How Robust is the Wisdom of the Crowds?

Noga Alon
Tel Aviv University and
Microsoft Research, Israel
nogaa@tau.ac.il

Michal Feldman
Tel Aviv University and
Microsoft Research, Israel
michal.feldman@cs.tau.ac.il

Omer Lev
Hebrew University and
Microsoft Research, Israel
omerl@cs.huji.ac.il

Moshe Tennenholtz
Technion, Israel
moshet@ie.technion.ac.il

Abstract

We introduce the study of adversarial effects on wisdom of the crowd phenomena. In particular, we examine the ability of an adversary to influence a social network so that the majority of nodes are convinced by a falsehood, using its power to influence a certain fraction, $\mu < 0.5$ of N experts. Can a bad restaurant make a majority of the overall network believe in the quality of that restaurant by misleading a certain share of food critics into believing its food is good, and use the influence of those experts to make a majority of the overall network to believe in the quality of that restaurant? We are interested in providing an agent, who does not necessarily know the graph structure nor who the experts are, to determine the true value of a binary property using a simple majority. We prove bounds on the social graph's maximal degree, which ensure that with a high probability the adversary will fail (and the majority vote will coincide with the true value) when he can choose who the experts are, while each expert communicates the true value with probability $p > 0.5$. When we examine expander graphs as well as random graphs we prove such bounds even for stronger adversaries, who are able to pick and choose not only who the experts are, but also which ones of them would communicate the wrong values, as long as their proportion is $1 - p$. Furthermore, we study different propagation models and their effects on the feasibility of obtaining the true value for different adversary types.

1 Introduction

Understanding the way opinions are formed and disseminated throughout a social setting has been explored for the past few decades, following the key insight by Rogers [Rogers, 2003] that new and unfamiliar products (and ideas) are spread by a core group of "early adopters" which, depending on their size, their positions in the social graph, and their inclinations can either make a product successful or doom it to oblivion.

A similar dynamic can be observed when people form an opinion regarding uncertain properties of some product or service, such as the quality of a new restaurant. Members of a social network are usually either experts — those who form their opinions based on first-hand experience — or non-experts who are influenced by the opinions of their expert friends. For example, people forming opinions about a new restaurant will often rely on a set of experts, such as food critics, or first-hand observers, to form their own opinion. Once experts report their opinions, it is reasonable for them to form their own opinions by conforming with the majority opinion of their expert friends. This phenomenon is magnified in online social networks, where new ideas, opinions and technologies spread easily through the network, and experts can have a huge effect on the population.

In this work we take an adversarial approach, where an adversary attempts to intentionally mislead the population and disseminate a falsehood throughout the network. Considering again the restaurant example, a mediocre restaurant (that can improve its quality with some additional effort) might try to choose the restaurant reviewers (=experts) that come in during its warm-up period in order to influence them. We consider different adversarial (and non-adversarial) models that differ from each other in the power of the adversary to directly approach specific individuals and affect their opinions.

As opinions propagate throughout the graph, the following question arises: how can an oblivious observer infer the ground truth from the network? That is, while agents who know who are the experts in the system can approach them when forming an opinion, an oblivious observer might not even be able to identify these experts. Moreover, an oblivious observer might not be familiar with the network structure. All that our observer can see is the number of "voters" for each opinion. Thus, it is of great interest to make sure the majority of agents hold the ground truth. In this work, we study network properties that make the network *robust* against adversarial attempts to mislead the population. Specifically, we wish to identify network topologies and opinion formation models in which, with high probability, the majority of the population believe in the ground truth.

Our model We consider a social network, given by an undirected graph $G = (V, E)$, with $|V| = n$ nodes, corresponding to agents. The agents live in a world with a ground truth, which is either *red* (R) or *blue* (B). It will be convenient to assume, without loss of generality, that the ground truth is R. Agents form opinions about the state of the world through a

process we shall describe soon. We denote by $c(v) \in \{R, B\}$ the opinion of node v.

A set $V' \subseteq V$ of size $\mu|V|$ (for a fixed μ), constitutes the *expert* set. Experts can be chosen randomly or adversarially. For both adversary types considered here, it is assumed the adversary has complete information of the graph, and we assume the adversary seeks to fool our observer, and make it seem as if the underlying truth is different than it really is. We distinguish between three models of expert formation.

