Why Rumors Spread Fast in Social Networks, and How to Stop It*

Ahad N. Zehmakan[1] , **Charlotte Out**[2] , **Sajjad Hesamipour Khelejan**[3]
[1] School of Computing, The Australian National University
[2] Department of Computer Science & Technology, University of Cambridge
[3] School of Computer Science & Statistics, Trinity College Dublin
ahadn.zehmakan@anu.edu.au, ceo33@cam.ac.uk, hesamips@tcd.ie

Abstract

We study a rumor spreading model where individuals are connected via a network structure. Initially, only a small subset of the individuals are spreading a rumor. Each individual who is connected to a spreader, starts spreading the rumor with some probability as a function of their trust in the spreader, quantified by the Jaccard similarity index.Furthermore, the probability that a spreader diffuses the rumor decreases over time until they fully lose their interest and stop spreading.

We focus on determining the graph parameters which govern the magnitude and pace that the rumor spreads in this model. We prove that for the rumor to spread to a sizable fraction of the individuals, the network needs to enjoy "strong" expansion properties and most nodes should be in "well-connected" communities. Both of these characteristics are, arguably, present in real-world social networks up to a certain degree, shedding light on the driving force behind the extremely fast spread of rumors in social networks.

Furthermore, we formulate a large range of countermeasures to cease the spread of a rumor. We introduce four fundamental criteria which a countermeasure ideally should possess. We evaluate all the proposed countermeasures by conducting experiments on real-world social networks such as Facebook and Twitter. We conclude that our novel decentralized countermeasures (which are executed by the individuals) generally outperform the previously studied centralized ones (which need to be imposed by a third entity such as the government).

1 Introduction

With the rapid development of the Internet, social media has become a convenient online platform for users to obtain information, express and exchange opinions and stay in touch with friends. However, online social networks also pave the road for the propagation of misinformation, particularly rumors (commonly defined as unverified information or deliberately falsified news). It is usually difficult for the public to recognize the falsehood of a rumor, especially if it is designed skillfully, cf. [Vosoughi *et al.*, 2018]. The spread of rumors can mislead people to behave in irrational ways, which can cause a series of undesirable consequences, such as public panic, virtual assets losses, manipulation of the outcome of political events, and economic damages. Consequently, there has been a growing demand and interest to gain insights into the rumor spreading dynamics and design powerful countermeasures to reduce the threats posed by rumors.

To shed some light on the fundamental characteristics and essential principles of rumor propagation phenomenon, scholars from a vast spectrum of backgrounds have introduced and studied various rumor spreading models, such as SIR model [Zhao *et al.*, 2012], Push-Pull protocol [Giakkoupis, 2011], DK model [Daley and Kendall, 1965] and the Independent Cascade (IC) model [Kempe *et al.*, 2003]. In most of these models, the interactions and influence between the individuals are modelled using a graph structure, which represents a social network (SN). The state of the individuals (e.g., informed/uninformed) is updated following a rumor spreading rule. The updating rules are tailored to capture various properties observed in real-world scenarios, usually conceptualized by social scientists.

One aspect of the rumor spreading dynamics which has attracted a substantial amount of attention is the design of effective countermeasures to stop or slow down the spread of rumors, e.g., blocking users, blocking connections, and spreading an "anti-rumor", cf. [He *et al.*, 2015].

In the present paper, we introduce a rumor spreading model which inherits characteristics of the IC, Push-Pull, and SIR model and additionally captures the well-established sociological concepts of the impact of trust and forgetting mechanism. In an attempt at fighting rumor spreading, we study six countermeasures. We establish four essential criteria that a good countermeasure should possess and evaluate the proposed countermeasures on those. We observe that the decentralized countermeasures perform generally better.

1.1 Preliminaries

Let $G = (V, E)$ be a simple connected undirected graph, where $n := |V|$ and $m := |E|$. For a node $v \in V$,

*All the eliminated proofs are given in the full version: https://arxiv.org/abs/2305.08558

$N(v) := \{v' \in V : \{v', v\} \in E\}$ is the *neighborhood* of v. Furthermore, $\hat{N}(v) := N(v) \cup \{v\}$ is the *closed neighborhood* of v. Let $d(v) := |N(v)|$ be the *degree* of v in G. We also define $d_A(v) := |N(v) \cap A|$ for a set $A \subseteq V$. Furthermore, let $\partial(A) := \{v \in V \setminus A : \{v', v\} \in E, v' \in A\}$ be the *node boundary* of $A \subset V$.

