the authors of [3, 4] showed that a surprising phenomenon of herding (or cascading) occurs: with probability one, there is a finite time such that all agents that appear after this time choose to ignore their private observations and simply copy the action of the previous agent. This model was later generalized in [20, 1], and similar cascading phenomena were shown to exist in these general models as well.

Agents herding to the wrong action may be highly undesirable for a society, and there has been a considerable amount in work in developing strategies that avoid herding. Peres et al. in [18] consider this question for the model in [3, 4]. They propose a stochastic strategy such that each user independently and randomly discards the information of actions of previous players, and plays an action solely based on its private information. It is shown that this strategy avoids cascades and that an asymptotic rate for the probability of learning by the users as a team may be calculated. Authors in [22] study an ergodic version of this model where users also have the ability to post a review about the product at some cost. They design incentives to be given by social planner that aims to align the goals of strategic players with the team objective, and incentivizes users to reveal their private information through the reviews. Acemoglu et al. in [2] study learning dynamics under two classes of rating systems: full history, where customers see the full history of reviews, and summary statistics, where the platform reports some summary statistics of past reviews. They provide conditions for asymptotic learning under both cases. Harel et al. in [10] study the speed of learning with myopically selfish agents acting repeatedly and show that only a fraction of players private information is transmitted through their actions, where this fraction goes to zero as the number of players go to infinity, demonstrating *groupthink* behavior. Dhounchak et al. in [7] explore the use of multi-type branching processes to analyze viral properties of content in a social network, with and without competition from other sources. Vasal and Anastasopoulos in [21] consider a dynamic model where users with independent Markovian types make noisy observations of their types privately and publicly observe actions of everybody else. The users are not myopic anymore, and participate throughout the game through an endogenous or exogenous process. Based on a sequential decomposition methodology to compute perfect Bayesian equilibria, they characterize set of informational cascades for their dynamic model.

In this paper, we study the problem of social learning and specifically herding behavior when the social network corresponding to the agents is a tree. Herding or cascading on such a graph may be interpreted as spreading of a rumor, or infection on the graph. Infection spread processes have been studied extensively in network science [12, 17], with problems of interest being maximizing (or minimizing) the spread of an infection [13], or identifying (or hiding) sources of infections [19, 8, 5]. In our model, we consider a random rooted tree where each node corresponds to an agent. We call the root of this tree the source (or the ancestor). Agents make private observations of the state of the world and these observations are independent and identically distributed conditioned on the state of the world. Furthermore, an agent can observe the (possibly noisy) history of actions of all its ancestors; i.e., of all the agents that lie on the path joining the source node to the agent's node. Based on the agent's private observation and observation past actions, the agent performs an action, such as to buy or not buy a product. As in the previously described models, herding occurs if an agents performs the same action as their parent node, without considering their private observation. We show that probability of users herding reduces to studying extinction probability in a branching process, which we characterize using Galton-Watson branching process theory. In the setting we consider, users may not herd to the same action: different branches of the tree may herd to different actions. In the infection spread interpretation, this analysis yields which of two competing infections (or products) becomes eventually dominant. It is known that random trees are good local approximation for random graphs, for instance a graph locally behaves like a random tree whose degree distribution is Binomial($n, \frac{p}{n}$) if the probability of connection in an Erdös Rényi graph[1] is p [6].

Our model can be considered as a subset of the random observation model considered in [1]. Authors in [1] consider a sequential model where every user observes a random subset of the actions of the previous users. They provide sufficient conditions of *expansive observations* for probability of asymptotic learning (where users correctly learn the state of the system) to be 1. In our model, we first consider a general observation model which we later specialize to a binary symmetric channel and provide conditions when users fall into a cascade (which may not correspond to the right action) is 1 and when it is strictly smaller than 1. Thus our focus is not on asymptotic learning, rather on information cascades.

The paper is structured as follows. In Section 4, we define equilibrium strategies of the players and other preliminary notation used later in the paper. In Section 3, we use the theory of Multi-type Galton-Watson process to study this general model. In Section 5, we specialize our results to a binary symmetric channel with erroneous actions where players observe actions of their ancestors through a binary symmetric channel. We conclude in Section 6. All proofs are presented in Appendices.