- Strong adversary: an adversary chooses an expert set $V' \subseteq V$ (such that $|V'| = \mu|V|$), and assigns opinions to agents in V' satisfying the following equations: $|\{v \in V'|c(v) = R\}| = (\frac{1}{2} + \delta)|V'|$ and $|\{v \in V'|c(v) = B\}| = (\frac{1}{2} - \delta)|V'|$, for some fixed δ.
- Weak adversary: an adversary chooses an expert set $V' \subseteq V$ (such that $|V'| = \mu|V|$). Experts receive signals about the state of the world, and are more likely to be correct than incorrect. Specifically, for every agent $v \in V'$ independently, it holds that $c(v) = R$ with probability $1/2 + \delta$ and $c(v) = B$ with probability $1/2 - \delta$, for some fixed δ.
- Random process: a set of $\mu|V|$ nodes are chosen uniformly at random forming expert set V'. Opinion formation of agents in V' is as in the weak adversary model.

We next describe the dissemination process. For every agent v, let $N(v)$ denote the set of agent v's neighbors; i.e., $N(v) = \{u|(v, u) \in E\}$. Every agent $v \notin V'$ forms its opinion by conforming with the majority opinion of its expert neighbors. That is, if $|\{u \in N(v) \cap V'|c(u) = R\}| > |\{u \in N(v) \cap V'|c(u) = B\}|$, then $c(v) = R$. If the inequality is reversed, then $c(v) = B$. Finally, if it holds with equality, then $c(v) = R$ with probability $1/2$ and $c(v) = B$ with probability $1/2$. Note that this tie breaking rule implies that nodes with no neighbors in V' form their opinion uniformly at random. In Section 4, we examine whether a more "viral" propagation model is more beneficial than this one.

A network is said to be *robust* against a particular adversary model if, with high probability, the majority of agents hold the true opinion, despite an adversary's attempt to deceive.

Our results We find that the power of the adversary is strongly affected by the network structure. Specifically, in the case of a weak adversary (i.e., one who can choose the set of experts, but cannot determine the partition of opinions within the set), we show that any network is robust as long as the highest degree does not exceed some upper bound[1]. For expanders and random graphs, we establish robustness results even with respect to a strong adversary (i.e., one who can choose the set of experts, and also determine the partition of opinions within the set). Finally, we study similar questions under an extended propagation model, where similar dynamics take place iteratively (so that experts' opinions propagate beyond their direct friends in the network). We find that an iterative propagation can either help or harm a weak adversary. Similarly, under a random process, iterative propagation can be either helpful or harmful. In the case of a strong adversary, we conjecture that an iterative propagation can never harm the adversary, and we give an example where it can be helpful.

[1]Bounds which also seem to be ensured, for example, by Facebook's maximal bound for friend numbers (though the more recent and unlimited "follower" status upends this property, naturally).

Related Work The discussion of how opinions and ideas spread through society has been an active research field since the seminal work of Rogers 2003, which introduced the concept of "early adopters" as a vanguard from which an invention may spread to the rest of society. From that line of research, various directions were adopted to try and understand – both theoretically and empirically – how agents might adopt a certain property depending on how many were adopting it in the society surrounding them.

On issues on which agents have an innate opinion while being influenced externally, research expanded to cases where agents do not have full information, first for very limited settings [Farrell and Saloner, 1985], and then for richer ones, which involved finding equilibria in these scenarios [Katz and Shapiro, 1985], and creating models to incorporate influence of agents on each other, first on limited, lattice-like, graphs [Blume, 1993; Ellison, 1993], and then on general graphs [Morris, 2000; Young, 2000; Tangand *et al.*, 2009; Kameda *et al.*, 1997; López-Pintado and Watts, 2008]. A particular strand of this research focused on "informational cascade" or "herd mentality" when choices are made sequentially in political settings (where there is a ground truth, which agents aim to reach) or market cases (where there is no underlying truthful choice) [Bikhchandani *et al.*, 1992; Banerjee, 1992; Arthur, 1989], with more recent research trying to find various equilibria in such cascades [Alon *et al.*, 2012], and examining such cascades on graph structure, including, as in our scenario, on random graphs [Watts, 2002].

In particular, two fairly recent papers dealing with cascades when there is a ground truth are related to ours: Both Mossel et al. 2014 and Feldman et al. 2014 include a ground truth which agents have a higher probability of supporting. Despite different synchronous modes between these papers, both utilize the same dynamic we explore: nodes adopting the color of the majority of their already colored neighbors. However, both papers focus on reaching a consensus in the social network (as we don't allow agents to change their views, this goal is irrelevant in our model), and strategies for a truthful consensus, while we only strive to have the majority of the agents be truthful, and include an active adversary, trying to prevent acceptance of the ground truth.

A closely related track of research explores "word of mouth" models for information diffusion, where agents have no particular opinion. Exploring word of mouth travel involved empirical work [Brown and Reingen, 1987], as well as theoretical one, examining propagation models [Granovetter, 1973; Goldenberg *et al.*, 2001; Young, 2009], attempting to explain how marketing works using combination of ads (to "early adopters") and word of mouth (an overview of much of the research can be seen in Mahajan et al. 1990 and Young 2009), and trying to find influential agents in the network [Kempe *et al.*, 2003]. Some of the research in this direction has evolved into work on recommendation

systems, where agents have trust relationships according to which they accept recommendations [Andersen *et al.*, 2008; Domingos and Richardson, 2001; Richardson and Domingos, 2002]. Similar to these models (in particular, Andersen et al. 2008), one can look at our experts as the opinionated nodes in these recommendation systems.