We define a *coloring* to be a function $\mathcal{C} : V \to \{r, u, o\}$, where r, u, and o represent *red*, *uncolored*, and *orange* respectively. For a node $v \in V$, the set $N_a^{\mathcal{C}}(v) := \{v' \in N(v) : \mathcal{C}(v') = a\}$ includes the neighbors of v which have color $a \in \{r, u, o\}$ in coloring $\mathcal{C}$.

Rumor Spreading Model. Consider an initial coloring $\mathcal{C}_0$ of a graph G. In each round, all nodes simultaneously update their color according to the following updating rule:

$$\mathcal{C}_t(v) = \begin{cases} o & \text{if } \mathcal{C}_{t-1}(v) = o \\ r & \text{if } \mathcal{C}_{t-1}(v) = r \text{ and } \mathcal{J}_t(v) < k \\ o & \text{if } \mathcal{C}_{t-1}(v) = r \text{ and } \mathcal{J}_t(v) = k \\ u & \text{if } \mathcal{C}_{t-1}(v) = u \text{ w.p. } p^*(v) \\ r & \text{if } \mathcal{C}_{t-1}(v) = u \text{ w.p. } 1 - p^*(v) \end{cases}$$

where $\mathcal{C}_t(v)$ is the color of node v in the t-th round, integer k is a model parameter, $\mathcal{J}_t(v)$ is the number of rounds v has been red until round t, $\mathcal{S}(v, v') := |\hat{N}(v) \cap \hat{N}(v')|/|N(v) \cup N(v')|$ for $v, v' \in V$, and $p^*(v) := \prod_{v' \in N_r^{\mathcal{C}_{t-1}}(v)} \left(1 - \frac{S(v,v')}{2^{\mathcal{J}_t(v')}}\right)$. A red node corresponds to an individual who is *informed* of the rumor. An informed node stops spreading the rumor after k rounds and turns orange (*uninterested*), which it remains forever. An uncolored node corresponds to an *uninformed* individual. If an uninformed (uncolored) node v is adjacent to an informed (red) node v', then v' turns v into red w.p. $S(v, v')/2^{\mathcal{J}_t(v')}$ independently. Thus, v becomes red in the next round w.p. $1-p^*$ and remains uncolored w.p. p^*. The coefficient $1/2^{\mathcal{J}(v')}$ corresponds to the probability that v' spreads the rumor and $\mathcal{S}(v, v')$ is the probability that v accepts it. The value of $1/2^{\mathcal{J}(v')}$ accounts for the fact that v' might not necessarily spread the rumor w.p. 1 and the probability decreases exponentially in the number of rounds v' has been informed of the rumor, reflecting the fact that an individual loses interest in a rumor over time, cf. [Zhao *et al.*, 2013]. The coefficient $\mathcal{S}(v, v')$ (the Jaccard index) which measures the similarity between two nodes reflects the fact that people are more likely to accept information from their trusted connections [Figeac and Favre, 2021]. In the numerator, we use $\hat{N}(v)$. This is to ensure that for two adjacent nodes v, u, $S(v, u)$ (the accepting probability) is not zero. We could analogously define $S(v, u) = (|N(v) \cap N(u)| + 2)/|N(v) \cup N(u)|$ since we are always concerned about adjacent nodes. We note that we can also view $S(v, v')$ as the weight of the edge $\{v, v'\}$.

Our model is different from the IC model in two ways: In the IC model (i) k is always set to 1 (i.e., a red node becomes orange after one round) (ii) the weights are usually assigned randomly.

Starting from any initial coloring, the process eventually reaches a *fixed* coloring where all nodes are orange or uncolored. If the process reaches a coloring with a constant fraction of orange nodes, say 10%, then we say that the rumor *spreads*, and it does not otherwise. (There is nothing unique about 10% and our results hold for similar fixed values.)