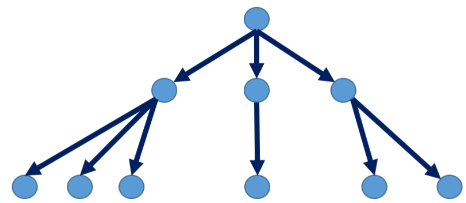

Fig. 1: Galton-Watson Branching process

2 Model

A Galton-Watson tree is defined as a branching process which starts with one node and each node of the tree gives birth to D children, where D is a discrete random variable with known probability generating function (PGF) ϕ_D i.e. $\phi_D(s) = \mathbb{E}[s^D]$. Let there be a product whose value is $V \in \{0,1\}$ which is random such that $P(V=0) = P(V=1) = 1/2$. There exists infinite selfish players who appear sequentially in a predefined order such that each player acts on one of the node of a tree with a random degree whose probability generating function (PGF) is given by ϕ_D, where $P(D=0) = 0$.[2] We denote k^{th} player at stage t of the tree by p_t^k. Player p_t^k privately receives an observation $x_t^k \in \mathcal{X}$ about the value of the product through a general channel $Q(x_t^k|V=v)$.[3] Equivalently, we can assume that players observe $q_t^k := P(V=1|x_t^k)$ such that q_t^k conditioned on value v is distributed according to CDF $F^v, v \in \{0,1\}$, (i.e. $F^v(q) := P(q_t^k \leq q|V=v)$) where $supp(F^v) = [\underline{b}, \bar{b}] \subseteq [0,1]$, and $0 \leq \underline{b} < 1/2 < \bar{b} \leq 1$. We also assume that F^0, F^1 are not identical and are absolutely continuous with respect to each other so that no signal is fully revealing. Each player appears exactly once in the game and shares its actions with all its descendants on the tree. Thus each player observes actions of its ancestors, denoted by $a^{\mathcal{P}(p_t^k)}$, and its own private observation, x_t^k or q_t^k. Based on this information, it takes an action $a_t^k \in \{0,1\}$ to either buy the product ($a_t^k = 1$) or not buy the product ($a_t^k = 0$) and then it leaves the system, implying that it is myopic in nature. It gets a reward $R_t^k(a_t^k, v) = \begin{cases} 1 \text{ if } a_t^k = 1, v = 1, \\ -1 \text{ if } a_t^k = 1, v = 0, \\ 0 \text{ if } a_t^k = 0 \end{cases}$.

3 Preliminaries: Multi-type Galton-Watson process

In this section, we will review the theory of Multi-type Galton Watson population process [16] which will be used in Section 4 to study the probability of falling into cascades. A Galton-Watson tree is defined as a branching process which starts with one

[1] Erdös Rényi random graph with parameter p is defined as an undirected random graph where each pair of vertices is connected independently with probability p.

[2] This assumption excludes the possibility the tree goes extinct due to lack of child and thus implies that probability of extinction is the same as probability of a cascade (as justified later).

[3] We assume that the Q kernels have monotonicity property, as in [4], which implies $\frac{Q(X_t^k=x|V=1)}{Q(X_t^k=x|V=0)}$ is increasing in x. This fact is used in the proof of Lemma 1.

node and each node of the tree gives birth to D children, where D is random with known probability generating function (PGF) ϕ_D. Galton and Watson in [14] used this branching process to study extinction rate of family names. Since then it has been used to study nuclear reactions, population growth, and cosmic rays [11]. In a single type Galton-Watson process, all nodes are of the same type. It is known that the tree goes extinct almost surely if $\mathbb{E}[D] < 1$ and it goes extinct with a probability strictly less than 1 if $\mathbb{E}[D] > 1$, where this probability is given by the smallest solution of the fixed-point equation $y = \phi_D(y)$. A multi-type Galton-Watson process is defined as a branching process where each node of the tree could be of multiple type (finite, countable or uncountable) and gives birth to children of any of the types with certain probability.

Moyal in [16] studied this process for arbitrary type in [0,1]. Suppose every node has a type $x \in [0,1]$ and it gives birth to n children of types $(y_1, \dots, y_n)$ with probability $P_1^{(n)}(dy^n|x)$, where $y_i \in [0,1]$. Here the subscript (1) in $P_1^{(n)}(dy^n|x)$ defines the depth of the tree and superscript (n) defines the number of children. Then for any function $\xi : [0,1] \to [0,1]$, the probability generating functional $G_1 : ([0,1] \to [0,1]) \times [0,1] \to [0,1]$ of this process at depth 1 with 1 ancestor be defined as follows,

$$G_1(\xi|x) := \sum_{n=0}^{\infty} \int_{X^n} \xi(y_1) \dots \xi(y_n) P_1^{(n)}(dy^n|x) \tag{1}$$

and the probability generating functional of this process at depth k with r ancestors, where $x^r = (x_1, x_2, \dots, x_r)$, is defined as

$$G_k(\xi|x^r) := \sum_{n=0}^{\infty} \int_{X^n} \xi(y_1) \dots \xi(y_n) P_k^{(n)}(dy^n|x^r) \tag{2}$$

Then it is shown in [16] that for every stage, the probability generating functional is multiplicative in the number of ancestors, i.e.

$$G_1(\xi|x^r) = \prod_{i=1}^{r} G_1(\xi|x_i) \tag{3}$$

and for each ancestor, it is compositional in the depth of the tree, i.e.

$$G_k(\xi|x) = G_j\{G_{k-j}[\xi|\cdot]|x\} \tag{4}$$

Moreover, the asymptotic extinction probability, $q(x)$, is given by the minimal non-negative solution of the functional equation

$$\xi(x) = G_1(\xi|x) \tag{5}$$

4 Analysis : Informational cascades as Extinction probability

In this section, we show that the problem of occurrence of informational cascades on random trees is equivalent to extinction probability of an appropriately defined random tree. Then using the results from Multi-type Galton-Watson process, we find conditions when the random tree falls into an informational cascade.

In our model description, we define two equivalent observation models, one through observation kernel $Q(x_t^k|V=v)$ and other through the CDF F^v of q_t^k conditioned on state v, where $q_t^k \in [0,1], q_t^k := P(V=1|x_t^k)$. In the following lemma, we first show how CDF of observations x_t^k are related to F^v.