2 Weak Adversaries

In this section we consider the robustness of networks to weak adversaries. We establish a property that ensures that the ground truth is held by the majority of the population with high probability. We begin with a few simple examples.

Example 1. *Consider the clique graph. There is a higher probability a majority of V' would be Red, and then, all nodes that are not in V' will turn Red. It is easy to verify that in this example the majority of the population will turn Red with high probability.*

Example 2. *Consider a "star" network, with a single central mode connected to all other nodes. If the central mode belongs to V', then, with probability $\frac{1}{2} - \delta$ it will be $Blue$, thereby causing all nodes not in V' to turn $Blue$. Thus, in this example, the majority will be $Blue$ with probability close to $1/2$, so the network is not robust against a weak adversary.*

We shall now show that for any $\epsilon < \mu, \delta$, for $0 < \mu, \delta < \frac{1}{2}$, if n is sufficiently large, we have a sufficient criterion for majority to reflect the truth.

Theorem 1. *For $0 < \epsilon < \mu, \delta < \frac{1}{2}$, if n is sufficiently large, there is an absolute positive constant c_1 so that if the highest degree Δ satisfies*

$$\Delta \leq c_1 \frac{\epsilon \delta^4 \mu n}{\log(1/\epsilon)}$$

then majority over all vertices gives the truth with probability at least $1 - \epsilon$.

This is nearly tight, (though the precise best possible dependence on δ and ϵ is not), as shown by the following.

Proposition 1. *There is an absolute positive constant c_2 such that for all ϵ, δ, μ and all large n there is an example of a graph $G = (V, E)$ with $|V| = n$, highest degree*

$$\Delta \leq c_2 \frac{\delta^2 \mu n}{\log(1/\epsilon)}$$

and a choice of $V' \subset V$, $|V'| = \mu n$ for which the majority fails to give the truth with probability exceeding ϵ.

Proof of Theorem 1: Put $V'_T = \{v \in V' : c(v) = R\}, V'_F = \{v \in V' : c(v) \neq R\}$. By Chernoff's Inequality (c.f. [Alon and Spencer, 2008], appendix A), if n is sufficiently large

$$Prob(|V'_T| - |V'_F| < \delta \mu n) < \epsilon/4. \quad (1)$$

Split the vertices of $V - V'$ into three groups, V_H, V_L and V_N, according to the number of their neighbors in V', as follows. Put $M = c_3 \frac{1}{\delta^2} \log(1/\epsilon)$ (with c_3 an absolute constant chosen appropriately).

$$V_H = \{v \in V - V' : |N_v \cap V'| \geq M\}$$
$$V_L = \{v \in V - V' : 1 \leq |N_v \cap V'| < M\}$$
$$V_N = \{v \in V - V' : N_v \cap V' = \emptyset\}$$

Put $V_{HT} = \{v \in V_H : c(v) = R\}$, $V_{HF} = V_H - V_{HT}$ and define $V_{LT}, V_{LF}, V_{NT}, V_{NF}$ similarly.

By Chernoff, again, the opinions in V_N are balanced with high probability and in particular, for sufficiently large n

$$Prob(|V_{NF}| - |V_{NT}| > \delta \mu n/4) < \epsilon/4. \quad (2)$$

Fix a vertex $v \in V_H$. The probability that the opinions of at least half of his neighbors are B is the probability that a binomial random variable with parameters $\ell \geq M$ and $p = \frac{1}{2} + \delta$ has a value of at most $\frac{\ell}{2}$, which is, by Chernoff, at most

$$e^{-c_4 \delta^2 M} < \epsilon^3/16 \leq \frac{\epsilon \mu \delta}{16},$$

where the first inequality follows by choosing c_3 (in the definition of M) appropriately, and the second by the fact that $\epsilon < \mu$ and $\epsilon < \delta$. It follows that the probability that the opinion of v is $c(v) = B$ is at most $\epsilon \mu \delta/16$, and hence, by linearity of expectation, the expected size of V_{HF} is at most $|V_H| \epsilon \mu \delta/16$. By Markov's Inequality this implies

$$Prob(|V_{HF}| - |V_{NT}| > \frac{\delta \mu n}{4}) < Prob(|V_{HF}| > \frac{\delta \mu |V_H|}{4}) < \frac{\epsilon}{4} \quad (3)$$