Graphs. Let $\mathcal{G}_{n,p}$ denote the Erdős-Rényi (ER) random graph, which is the random graph on n nodes, where each edge is present independently w.p. p. For integers n and r, we define the (n,r)-*flower* graph in the following way. Consider a cycle $C_N = v_1, \cdots, v_N$ for $N = n/r$. For each node v_i, add a distinct clique of size $r-1$ to the graph and add an edge between v_i and every node in the clique. We refer to each node v_i and its clique as a *super node* and v_i is called the *boundary* node of the super node. We are particularly interested in the case of $r = \log^2(n)$, which is simply called the n-*flower* graph. (Note that $(n, 1)$-flower graph is simply a cycle graph with n nodes.)

To measure the expansion of a graph, we consider an algebraic characterization of expansion. Let $\lambda(G)$ be the second-largest absolute eigenvalue of the adjacency matrix of G. Small values of $\lambda(G)$ imply that G has strong expansion properties (i.e., is well-connected). For integers n, d, we define the (n, d)-*moderate expander* graph in the following way, where we always assume that n is "significantly" larger than d. Let H be a N-node, D-regular graph such that $\lambda(H) \leq C\sqrt{D}$, where $N = \frac{n}{\log^2(n)}$ and $D = d \cdot \log^2(n)$ and C is a positive constant. Replace every node x in H with a clique of size $\log^2 n$ and then evenly distribute the D edges of x among these $\log^2 n$ nodes. The obtained n-node $(\log^2 n + d - 1)$-regular graph is an (n,d)-moderate expander, which is denoted by $\mathcal{M}_{n,d}$. Similar to the (n, r)-*flower* graph, the set of $\log^2 n$ nodes in each of the N cliques is called a *super node*. (Note that moderate expanders are not meant to mimic real-world SNs. They are solely designed to maximize the spread of rumors and are objects of theoretical interest.)

Experimental Setup. For our experiments, we rely on publicly available graph data from [Leskovec and Krevl, 2014]. Our experiments were conducted on the following SNs: Twitter (81306 nodes and 1342310 edges), Facebook (4039 nodes and 88234 edges), Google+ (107614 nodes and 13673453 edges), Twitch Germany (9498 nodes and 153138 edges), and Twitch France (6549 nodes and 122666 edges). We use shorthand TW, FB, G+, T-GE, T-FR, respectively. We also conducted experiments on ER random graph and Hyperbolic random graph (HRG). The parameters in these graphs were set such that the (expected) number of nodes/edges is comparable to the ones in the aforementioned real-world networks. For HRG, one also needs to provide the exponent of the power-law degree distribution β and the temperature T as the input parameters. We set $\beta = 2.5$ and $T = 0.6$. We used the algorithm of [Staudt *et al.*, 2016] for the generation of HRG random graphs. Furthermore, the experiments which required random choice of edges or colors were executed 100 times. and then the average output was considered. The code for the experiments is available at https://github.com/charlotteout/RumourSpreading.

Assumptions. All logarithms are to base e, unless pointed out otherwise. We let n tend to infinity and say an event $\mathcal{E}$ happens with high probability (w.h.p.) if it occurs w.p. $1 -$

$o(1)$. We always assume that initially one randomly chosen node is red, and all other nodes are uncolored, otherwise, it is stated explicitly. Furthermore, we suppose the parameter k is a small integer, say $k = 5$, but our results would hold for any constant value of k.

1.2 Our Contribution

We study a rumor spreading model which captures fundamental characteristics such as the randomized spreading mechanism and various agent types as introduced in the IC, Push-Pull and SIR model, as well as sociological concepts such as the impact of homophily on trust [Granovetter, 1973], formulated by the Jaccard index, and the forgetting mechanism [Zhao *et al.*, 2013].

Firstly, we address the question: What are the graph structures for which the rumor spreads (in other words, what graph parameters govern the spread of rumors)? It has previously been argued that that information disseminates quickly when the graph has strong expansion properties (i.e., is well-connected), cf. [Giakkoupis and Sauerwald, 2012]. However, for our model expansion is not solely sufficient for a rumor to spread, especially if the graph is sparse which is usually the case in the real-world SNs. In particular, we prove that in our model on the ER random graph $\mathcal{G}_{n,p}$ (which enjoys strong expansion properties, cf. [Le *et al.*, 2017]) for p sufficiently smaller than $1/\sqrt{n}$, the rumor does not spread with a constant probability.