Lemma 1.

$$F^v(q) = F_X^v\left(\left(\frac{Q(\cdot|V=1)}{Q(\cdot|V=0)}\right)^{-1}\left(\frac{q}{1-q}\right)\right) \tag{6}$$

Proof. Please see Appendix A.

4.1 Perfect Bayesian equilibrium

In our model, user p_t^k observes actions of its ancestors $a^{\mathcal{P}(p_t^k)}$, and its own private observation, x_t^k and therefore takes decision a_t^k through a policy g_t^k of the form $a_t^k = g_t^k(a^{\mathcal{P}(p_t^k)}, x_t^k)$. Its objective is to maximize its expected reward $\mathbb{E}^{g_t^k}\{R_t^k\}$. Since users are (myopically) strategic and there is asymmetry of information, this results in a dynamic game of asymmetric information. An appropriate notion of equilibrium for such games is perfect Bayesian equilibrium [9], which consists of an equilibrium belief profile and an equilibrium strategy profile. In general, such equilibria are difficult to compute because of interdependence of equilibrium strategies and equilibrium beliefs, which can not be sequentially decomposed, rendering computing equilibria for such problems difficult (see [9, ch. 8]). However, for this game (and other such games in the literature where users are myopic), computing such equilibria is easier as equilibrium strategies and beliefs can be easily decoupled shown as follows. User p_t^k's objective to maximize $\mathbb{E}^{g_t^k}\{R_t^k\}$ over strategies g_t^k is equivalent to maximizing $\mathbb{E}\{R_t^k|a^{\mathcal{P}(p_t^k)}, x_t^k, a_t^k\}$ over actions a_t^k for every history $(a^{\mathcal{P}(p_t^k)}, x_t^k)$. Thus

$$\mathbb{E}\{R_t^k|a^{\mathcal{P}(p_t^k)}, x_t^k, a_t^k\}$$
$$= \begin{cases} 1.P(V=1|a^{\mathcal{P}(p_t^k)}, x_t^k) - 1.P(V=0|a^{\mathcal{P}(p_t^k)}, x_t^k) \text{ if } a_t^k = 1 \\ 0 \text{ if } a_t^k = 0 \end{cases} \tag{7a}$$
$$= \begin{cases} 2P(V=1|a^{\mathcal{P}(p_t^k)}, x_t^k) - 1 \text{ if } a_t^k = 1 \\ 0 \text{ if } a_t^k = 0. \end{cases} \tag{7b}$$

Equilibrium decision strategy: (7) implies that user p_t^k takes action $a_t^k = 1$ if $P(V = 1|a^{\mathcal{P}(p_t^k)}, x_t^k) > 1/2$. We assume that in case of ties, i.e. when $P(V=1|a^{\mathcal{P}(p_t^k)}, x_t^k) = 1/2$, it chooses an action based only on its private information (as shown below in (8)).

We next define public belief $\pi_t^k \in [0,1]$ as $\pi_t^k := P(V=1|a^{\mathcal{P}(p_t^k)})$. Similar to [1, Theorem 1], it can be shown that the decision rule $P(V=1|a^{\mathcal{P}(p_t^k)}, x_t^k) > 1/2$ is equivalent to $P(V=1|a^{\mathcal{P}(p_t^k)}) + P(V=1|x_t^k) > 1$ i.e. $\pi_t^k + q_t^k > 1$. A proof is provided for convenience in Appendix C.

Thus player p_t^k takes action according to the following equilibrium decision rule g, where

$$a_t^k = \begin{cases} 1 \text{ if } q_t^k > 1 - \pi_t^k \text{ or } q_t^k = 1 - \pi_t^k \text{ and } q_t^k \geq 1/2 \\ 0 \text{ if } q_t^k < 1 - \pi_t^k \text{ or } q_t^k = 1 - \pi_t^k \text{ and } q_t^k < 1/2. \end{cases} \tag{8}$$

In the following lemma, we find the update of the belief π_t^k under above strategy.

Lemma 2. *Under the equilibrium policy g defined above and for $a \in \{0,1\}$, there exist functions $\psi_a : [0,1] \to [0,1]$ such that for any action a_t^k at time t, the common belief π_t^k is updated as*

$$\pi'_{t+1} = \psi_{a_t^k}(\pi_t^k). \tag{9}$$

Proof. Please see Appendix B.

The update of π_t^k satisfies the following property that when $(1-\pi_t^k)$ is outside the support of F^v, i.e. for $(1-\pi_t^k) \notin [\underline{b}, \bar{b}]$, then the update of π_t^k for both actions leads to same belief i.e. $\pi'_{t+1} = \psi_{a_t^k}(\pi_t^k) = \pi_t^k$. Furthermore, $\psi_{a_t^k}(0) = 0, \psi_{a_t^k}(1) = 1 \forall a_t^k$.