It remains to estimate the contribution of V_L vertices opinions. This is done using the second moment method described, for example, in [Alon and Spencer, 2008], Chapter 4:

For each vertex $v \in V_L$, let X_v be the indicator random variable with value 1 iff $c(v) = B$ (and 0 otherwise). Note that since v has at least one neighbor in V', the probability that $X_v = 1$ is at most $1/2 - \delta$. Put $X = \sum_{v \in V_N} X_v$, then X is the random variable whose value is exactly $|V_{LF}|$. Put $m = |V_L|$ and note that by linearity of expectation the expected value of X satisfies $E[X] \leq m(1/2 - \delta)$. We next bound the variance of X. For $v, v' \in V_L$, let $v \sim v'$ denote that $v \neq v'$ and v, v' have at least one common neighbor in V'. Note that if $v \neq v'$ do not satisfy $v \sim v'$ then the random variables $X_v, X_{v'}$ are independent, and hence their covariance is 0. We thus have

$$Var[X] = \sum_{v \in V_N} Var(X_v) + \sum_{v, v' \in V_L, v \sim v'} Cov(X_v, X'_v),$$

where the sum ranges over all ordered pairs $v, v' \in V_L, v \sim v'$.

As each X_v is an indicator random variable, its variance is at most its expectation. Similarly, for $v, v' \in V_L, v \sim v'$:

$$Cov(X_v, X'_v) = E[X_v \cdot X_{v'}] - E[X_v]E[X_{v'}] \leq E[X_v \cdot X_{v'}] < \frac{1}{2}.$$

Note, crucially, that for each $v \in V_L$, the number of $v' \in V_L$ so that $v \sim v'$ is smaller than $M(\Delta - 1)$, as v has less than M neighbors in V', and each of them can have at most $\Delta - 1$ other neighbors in V_L. We thus conclude that

$$Var[X] \leq E[X] + |V_L|(\Delta - 1)M \cdot \frac{1}{2} <$$
$$m/2 + m(\Delta - 1)M \cdot \frac{1}{2} < m\Delta M/2$$

Note that the probability that $X = |V_{LF}|$ is at least $m/2 + \delta \mu n/4$ is at most the probability that it exceeds its expectation, which is at most $m(1/2 - \delta)$, by at least

$$\delta m + \delta \mu n/4 \; (\; \geq \delta \sqrt{\mu m n} \;).$$

By Chebyshev's Inequality we conclude that if

$$\Delta \leq c_1 \frac{\epsilon \delta^4 \mu n}{\log(1/\epsilon)}$$

then

$$\begin{aligned} &Prob(|V_{LF}| - |V_{LT}| \geq \delta\mu n/2) = \\ &Prob(|V_{LF}| \geq m/2 + \delta\mu n/4) \leq \frac{Var[X]}{\delta^2 \mu m n} \leq \\ &\frac{m\Delta M}{2\delta^2 \mu m n} = c_5 \frac{\Delta \log(1/\epsilon)}{\delta^4 \mu n} \leq \epsilon/4 \end{aligned} \tag{4}$$

for an appropriate choice of c_1 in Theorem 1.

Combining (1),(2), (3) and (4) we conclude that with probability at least $1-\epsilon$ none of the events in these inequalities holds. It is easy to see that if this is the case then the majority opinion is indeed R, as needed.

Proof of Proposition 1: Let $G = (V, E)$ be a graph consisting of vertex disjoint stars, each of size $t = c_6 \frac{\delta^2 \mu n}{\log(1/\epsilon)}$. For the set V' we choose the centers of $q = c_7 \frac{1}{\delta^2} \log(1/\epsilon)$ of the stars, as well as $\mu n - q$ additional vertices in some $\lceil (\mu n - q)/t \rceil$ other stars. For the right choice of c_6, c_7, the probability that at least $q/2 + 2\delta q$ of the centers in V' will have the wrong opinion B is at least 2ϵ. They will affect all the leaves of the stars giving an advantage of at least $4\delta\mu n$ to the wrong opinion over the truth T among the vertices of these q stars. The probability that the other opinions will change the majority is smaller than ϵ, implying the desired result. The detailed computation is omitted.

3 Strong Adversaries

Here we examine the robustness of networks against a strong adversary. We begin with an example demonstrating the potential vulnerability of networks against a strong adversary.