Additionally, we show that an abundance of very well-connected local communities (which result in large values of $\mathcal{S}(v, v')$ for adjacent nodes v, v') alone also cannot guarantee extensive spread of rumors. In particular, we prove that on an (n, r)-flower graph, where $\mathcal{S}(v, v') = 1$ for almost every two adjacent nodes v, v', the rumor does not spread w.h.p. for $r \leq n^{1-\epsilon}$ and $\epsilon > 0$ (even when we start with $o(\log n)$ red nodes).

However, we show that the combination of these two properties guarantees an extremely fast spread of rumors. More precisely, we prove for even very sparse moderate expander graphs, the rumor spreads in logarithmically many rounds. Roughly speaking, the strong local communities help the rumor to spread quickly inside a community once it reaches a node in that community and expansion ensures that it breaks out into other communities invasively. (We emphasize that the average degree of moderate expanders in this set-up is in the order of $\log^2 n$, which is much smaller than the average degree of $\sqrt{n}$ required in ER graphs for spreading.)

A natural question to ask is whether the rumor spreads on real-world SNs in our model. Our experiments on real-world graph data such as Twitter and Facebook demonstrate that the rumor indeed spreads to a very large body of the network in a short period of time. While the social graphs which emerge in the real world do not have the expansion and community structure tailored for the moderate expanders, they still enjoy a certain level of expansion, and well-connected communities are present in abundance. Note that this is an indication that our model is more realistic than previous models such as Push-Pull models, which advocate strong expansion properties as necessary and sufficient condition for fast spread of rumors, as we know that in real life rumors spread very fast in real-world SNs, and they are not strong expanders.

Moreover, we formulate and investigate several countermeasures. Some of them (e.g., blocking nodes and edges) need to be implemented by a third entity such as the government, and we refer to them as centralized countermeasures. On the other hand, the decentralized ones are executed by the members of the network. It turns out that the proposed decentralized countermeasures not only enjoy several desirable criteria such as not interfering with freedom of expression and not being too intrusive, but also significantly outperform the centralized ones in stopping the spread of the rumor according to our experiments. The prior work has focused on the development of centralized countermeasures, see Section 1.3 (which are also implemented in practice up to some degree, e.g., by blocking accounts). Our work aspires to send out the message that the focus should be shifted towards the development of decentralized countermeasures, which can be achieved for instance through educating the members rather than forceful actions of a third entity.

1.3 Prior Work

A plethora of rumor spreading models have been developed and studied in recent years. Here, we focus on the most fundamental and relevant models, which have inspired our work.

Push-Pull Models. In this set-up, each node is either red or uncolored. In each round, every red node makes a randomly chosen neighbor red (Push model), or every uncolored node adopts the color of a randomly chosen neighbor (Pull model), or both (Push-Pull model). Since there is no forgetting mechanism in place, all nodes eventually become red (i.e., the rumor spreads). Thus, a natural question is how long this takes. For the Push model, the spreading time is known [Feige *et al.*, 1990] to be $\mathcal{O}(\Delta \cdot (\Lambda + \log(n))$, where Δ and Λ are the maximum degree and diameter of the underlying graph. For the Push-Pull model, after a long line of research, the bound $\mathcal{O}\left(\Phi^{-1} \log(n)\right)$, for Φ being the conductance of the graph, was proven [Giakkoupis and Sauerwald, 2012].

Independent Cascade (IC) Model. In the IC model [Goldenberg *et al.*, 2001], in each round every red node v makes an uncolored node u in its neighborhood red w.p. p_{vu}. A red node becomes orange after one round, which is similar to setting $k = 1$ in our model. However, in the IC model, the probabilities p_{vu} are chosen uniformly at random. Motivated by viral marketing, the main focus in this model is developing algorithms for finding subsets of nodes that maximize the spread of the red color, mostly exploiting monotonicity and submodularity properties (cf. [Mossel and Roch, 2007; Chen *et al.*, 2011]).