Thus, in this process, the probability of actions of the users is determined using the equilibrium strategies as follows. For $v \in \{0,1\}$, probability that a user takes action 0 when the value of the product is v is given by

$$\begin{aligned} &P^g(a_t^k = 0 | V = v, a^{\mathcal{P}(p_t^k)}) \\ &= P(q_t^k < 1 - \pi_t^k | V = v) + P(q_t^k = 1 - \pi_t^k, q_t^k \leq 0.5 | V = v) &(10) \\ &= P(q_t^k < 1 - \pi_t^k | V = v) + P(q_t^k = 1 - \pi_t^k, \pi_t^k \geq 0.5 | V = v) &(11) \\ &= \begin{cases} F^v(1-\pi_t^k) \text{ if } \pi_t^k \geq 0.5 \\ F_-^v(1-\pi_t^k) \text{ if } \pi_t^k < 0.5, \end{cases} &(12) \end{aligned}$$

where $F_-^v(x) = \lim_{\epsilon \to 0} F^v(x - \epsilon)$, and probability that a user takes action 1, $P^g(a_t^k = 1 | V = v, a^{\mathcal{P}(p_t^k)}) = 1 - P^g(a_t^k = 0 | V = v, a^{\mathcal{P}(p_t^k)})$.

4.2 Information cascades as extinction probability in a branching process

Based on the common belief, we define a branching process as follows. We call π_t^k as type of player p_t^k. A player of type m, where $1 - m \in [\underline{b}, \bar{b}]$, has D children of type $\psi_0(m)$ with probability $F^1(1-m)$, or D children of type $\psi_1(m)$ with probability $(1 - F^1(1-m))$[4]. For $1 - m < \underline{b}$, player p_t^k has D children of type of type $\psi_1(m)$ with probability 1, and for $1 - m > \bar{b}$, player p_t^k has D children of type of type $\psi_0(m)$ with probability 1. We define types $1 - m \notin [\underline{b}, \bar{b}]$ as "extinct" from the point of view of the branching process.

Lemma 3. *Informational cascades are equivalent to extinction probability in the branching process defined above.*

[4] Note that in the case of more than 2 actions, there will be corresponding number of types of children.

Proof. Since player of any other type m such that $1-m \notin [\underline{b}, \bar{b}]$ is in cascade, it will only have one type of children with the same type. This is because $\psi_0(m) = \psi_1(m) = m$ for $1-m \notin [\underline{b}, \bar{b}]$ as noted earlier. Thus types in $1-m \notin [\underline{b}, \bar{b}]$ are absorbing types, whereas for $1-m \in [\underline{b}, \bar{b}]$, $\psi_0(m) < m < \psi_1(m)$. This implies the statement of the lemma.

Thus using multi-type Galton-Watson branching process theory described in Section 3, starting from type x, the probability of falling into a cascade, $q(\cdot)$, is given by the minimal non-negative solution of the functional equation

$$\xi(x) = G_1(\xi|x) \tag{13}$$

where

$$G_1(\xi|x) = \sum_{n=0}^{\infty} \int_{X^n} \xi(y_1)\ldots\xi(y_n) P_1^{(n)}(dy^n|x) \tag{14a}$$

$$= \sum_{d=0}^{\infty} P(D=d)[F^1(1-x)\xi(\psi_0(x))^d + (1-F^1(1-x))\xi(\psi_1(x))^d] \tag{14b}$$

$$= F^1(1-x)\phi_D(\xi(\psi_0(x))) + (1-F^1(1-x))\phi_D(\xi(\psi_1(x))). \tag{14c}$$

Since $G_1(\cdot)$ is a probability generating functional, $\xi(x) = 1$ is always a solution of the fixed-point equation (13). Furthermore, since we assume that $P(V=1) = 1/2$, the probability of cascades is given by $q(1/2)$, where $q(\cdot)$ is given by the minimal non-negative solution of the functional equation (13).

4.3 Finite types

Definition 1. *Let $\mathcal{Q}$ be the class of observation channels such that for every given $Q \in \mathcal{Q}$ with support $[\underline{b}, \bar{b}]$, there exists a finite set $\mathcal{Z}_f(Q) \subset (\underline{b}, \bar{b})$ such that $0.5 \in \mathcal{Z}_f(Q)$ and $\forall a \in \{0,1\}$ and $\forall x \in \mathcal{Z}_f(Q)$, $\psi_a(x) \in \mathcal{Z}_f(Q)$.*

We conjecture that such channels can be characterized as $\psi_1^k(x) = \psi_0^{-1}(x)$ for some k such that $\psi_1^k(0.5) < \bar{b}$, or $\psi_0^k(x) = \psi_1^{-1}(x)$ for some k such that $\psi_0^k(0.5) > \underline{b}$. Note that one example of such a channel is the case of the binary symmetric channel (BSC), as described below. In general, it appears hard to construct a channel with finite types. Note however that the general theory from subsection *B* always holds, except that the probability of extinction is hard to compute when the channel is not finite type.

Analyzing the BSC: Consider the binary symmetric channel with cross over probability $p \in [0.5, 1]$. It is easy to show that this channel belongs to the set $\mathcal{Q}$ with $\mathcal{Z}_f(Q) = \{\frac{\bar{p}^2}{p^2+\bar{p}^2}, \bar{p}, 0.5, p, \frac{p^2}{p^2+\bar{p}^2}\}$. Let $r_i, i \in \{1,\ldots,5\}$ be the elements of $\mathcal{Z}_f(Q)$. Then

$$F^0(x) = \begin{cases} 0 \text{ if } x < \bar{p} \\ p \text{ if } \bar{p} \le x < p \\ 1 \text{ if } p < x \end{cases}, F^1(x) = \begin{cases} 0 \text{ if } x < \bar{p} \\ \bar{p} \text{ if } \bar{p} \le x < p \\ 1 \text{ if } p < x \end{cases} \tag{15}$$