Example 3. *We show here a case where a network is more robust against a weak adversary than against a strong one: Let $\delta = \frac{1}{4}$, $\mu = \frac{1}{10}$. The graph consists mostly of atoms (i.e., nodes not connected to any other node), except for $\frac{5}{40}n$ nodes, which are divided into quintets – every 5 nodes form a clique. As the maximal degree is fixed at 4, according to Theorem 1, for a large enough n, majority will prevail and reflect the underlying truth (i.e., Red) with very high probability.*

However, a strong adversary shall choose one node in each of the quintets to be an expert ($\frac{1}{40}n$ nodes) which it will turn $Blue$, and $\frac{3}{40}n$ other nodes which are atoms, to be experts which will turn Red. The left over atoms ($\frac{4}{5}n$ nodes) will be, by Chernoff, roughly equally divided into Red and $Blue$ (about $\frac{2}{5}n$ nodes each) with high probability, while all the quintets will turn $Blue$. Ultimately, we have, with high probability, $(\frac{19}{40}+o(1))n$ Red nodes and $(\frac{21}{40}+o(1))n$ $Blue$ nodes – making $Blue$, the adversary's choice, the majority winner.

3.1 Expanders

An (n, d, λ)-graph is a d-regular graph on n vertices in which the absolute value of every eigenvalue besides the first is at most λ.

The following theorem will be useful in establishing the robustness of expanders against strong adversaries.

Theorem 2. *Let $G = (V, E)$ be an (n, d, λ)-graph, let A and B be subsets of V and assume that $|A| > |B|$. Let X be the set of all vertices v of G satisfying $|N(v) \cap B| \geq |N(v) \cap A|$, where $N(v)$ is the set of neighbors of v in G. Then*

$$|X| \leq \frac{2\lambda^2}{d^2} \frac{|A|(1-|A|/n) + |B|(1-|B|/n)}{(|A|-|B|)^2} n^2.$$

To prove the theorem we need the following known result that shows that if λ is much smaller than d, then for every set of vertices A, most vertices have roughly the "right" number of neighbors in A.

Lemma 1 ([Alon and Spencer, 2008], Theorem 9.2.4)**.** *Let $G = (V, E)$ be an (n, d, λ)-graph, and let $A \subset V$ be an arbitrary set of vertices of G, then*

$$\sum_{v \in V} (|N(v) \cap A| - \frac{d|A|}{n})^2 \leq \lambda^2 |A|(1 - \frac{|A|}{n})$$

We also need the following simple fact.

Lemma 2. *Let $a > b$ be two reals and suppose $x \geq y$. Then $(x-b)^2 + (y-a)^2 \geq (a-b)^2/2$.*

Proof of Theorem 2: By Lemma 1

$$\sum_{v \in V} (|N(v) \cap A| - \frac{d|A|}{n})^2 \leq \lambda^2 |A|(1 - \frac{|A|}{n})$$

and

$$\sum_{v \in V} (|N(v) \cap B| - \frac{d|B|}{n})^2 \leq \lambda^2 |B|(1 - \frac{|B|}{n}).$$

therefore

$$\begin{aligned} &\sum_{v \in V} (|N(v) \cap A| - \frac{d|A|}{n})^2 + (|N(v) \cap B| - \frac{d|B|}{n})^2 \\ &\leq \lambda^2 [|A|(1 - \frac{|A|}{n}) + |B|(1 - \frac{|B|}{n})]. \end{aligned}$$

By Lemma 2 (with $a = \frac{d|A|}{n}$ and $b = \frac{d|B|}{n}$), each vertex $v \in X$ contributes to the left hand side of the last inequality at least $\frac{d^2(|A|-|B|)^2}{2n^2}$ and we thus conclude that

$$|X| \frac{d^2(|A|-|B|)^2}{2} \leq \lambda^2 [|A|(1 - \frac{|A|}{n}) + |B|(1 - \frac{|B|}{n})] n^2,$$

completing the proof.

Theorem 2 implies that if the network in our social voting game is an (n, d, λ)-graph with λ much smaller than d and d sufficiently large as a function of μ and δ, then majority gives the truth (deterministically) even against a *strong adversary*, that is, an adversary who is allowed to select a set V' of $\mu|V|$ experienced vertices, and is also allowed to select any partition of it into two disjoint sets A and B with $|A| = (\frac{1}{2}+\delta)|V'|$ and $|B| = (\frac{1}{2} - \delta)|V'|$, where all members of A get the truth Red and all those in B get $Blue$.

Theorem 3. *Let $G = (V, E)$ be an (n, d, λ)-graph and suppose that*

$$\frac{d^2}{\lambda^2} > \frac{1}{\delta^2 \mu (1 - \mu + 2\delta\mu)}.$$

Then for any strong adversary as above the majority gives the truth. In particular, if G is a Ramanujan graph, that is, $\lambda \leq 2\sqrt{d-1}$ this is the case provided

$$d \geq \frac{4}{\delta^2 \mu (1 - \mu + 2\delta\mu)}.$$

Proof. It suffices to check that if A and B are disjoint sets of vertices satisfying $|A| = (\frac{1}{2}+\delta)\mu n$ and $|B| = (\frac{1}{2}-\delta)\mu n$, then the number of vertices v outside $A \cup B$ for which $|N(v) \cap B| \geq |N(v) \cap A|$ is smaller than $(\frac{1-\mu}{2} + \delta\mu)n$. By Theorem 2 this number is smaller than