Weighted Connections. Recall that in the IC model (and other similar models) weights are assigned to the edges randomly. As this is not entirely realistic, it would be relevant to introduce meaningful weight assignment mechanisms. Using the communication information of individuals on various real-world networks, [Onnela *et al.*, 2007] and [Goyal *et al.*, 2010] observed that there is a strong correlation between the number of shared friends of two individuals and their level of communication. Consequently, they proposed the usage of

similarity measures, such as Jaccard-like parameters, to approximate the weights of connections between nodes. This is also aligned with the well-studied strength of weak ties hypothesis [Granovetter, 1973]. This line of research has inspired the choice of Jaccard index in our model.

Countermeasures. A large part of the research efforts for developing countermeasures is concentrated around blocking nodes and edges. However, since in most models finding the most "influential" nodes/edges is NP-hard, cf. [Kempe *et al.*, 2003], the focus has been on approximate blocking strategies, which use structural properties. For nodes, various algorithms such as blocking nodes with the highest degree, betweenness, and closeness have been investigated, cf. [He *et al.*, 2015; Wang *et al.*, 2015; Yu *et al.*, 2008]. Furthermore, for different greedy-based edge blocking strategies to minimize the spread in the IC model, see [Kimura *et al.*, 2008; Yan *et al.*, 2019]. Other studied countermeasures are spreading the truth as an anti-rumor, cf. [Tripathy *et al.*, 2010; Ding *et al.*, 2020], inoculation strategies (which rest on the idea that if people are forewarned that they might be misinformed, they become more immune), cf. [Lewandowsky and Van Der Linden, 2021], and accuracy flags, cf. [Gausen *et al.*, 2021]. For more results on countermeasures also see [Coro *et al.*, 2020; Bredereck *et al.*, 2021; Zheng *et al.*, 2022; Qian *et al.*, 2018; Ma *et al.*, 2016].

2 When Does a Rumor Spread?

2.1 Erdős-Rényi Random Graph

Theorem 1. *Consider the coloring where only one node is red (the rest is uncolored) on $\mathcal{G}_{n,p}$ with $p \leq \frac{1}{n^{\frac{1}{2}+\epsilon}}$ for any constant $\epsilon > 0$. The rumor does not spread with a constant probability.*

Proof. Define $s := \lceil 1/\epsilon \rceil + 1$. For a pair of distinct nodes v and u, the probability that the inequality $|N(v) \cap N(u)| \geq s$ holds is upper-bounded by $\binom{n-2}{s} p^{2s}$. Let X be the number of pairs which satisfy the above inequality. Then, we have $\mathbb{E}[X] \leq \binom{n}{2} \binom{n-2}{s} p^{2s} \leq n^{s+2} p^{2s} \leq \frac{n^{s+2}}{n^{s+2s\epsilon}} = o(1)$, where we used that $p \leq 1/n^{\frac{1}{2}+\epsilon}$ and $s\epsilon > 1$, respectively. Hence, by Markov's inequality (cf. [Dubhashi and Panconesi, 2009]), $\Pr[\mathcal{A}] = \Pr[X \geq 1] \leq o(1)$, where $\mathcal{A}$ is the event that $X \geq 1$ (and $\bar{\mathcal{A}}$ is the complement of $\mathcal{A}$).

Let v be the only node which is colored red in $\mathcal{C}_0$. For each node $u \in N(v)$, we have $\Pr[\mathcal{C}_1(u) = r | d(v) = d \wedge \bar{\mathcal{A}}] = \frac{|\hat{N}(v) \cap \hat{N}(u)|}{2|N(v) \cup N(u)|} \leq \min\left(\frac{s+2}{2d}, \frac{1}{2}\right)$. For $(s+2)/(2d)$, we used that $|\hat{N}(v) \cap \hat{N}(u)| \leq |N(v) \cap N(u)| + 2 \leq s + 2$ and $|N(v) \cup N(u)| \geq d(v) = d$. The upper bound of $1/2$ holds because $|\hat{N}(v) \cap \hat{N}(u)| \leq |N(v) \cup N(u)|$.