(Note that F^1 stochastically dominates F^0 as expected). Then

$$\psi_0(x) = \begin{cases} x & 0 \le x < \bar{p} \\ \frac{x\bar{p}}{\bar{x}p + x\bar{p}} & \bar{p} \le \bar{x} < p \\ x & p \le x \le 1 \end{cases}, \psi_1(x) = \begin{cases} x & 0 \le x < \bar{p} \\ \frac{xp}{\bar{x}\bar{p} + xp} & \bar{p} \le x < p \\ x & p \le x < 1 \end{cases} \tag{16}$$

and

$$G_1(r|i) = F^1(1 - r_i)\phi_D(r_{i-1}) + (1 - F^1(1 - r_i))\phi_D(r_{i+1}). \tag{17}$$

5 BSC with erroneous actions

In this section, we consider the observation model of [15] where users read actions of their ancestors erroneously through a binary symmetric channel with error probability $\epsilon \in [0, 0.5]$. Specifically, each user observes o_t^k instead of a_t^k where $P(O_t^k \neq A_t^k) = \epsilon$. It is shown in [15] that $z_n := (\#1's - \#'0s)$ in $o_{1:n}$ is a sufficient statistics for the history $o_{1:n}$. Let $a = \epsilon(1-p) + (1-\epsilon)p$, $\alpha = \frac{1-p}{p}$, and π_t is a finite birth-death Markov chain with transition probabilities as $P(\pi_{t+1} = s + 1 | \pi_t = s) = a$ and $P(\pi_{t+1} = s - 1 | \pi_t = s) = 1 - a$. It is also shown that a positive cascade occurs when z_n becomes greater than or equal to $k := \lfloor \log_{\frac{1-a}{a}} \alpha \rfloor + 1$ and a negative cascade occurs when z_n becomes smaller than or equal to $-k$, and thus $\pm k$ are absorbing states of the chain. We study this model using Multi-type Galton-Watson process theory as follows. From (14c), $G_1(s|i), i \in \{-k, \ldots, k\}$ be defined as follows.

$$G_1(s| - k) = \phi_D(s_{-k}) \tag{18}$$
$$G_1(s|k) = \phi_D(s_k) \tag{19}$$

and for $i \in \{-k + 1, \ldots, k - 1\}$,

$$G_1(s|i) = \sum_{d=0}^{\infty} P(D = d)(a s_{i-1}^d + \bar{a} s_{i+1}^d) \tag{20}$$
$$= a\phi_D(s_{i-1}) + \bar{a}\phi_D(s_{i+1}). \tag{21}$$

Let $q := [-q_k, \ldots, q_k]$ where q_i represent the probabilities of falling into a cascade starting from type i. Using (13) the minimal solution of the following fixed-point represents this probability,

$$q_i = G_1(q|i) \quad \forall i \tag{22}$$

i.e. for $i = \{-k + 1, \ldots, k - 1\}$,

$$q_i = a\phi_D(q_{i-1}) + \bar{a}\phi_D(q_{i+1}) \tag{23}$$

and $q_{-k} = q_k = 1$.

Figures 2, and 3 shows the probability of cascades w.r.t. ϵ and p when the tree is a $d-$regular graph for degrees 1 and 2 respectively. Figures 4 and 5 show the same probability tree has degree distribution as Poisson with parameters 0.5 and 2 respectively.

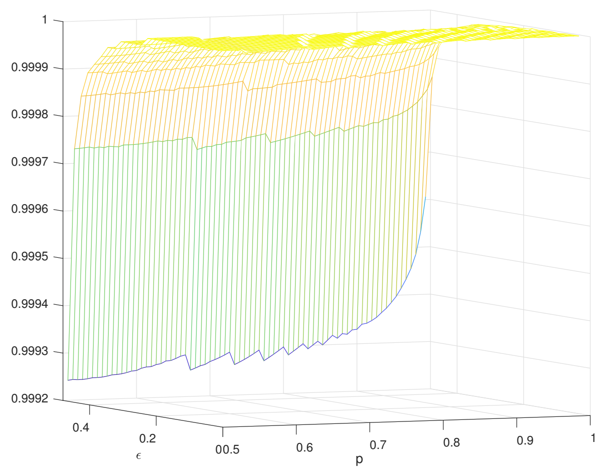

Fig. 2: Probability of tree falling into cascade for d-regular tree with d=1 where $0.5 < p < 1$ and $0 < \epsilon < 0.5$.

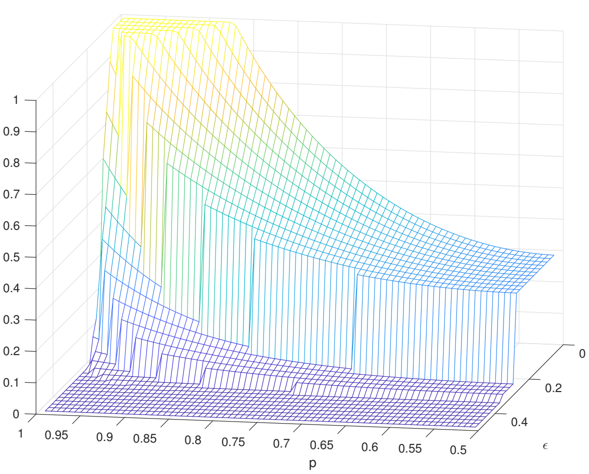

Fig. 3: Probability of tree falling into cascade for d-regular tree with d=2 where $0.5 < p < 1$ and $0 < \epsilon < 1$. Notice for smaller values of the channel and observation noise, the tree cascades with a higher probability.