$$\frac{2\lambda^2}{d^2}\frac{\mu n}{4\delta^2\mu^2 n^2}n^2$$

which satisfies the required bound provided the assumption on d^2/λ^2 holds. □

3.2 Binomial Random Graphs

$G = G(n,p)$ is the binomial random graph, in which each edge has a probability of p of existing in the graph. As with expander graphs, these graphs have particular properties that let us elaborate on their robustness in the face of strong adversaries. Specifically, with high probability (that is, with probability that tends to 1 as n tends to infinity) the statement of Theorem 3 holds even for average degree that is a bit smaller than the one above. For this model the following holds:

Theorem 4. *There exists an absolute constant c so that if $\mu < 1/2$ and*

$$d = np \geq c \cdot max\ \{\ \frac{\log(1/\mu)}{\delta^2}, \frac{1}{\mu\delta}\ \}$$

then with high probability if $G = G(n,p)$ then for any strong adversary as above the majority gives the truth.

The proof is by showing that for any fixed two disjoint sets of vertices A and B of sizes as in the previous section, the probability that there are too many vertices having more neighbors in B than in A is sufficiently small to ensure that even after multiplying it by the number of possible sets A and B the number obtained still tends to zero as n tends to infinity. Its details are omitted due to space constraints

4 Iterative Propagation

We now seek to understand how changing the propagation model from the one we have used so far to a more "viral" one affects the adversary, and how that effect changes as we accord different powers to the adversaries — random, weak and strong.

So far, we have dealt with a limited propagation, which only handles the influence experts have over the agents directly connected to them. An alternative model, *iterative propagation* tries to incorporate the influence agents have over others in the social network. I.e., once the experts' view propagates to the agents connected to it, those agents propagate their view to those connected to them, and so on, until the whole connectivity component is influenced.

More formally, starting from V', if $|\{u \in N(v) \cap V' | c(u) = R\}| > |\{u \in N(v) \cap V' | c(u) = B\}|$, then $c(v) = R$. If the inequality is reversed, then $c(v) = B$, and if it holds with equality, then $c(v) = R$ with probability $\frac{1}{2}$ and $c(v) = B$ with probability $\frac{1}{2}$ (same as with the limited propagation we used). Now, defining V^1 as $V' \cup \bigcup_{v \in V'} N(v)$, we repeat this process, but with V^1 instead of V', creating V^2, and then over and over until V^k includes all vertices of the connected components containing vertices of V.

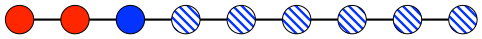

Figure 1: Iterative propagation will cause the line graph to turn $Blue$ if last node is colored $Blue$, ensuring the color's propagation (striped nodes indicate propagation).

4.1 Strong Adversaries

Recall that strong adversaries can choose V' and the color of each of its nodes, subject to having $(1/2+\delta)|V'|$ red nodes and $(1/2-\delta)|V'|$ blue ones. We conjecture that in this case iterative propagation is always beneficial. We shall demonstrate its usefulness in a specific case.

Example 4. Iterative Propagation helps the adversary

Consider a graph consisting of a path on $m = 0.024n$ nodes together with $0.976n$ isolated nodes. Suppose $\mu = \delta = \frac{1}{10}$. Thus, the fraction of blue nodes in V' is $\frac{1}{2} - \delta = \frac{2}{5}$.

With limited propagation, it is not difficult to check that the best strategy of the adversary is to place $m/3 = 0.008n$ of the blue nodes along the path ensuring that all of it becomes blue, and place the remaining $0.04n - m/3 = 0.032n$ blue nodes and $0.06n$ red nodes outside the path. (For simplicity we omit all floor and ceiling signs whenever these are not crucial). By Chernoff, with high probability the total number of blue nodes will be $0.024n + 0.032n + (1/2 + o(1))0.884n = (0.498 + o(1))n$. Thus, in this case if n is sufficiently large then with high probability the majority is red, meaning that the adversary fails.

On the other hand with iterative propagation the adversary can place one blue node in the path and place all the other red and blue nodes outside it. This will ensure, by Chernoff, that with high probability the total number of blue nodes will be $0.024n + 0.04n - 1 + (1/2 + o(1))0.876n = (0.502 + o(1))n$. Thus, with high probability with iterative propagation $Blue$ attains the majority of nodes.

4.2 Weak Adversaries

Here we consider weak adversaries, and in this case, iterative propagation can both help and hinder the adversary.

Example 5. Iterative Propagation can help the adversary:

Consider a path on n nodes with $\mu = \frac{1}{10}$ and $\delta = \frac{1}{10}$.