Let $\mathcal{E}_i$, for $1 \leq i \leq k$, denote the event that v does not make any of its neighbors red in the i-th round. Then, $\Pr[\mathcal{E}_1 | d(v) = d \wedge \bar{\mathcal{A}}] \geq \left(1 - \min\left(\frac{s+2}{2d}, \frac{1}{2}\right)\right)^d$. If $\frac{s+2}{2d} < 1/2$ then $\left(1 - \frac{s+2}{2d}\right)^d \geq \left(\frac{1}{4}\right)^{(s+2)/2}$ (which gives a constant lower bound) using the estimate $(1-x) \geq \left(\frac{1}{4}\right)^x$ for $x < 1/2$. If $\frac{s+2}{2d} \geq 1/2$, then $d \leq s + 2$, which implies that $(1/2)^d$ is a constant. Therefore, in both cases, we can lower bound $\left(1 - \min\left(\frac{s+2}{2d}, \frac{1}{2}\right)\right)^d$ with some constant $C > 0$.

$$\Pr[\mathcal{E}_1] = \Pr[\bar{\mathcal{A}}] \cdot \Pr[\mathcal{E}_1 | \bar{\mathcal{A}}] + \Pr[\mathcal{A}] \cdot \Pr[\mathcal{E}_1 | \mathcal{A}] \geq$$
$$\Pr[\bar{\mathcal{A}}] \cdot \sum_{d=0}^{n-1} \Pr[\mathcal{E}_1 | d(v) = d \wedge \bar{\mathcal{A}}] \cdot \Pr[d(v) = d] \geq$$
$$\Pr[\bar{\mathcal{A}}] \cdot \sum_{d=0}^{n-1} C \cdot \Pr[d(v) = d] = (1 - o(1)) \cdot C \geq \frac{C}{2}.$$

With a similar argument, we can prove that $\Pr[\mathcal{E}_i | \mathcal{E}_{i-1} \wedge \cdots \wedge \mathcal{E}_1] \geq C/2$ for $2 \leq i \leq k$. Thus, we have $\Pr[\mathcal{E}_1 \wedge \cdots \wedge \mathcal{E}_k] = \Pr[\mathcal{E}_k | \mathcal{E}_{k-1} \wedge \cdots \wedge \mathcal{E}_1] \cdots \Pr[\mathcal{E}_2 | \mathcal{E}_1] \cdot \Pr[\mathcal{E}_1] \geq (C/2)^k$. This implies that w.p. at least $(C/2)^k = (C/2)^5$, no node becomes red during the first k rounds. In that case, the process ends with one orange node and $n-1$ uncolored nodes in k rounds. This bound on p turns out to be tight (a full proof is given in the extended version). □

2.2 Flower Graph

A super node whose all nodes are uncolored is called *uncolored* and *colored* otherwise. And it is said to be *red* if all its nodes are red.

Theorem 2. *Consider an (n,r)-flower graph for $r \leq n^{1-\epsilon}$ and constant $\epsilon > 0$. If initially there are $s(n) = o(\log n)$ red super nodes (and the rest is uncolored), the rumor does not spread w.h.p.*

Proof Sketch. A path of super nodes is a sequence of super nodes which form a path in the cycle obtained from collapsing each super node into a node. A path is *uncolored* if all its super nodes are uncolored. In a *reddish path*, there are no two adjacent uncolored super nodes and the endpoints are colored. We note that for any coloring of the (n,r)-flower graph, there is a set of maximal uncolored and reddish paths which partition the nodes in the graph.

Define a *phase* to be a sequence of k rounds. Let $\mathcal{C}$ be the coloring at the beginning of phase i. Consider all the endpoints of the uncolored paths in the aforementioned partitioning and define U to be their boundary nodes. Let $\mathcal{E}_i$ be the event that no node in U becomes red during the whole phase.

We observe that if the event $\mathcal{E}_i$ occurs, then all boundary nodes of the reddish paths endpoints become orange. Thus, all nodes which are not on any reddish path remain uncolored forever. Let us define $t^* := (1/C)^{2s(n)} \log(n)$, for a suitably chosen constant $0 < C < 1$, then with some relatively straightforward calculations, we can show that $\Pr[\wedge_{i=1}^{t^*} \bar{\mathcal{E}}_i] \leq \frac{1}{n}$. Thus, w.h.p. after at most t^* phases (i.e., kt^* rounds), we reach a coloring where all nodes which are not on any reddish path remain uncolored forever. Furthermore, we claim that the number of nodes on the reddish paths during the first kt^* rounds is sub-linear. Hence, the rumor does not spread w.h.p. (A full proof is given in the extended version.) □

2.3 Moderate Expander

Theorem 3. *Consider an (n, d)-moderate expander $\mathcal{M}_{n,d}$ with $d = \omega(1)$. If initially there is a red node (and the rest*

are uncolored), then the rumor spreads w.h.p. in $\mathcal{O}(\log_d n)$ rounds.