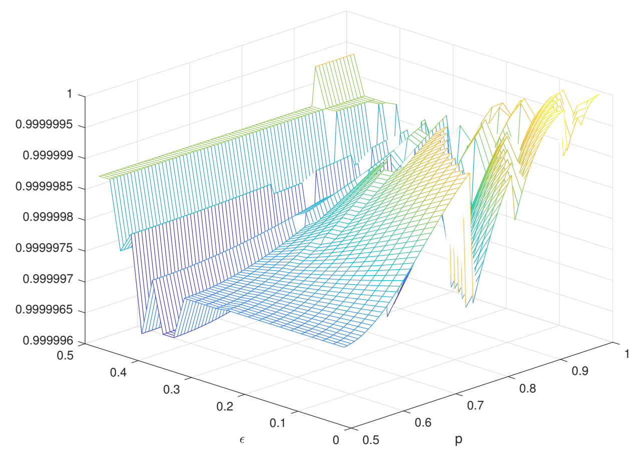

Fig. 4: Probability of tree falling into cascade for tree with degree distribution Poisson (0.2)

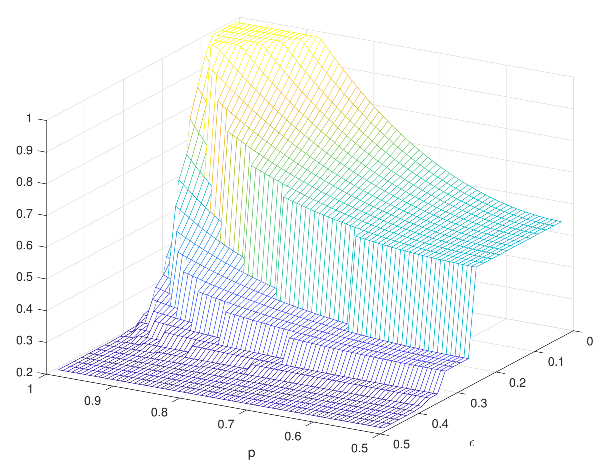

Fig. 5: Probability of tree falling into cascade for tree with degree distribution Poisson (2)

5.1 Special case: $\epsilon = 0$

We consider the case when $\epsilon = 0$ which represents the case when actions are perfectly observed. From the definition of α, a and k in the previous section, we note that $a = p$, and thus $k = 2$. Therefore there exist 5 possible states $\{-2, -1, 0, 1, 2\}$, where ± 2 are absorbing states so $q_{-2} = q_2 = 1$. Thus by repeated application of the above equations, one can deduce that

$$q_{-1} = p + \bar{p}\phi_D(q_0) \tag{24a}$$

$$q_0 = p\phi_D(q_{-1}) + \bar{p}\phi_D(q_1) \tag{24b}$$

$$q_1 = p\phi_D(q_0) + \bar{p}. \tag{24c}$$

Thus from (13), the probability of occurrence of an information cascades is given by the smallest non-negative solution of the fixed-point equation $y = p\phi_D(p + \bar{p}\phi_D(y)) + \bar{p}\phi_D(p\phi_D(y) + \bar{p})$.

Corollary 1. *The tree cascades with probability 1 if and only if* $\mathbb{E}[D] \leq \frac{1}{\sqrt{2p\bar{p}}}$.

Proof. Please see Appendix D.

Special Case: D=1 a.s. When $D = 1$ a.s., it represents the sequential decision making case as considered in [4] for the special case of observations through binary symmetric channel. Since $p\bar{p} \leq 1/4$ always hold true for $p \in [0.5, 1]$, the tree cascades with probability 1. We note that this is an alternate proof of occurrence of informational cascades of [4] for the BSC channel.

Special Case: D=2 a.s. Using the analysis above, the tree cascades with probability 1 if $2p\bar{p}2^2 < 1/2$ i.e. $\frac{2+\sqrt{3}}{4} < p \leq 1$. For $\frac{1}{2} < p < \frac{2+\sqrt{3}}{4}$, the probability of occurrence of informational cacades is the smallest fixed-point of the following equation,

$$y = p(p + \bar{p}y^2)^2 + \bar{p}(py^2 + \bar{p})^2. \tag{25}$$

6 Conclusion

In this paper, using the theory of Galton-Watson population processes and using common belief as type of every player, we studied occurrence of informational cascades in trees which are local approximations for a large Erdös-Rényi random graph network. We presented a general model to study such phenomena. We then specialized it for BSC with erroneous action observations, where we show that probability of falling into a cascade is analyzed using (finite) Multi-type Galton-Watson process. Furthermore for a binary symmetric channel, where we showed that if the expected degree of the node of the tree is less than $\frac{1}{\sqrt{2p\bar{p}}}$, then the tree cascades with probability 1, else it cascades with a probability strictly less than 1, which is given by the solution of a fixed-point equation. Our results indicate that the societies that are more tightly knit, i.e. have higher connection probability tend to herd more than the societies that are more individualistic, and have few social connections. Some interesting line of work include understanding other phenomena in such decisions on networks using other properties of Galton-Watson processes, and studying cascading for other graphical models. Some interesting cases one could consider could be Markovian branching processes or general lifetime distributions such as Bellman-Harris branching process.