With limited propagation, as the maximum degree in the graph is the constant 2, by Theorem 1 if n is sufficiently large Red wins with high probability, i.e., the adversary fails.

However, with iterative propagation, the adversary can choose V' to consist of the first $\mu n = n/10$ nodes of the path. The probability that the last among those is $Blue$ is $2/5$, and in this case we will have a majority of $Blue$ in the process (see Figure 1).

Example 6. Iterative propagation can harm the adversary:

Consider a graph which is the disjoint union of a star on $n/14$ nodes together with a 3-regular high girth expander on the remaining nodes. Let $\mu = \delta = 1/12$.

With limited propagation the adversary can place the center of the star in V', and place the remaining nodes of V' in the expander, so that no node of the expander has more than

one neighbor in V' (this is possible by a simple greedy procedure). Now there are $3(|V'|-1)$ expander nodes that are neighbors of the ones in V', and including the members of V' in the expander we have $n/3 - O(1)$ nodes. Thus, with high probability the difference between the number of Red and Blue nodes in the expander will be $(1+o(1))0.2 \cdot n/3 = (1/15+o(1))n$. However, with probability 0.4 the star's center is Blue, turning all the star blue, giving majority to Blue.

Consider now iterated propagation. Any node with a neighbor in V' (or more than one) also becomes Red with probability at least 0.6 during the first iteration. As nodes that have no common neighbor are independent this means that after the first iteration, with high probability the number of Red nodes in the expander exceeds that of the number of Blue nodes by at least a factor of $3/2 - o(1)$. A similar argument shows that with high probability this is also the case after any constant number of steps, where here it is convenient to use the assumption that the expander is of high girth. This ensures the neighborhood of any vertex is locally a tree, and if its distance from V' is r, it gets a color in iteration number r. Therefore, with high probability at least $0.6 - o(1)$ fraction of the expander's nodes will be Red, and even if the star is Blue, Red will still have majority. We omit the details.

4.3 Random Process

Finally we consider the random process model, with a passive adversary, in which both V' and the node color assignment is done randomly, and in this case iterative propagation can be both a curse and a blessing for the adversary:

Example 7. Iterative Propagation can help the adversary:
Take $n/60$ paths, each of length 3. Add a special node s and join it to the first node of each path. The graph we consider consists of this structure together with $19n/20 - 1$ additional isolated nodes. Take $\mu = \delta = 1/10$. With limited propagation Red is the majority with high probability. Indeed, by Chernoff Red has an advantage of at least $(1-o(1))0.2 \cdot 0.1 \cdot (59n/60 - 1) > (0.019 - o(1))n$ among the nodes besides s and its neighbors, hence even if s and all its $n/60$ neighbors become Blue, Red still has the majority.

With propagation, however, with probability $0.4 \cdot 0.1 = 0.04$ s is in V' and is colored Blue. Also, with high probability, more than 0.7 fraction of the paths connected to s have no member of V', and will thus all become Blue during the iterated propagation. This gives, with high probability, an advantage of more than $0.4 \cdot 3 \cdot n/60 = n/50$ to Blue among the non-isolated nodes, and by Chernoff in this case Blue has the majority with high probability.

Example 8. Iterative propagation can harm the adversary:
Consider a graph consisting of 10 stars, each of size $n/10$, where $\mu = \delta = 1/10$.

With limited propagation if V' contains exactly one center of a star and this node gets Blue, then with high probability, by Chernoff, Blue has the majority. This happens with probability bigger than $0.4/e > 0.1$ (see Figure 2).

With iterated propagation, Red gets, with high probability, a large majority in each star in which the center is not in V' and Blue, hence Red has the majority unless at least 5 centers belong to V' and are all Blue, and this happens with a much smaller probability (see Figure 3).

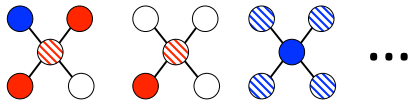

Figure 2: With limited propagation, the chance of a $Blue$ majority relies on $Blue$ being in a star center, propagating its color to the whole star, while the Red nodes of V' are not centered, limiting their propagation capacities (striped nodes indicate propagation).

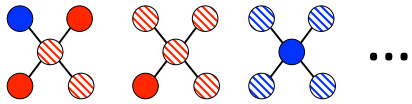

Figure 3: With iterative propagation, even non-center Red nodes have a significant impact, as much as star-centered $Blue$ nodes (striped nodes indicate propagation).

5 Discussion

In this paper we approach the issue of robustness – whether a social network is vulnerable to an adversarial attempt to propagate information through it. We found bounds that ensure a majority of network agents will not, with high probability, be duped by an adversary trying to manipulate agents. For expander graphs and random graphs, we even found such a limit for a strong adversary, which can decide which specific agents to deceive. Furthermore, we show that neither limited nor iterative propagation methods have deterministic influence on the capabilities of passive or weak adversaries.

This line of research opens various new questions and directions of work: we use a simple node majority as enabling an observer to attempt to find what the real truth is, yet one can imbue this agent with various capabilities, including some limited knowledge of the topological properties of the graph. Moreover, one can choose hybrid variants of adversaries (or a multitude of them), and examine how this affects the bounds and results we have shown.

Another natural extension deals with examining more particular sorts of graphs, perhaps relying on data on common graph structures in various social communities. As we have shown, assuming expander graphs or random ones allows us to show a bound for strong adversaries and, naturally, different types of graphs may be more or less robust to various sorts of adversaries and manipulations.

Acknowledgements

Moshe Tennenholtz carried out this work while at Microsoft Research, Israel. This work was partially supported by BSF grant 2012/107, by ISF grant 620/13, and by the Israeli I-Core program; by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement number 337122 and by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 274919.

References

[Alon and Spencer, 2008] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, 3rd edition, 2008.

[Alon *et al.*, 2012] Noga Alon, Moshe Babaioff, Ron Karidi, Ron Lavi, and Moshe Tennenholtz. Sequential voting with externalities: herding in social networks. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, page 36, Valencia, Spain, June 2012.

[Andersen *et al.*, 2008] Reid Andersen, Christian Borgs, Jennifer Chayes, Uriel Feige, Abraham Flaxman, Adam Kalai, Vahab Mirrokni, and Moshe Tennenholtz. Trust-based recommendation systems: an axiomatic approach. In *Proceedings of the 17th international conference on World Wide Web (WWW)*, pages 199–208, Beijing, China, April 2008.

[Arthur, 1989] W. Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394):116–131, March 1989.

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

[Bikhchandani *et al.*, 1992] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, October 1992.

[Blume, 1993] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424, June 1993.

[Brown and Reingen, 1987] Jacqueline Johnson Brown and Peter H. Reingen. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14(3):350–362, December 1987.

[Domingos and Richardson, 2001] Pedro Domingos and Matthew Richardson. Mining the network value of customers. In *International conference on Knowledge discovery and data mining (KDD)*, pages 57–66, San Francisco, California, August 2001.

[Ellison, 1993] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, September 1993.

[Farrell and Saloner, 1985] Joseph Farrell and Garth Saloner. Standartization, compatability, and innovation. *Rand Journal of Economics*, 16(1):70–83, 1985.

[Feldman *et al.*, 2014] Michal Feldman, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. Reaching consensus via non-bayesian asynchronous learning in social networks. In *Proceedings of the 17th. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, 2014.

[Goldenberg *et al.*, 2001] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.

[Granovetter, 1973] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):136–1380, May 1973.

[Kameda *et al.*, 1997] Tatsuya Kameda, Yohsuke Ohtsubo, and Masanori Takezawa. Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. *Journal of Personality and Social Psychology*, 73(2):296–309, 1997.

[Katz and Shapiro, 1985] Michael L. Katz and Carl Shapiro. Network externalities, competition, and compatibility. *The American Economic Review*, 75(3):424–440, June 1985.

[Kempe *et al.*, 2003] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *International conference on Knowledge discovery and data mining (KDD)*, pages 137–146, Washington D.C., August 2003.

[López-Pintado and Watts, 2008] Dunia López-Pintado and Duncan J. Watts. Social influence, binary decisions and collective dynamics. *Rationality and Society*, 20(4):399–443, November 2008.

[Morris, 2000] Stephen Morris. Contagion. *The Review of Economic Studies*, 67(1):57–78, January 2000.

[Mossel *et al.*, 2014] Elchanan Mossel, Joe Neeman, and Omer Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28(3):408–429, May 2014.

[Richardson and Domingos, 2002] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *International conference on Knowledge discovery and data mining (KDD)*, pages 61–70, Edmunton, Canada, July 2002.

[Rogers, 2003] Everett M. Rogers. *Diffusion of innovations*. Simon and Schuster, 5th edition, 2003.

[Tangand *et al.*, 2009] Jie Tangand, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *International conference on Knowledge discovery and data mining (KDD)*, pages 807–816, Paris, France, June 2009.

[Vijay Mahajan, 1990] Frank M. Bass Vijay Mahajan, Eitan Muller. New product diffusion models in marketing: A review and directions for research. *The Journal of Marketing*, 54(1):1–26, January 1990.

[Watts, 2002] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9):5766–5771, April 2002.

[Young, 2000] H. Peyton Young. The diffusion of innovations in social networks. Economic Working Paper 437, John Hopkins University, May 2000.

[Young, 2009] H. Peyton Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5):1899–1924, 2009.