Acknowledgement

The author would like to thank Varun Jog for interesting discussions and anonymous reviewers for the feedback.

A

Proof.

$$F^v(q) = P(Q_t^k < q|V = v) \tag{26a}$$
$$= P(P(V = 1|X_t^k) < q|V = v) \tag{26b}$$
$$= P\left(\frac{P(V = 1)Q(X_t^k|V = 1)}{P(V = 1)Q(X_t^k|V = 1) + P(V = 0)Q(X_t^k|V = 0)} < q|V = v\right) \tag{26c}$$
$$= P(\{x_t^k : \frac{Q(X_t^k = x_t^k|V = 1)}{Q(X_t^k = x_t^k|V = 0)} < \frac{q}{1-q}\}|V = v) \tag{26d}$$
$$= F_X^v\left(\left(\frac{Q(\cdot|V = 1)}{Q(\cdot|V = 0)}\right)^{-1}\left(\frac{q}{1-q}\right)\right) \tag{26e}$$

where $\left(\frac{Q(\cdot|V=1)}{Q(\cdot|V=0)}\right)$ inverse exists since it is non-decreasing as implied the monotone likeihood assumption, and $F_X^v(a) := P(X \le a|V = v)$.

B

Proof. Under the equilibrium policy g defined in (8), belief π_t^k can be updated using Bayes' rule as follows.

$$\pi_{t+1}^k = P^g(V = 1|a^{\mathcal{P}(p_{t+1}^k)}) \tag{27a}$$
$$= \frac{P^g(V = 1, a_t^k|a^{\mathcal{P}(p_t^k)})}{\sum_v P^g(V = v, a_t^k|a^{\mathcal{P}(p_t^k)})} \tag{27b}$$
$$= \frac{\pi_t^k P^g(a_t^k|V = 1, a^{\mathcal{P}(p_t^k)})}{(1-\pi_t^k)P^g(a_t^k|V = 0, a^{\mathcal{P}(p_t^k)}) + \pi_t^k P^g(a_t^k|V = 1, a^{\mathcal{P}(p_t^k)})} \tag{27c}$$
$$=: \psi_{a_t^k}(\pi_t^k) \tag{27d}$$

C

Proof.

$$\begin{aligned}
&P(V=1|a^{\mathcal{P}(p_t^k)}, x_t^k) \\
&= \frac{P(V=1, x_t^k|a^{\mathcal{P}(p_t^k)})}{P(V=0, x_t^k|a^{\mathcal{P}(p_t^k)}) + P(V=1, x_t^k|a^{\mathcal{P}(p_t^k)})} && \text{(28a)}\\
&= \frac{\pi_t^k P(x_t^k|V=1)}{(1-\pi_t^k)P(x_t^k|V=0) + \pi_t^k P(x_t^k|V=1)} && \text{(28b)}\\
&= \frac{\pi_t^k P(V=1|x_t^k)}{(1-\pi_t^k)P(V=0|x_t^k) + \pi_t^k P(V=1|x_t^k)} && \text{(28c)}\\
&= \frac{q_t^k \pi_t^k}{(1-q_t^k)(1-\pi_t^k) + q_t^k \pi_t^k} && \text{(28d)}
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{q_t^k \pi_t^k}{(1-q_t^k)(1-\pi_t^k) + q_t^k \pi_t^k} &\geq \frac{1}{2} && \text{(28e)}\\
q_t^k \pi_t^k &\geq (1-q_t^k)(1-\pi_t^k) && \text{(28f)}\\
\pi_t^k + q_t^k &\geq 1 && \text{(28g)}
\end{aligned}$$

D

Let $f(y) = p\phi_D(p + \bar{p}\phi_D(y)) + \bar{p}\phi_D(p\phi_D(y) + \bar{p})$. It is shown in Appendix E that $f(y)$ is monotonically increasing and strictly convex for $y > 0$.

Since $f(0) = p(\phi_D(p)) + \bar{p}\phi_D(\bar{p}) > 0$ and $f(y)$ is monotonically increasing and strictly convex, there exists another fixed point of $f(y) = y$ in the range $(0, 1)$ if and only if $\frac{d}{dy} f(y) \mid_{y=1} > 1$. The smallest fixed point is stable since the derivative of $f(y)$ is less than 1 at that point, and thus by Theorem 1, it represents the probability of cascading. The derivative of f(y) at $y = 1$ is given by,

$$\frac{d}{dy} f(y) = \frac{d}{dy} p(\phi_D(p + \bar{p}\phi_D(y))) + \bar{p}\phi_D(p\phi_D(y) + \bar{p}) \tag{29}$$

$$\frac{d}{dy} f(y) = p(\phi'_D(p + \bar{p}\phi_D(y)))\bar{p}\phi'_D(y) + \bar{p}\phi'_D(p\phi_D(y) + \bar{p})p\phi'_D(y) \tag{30}$$

$$\frac{d}{dy} f(y) \mid_{y=1} = 2p\bar{p}\phi'_D(1)\phi'_D(1) \tag{31}$$

Thus the tree cascades with probability 1 if and only if $2p\bar{p}(\phi'_D(1))^2 \leq 1$ i.e. $\phi'_D(1) = \mathbb{E}[D] \leq \frac{1}{\sqrt{2p\bar{p}}}$. Otherwise the tree cascades with a non-zero probability which is the smallest fixed-point of $f(y) = y$.

E

Lemma 4. *$f(y) = p\phi_D(p+\bar{p}\phi_D(y))+\bar{p}\phi_D(p\phi_D(y)+\bar{p})$ is monotonically increasing and strictly convex for $y > 0$.*

Proof. We will show that $p\phi_D(p+\bar{p}\phi_D(y))$ is strictly increasing and convex, and the proof of the other part is identical.

$$\frac{d}{dy}p\phi_D(p+\bar{p}\phi_D(y)) \tag{32a}$$
$$= p\phi_D^{'}(p+\bar{p}\phi_D(y))\bar{p}\phi_D^{'}(y) \tag{32b}$$
$$> 0 \tag{32c}$$

where the last inequality is true since $\phi_D(y) = \mathbb{E}[y^D] > 0$ and $\frac{d}{dy}\phi_D(y) = \mathbb{E}[Dy^{D-1}] > 0$ since $y > 0$ and $D \geq 1$ a.s..

$$\frac{d^2}{dy^2}p\phi_D(p+\bar{p}\phi_D(y)) \tag{32d}$$
$$= \frac{d}{dy}p\phi_D^{'}(p+\bar{p}\phi_D(y))\bar{p}\phi_D^{'}(y) \tag{32e}$$
$$= p\bar{p}\left(\phi_D^{'}(p+\bar{p}\phi_D(y))\phi_D^{''}(y) + \phi_D^{''}(p+\bar{p}\phi_D(y))\bar{p}(\phi_D^{'}(y))^2\right) \tag{32f}$$
$$> 0 \tag{32g}$$

where the last inequality is true by similar arguments as before.

References

1. Acemoglu, D., Dahleh, M.A., Lobel, I., Ozdaglar, A.: Bayesian learning in social networks. The Review of Economic Studies **78**(4), 1201–1236 (2011)
2. Acemoglu, D., Makhdoumi, A., Malekian, A., Ozdaglar, A.: Fast and slow learning from reviews. Tech. rep., National Bureau of Economic Research (2017)
3. Banerjee, A.V.: A simple model of herd behavior. The Quarterly Journal of Economics pp. 797–817 (1992)
4. Bikhchandani, S., Hirshleifer, D., Welch, I.: A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy **100**(5), pp. 992–1026 (1992), http://www.jstor.org/stable/2138632
5. Bubeck, S., Devroye, L., Lugosi, G.: Finding adam in random growing trees. Random Structures & Algorithms **50**(2), 158–172 (2017)
6. Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. Physical review E **70**(6), 066111 (2004)
7. Dhounchak, R., Kavitha, V., Altman, E.: A viral timeline branching process to study a social network. arXiv preprint arXiv:1705.09828 (2017)
8. Fanti, G., Kairouz, P., Oh, S., Ramchandran, K., Viswanath, P.: Hiding the rumor source. IEEE Transactions on Information Theory (2017)
9. Fudenberg, D., Tirole, J.: Perfect bayesian equilibrium and sequential equilibrium. journal of Economic Theory **53**(2), 236–260 (1991)

10. Harel, M., Mossel, E., Strack, P., Tamuz, O.: The speed of social learning. arXiv preprint arXiv:1412.7172 (2014)
11. Harris, T.E.: The theory of branching processes. Courier Corporation (2002)
12. Jackson, M.O.: Social and economic networks. Princeton university press (2010)
13. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146. ACM (2003)
14. Kendall, D.G.: Branching processes since 1873. Journal of the London Mathematical Society **1**(1), 385–406 (1966)
15. Le, T.N., Subramanian, V., Berry, R.: The impact of observation and action errors on informational cascades. In: Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on. pp. 1917–1922 (Dec 2014). https://doi.org/10.1109/CDC.2014.7039678
16. Moyal, J.: Multiplicative population chains. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. vol. 266, pp. 518–526. The Royal Society (1962)
17. Newman, M.: Networks: an introduction. Oxford university press (2010)
18. Peres, Y., Racz, M.Z., Sly, A., Stuhl, I.: How fragile are information cascades? arXiv preprint arXiv:1711.04024 (2017)
19. Shah, D., Zaman, T.: Rumors in a network: Who's the culprit? IEEE Transactions on information theory **57**(8), 5163–5181 (2011)
20. Smith, L., Sörensen, P.: Pathological outcomes of observational learning. Econometrica **68**(2), 371–398 (2000). https://doi.org/10.1111/1468-0262.00113, http://dx.doi.org/10.1111/1468-0262.00113
21. Vasal, D., Anastasopoulos, A.: Decentralized Bayesian learning in dynamic games. In: Allerton Conference on Communication, Control, and Computing (2016), https://arxiv.org/abs/1607.06847
22. Vasal, D., Subramanian, V., Anastasopoulos, A.: Incentive design for learning in user-recommendation systems with time-varying states. In: 2015 49th Asilomar Conference on Signals, Systems and Computers. pp. 1080–1084 (Nov 2015). https://doi.org/10.1109/ACSSC.2015.7421305