Similar to a flower graph, we call a super node x *uncolored* if all its nodes are uncolored. We say x is *strong red* if every node in it has become red at most three rounds before. A super node is *weak red* if it is neither strong red nor uncolored. Let u_t, s_t and w_t denote the number of uncolored, strong red, and weak red super nodes in the t-th round.

Recall that if we contract all $N = n/\log^2(n)$ super nodes in $\mathcal{M}_{n,d}$, we obtain a D-regular graph for $D = d\log^2(n)$. In Lemma 1 (whose proof is given in the extended version), we state that if a node in one of these super nodes is red, then the super node becomes red in 2 rounds. Then, in Lemma 3, we show that the number of strong red super nodes increases by roughly a d factor after every three rounds. Repeated application of Lemma 3 implies that the rumor spreads in $\mathcal{O}(\log_d n)$ rounds. (A more detailed discussion is given in the extended version, where we also argue that the bound $d = \omega(1)$ is necessary, i.e., the statement does not hold for constant d).

Lemma 1. *Consider a graph $G = (V, E)$ where nodes in $\mathcal{K} \subseteq V$ form a clique, $\kappa := |\mathcal{K}| \geq \log^2 n$, and for every $v \in \mathcal{K}$ $d(v) \leq 2\kappa$. If $\mathcal{C}_t(v) = r$ for some $v \in \mathcal{K}$ and all other nodes in $\mathcal{K}$ are uncolored, then there is no uncolored node in $\mathcal{K}$ in round $t + 2$ w.p. $1 - o(1/n)$.*

To prove Lemma 3, we need Lemma 2 and Observation 1. The proof of Lemma 2 is given in the extended version, which relies on the expander mixing lemma, cf. [Friedman, 2003].

Lemma 2. *Consider an N-node D-regular graph G, where $\lambda \leq C\sqrt{D}$, for some constant $C > 0$, and $D = \omega(1)$. If a node set A is of size at most $\frac{N}{10}$, then there is some constant $C' > 0$ such that $|\partial(A)| \geq \min(2N/5, |A|C'D)$.*

Observation 1. *Let x and y be two distinct super nodes in a moderate expander graph. Then, there is at most one edge between x and y, by construction.*

Lemma 3. *Consider an (n, d)-moderate expander $\mathcal{M}_{n,d}$ with $d = \omega(1)$. If $1 \leq s_t < C_1 N/D$, for a sufficiently small constant $C_1 > 0$, and $w_t = \mathcal{O}(s_t/d)$, then after three rounds there are $\Omega(s_t d)$ new strong red super nodes w.p. $1 - \exp(\Omega(-ds_t)) - o(1/\log n)$.*

Proof. Let $\mathcal{E}^*$ be the event that every uncolored super node becomes strong red in two rounds once it has at least one red node. Based on Lemma 1, $\mathcal{E}^*$ holds w.p. at least $1 - N \cdot o(1/n) \geq 1 - o(1/\log n)$ since there are N super nodes.

Furthermore, let q^* denote the probability that a node v, in a strong red super node, makes a node u, in an uncolored super node, red where there is an edge between v and u. Since $|\hat{N}(v) \cap \hat{N}(u)| \geq 2$, $|N(v) \cup N(u)| \leq 2(d + \log^2 n) \leq 2.5\log^2 n$ (using the assumption that d is significantly smaller than n), and v has been red for at most three rounds, we get the following upper-bound:

$$q^* \geq \frac{|\hat{N}(v) \cap \hat{N}(u)|}{2^3 |N(v) \cup N(u)|} \geq \frac{2}{8 \times 2.5 \log^2 n} = \frac{1}{10 \log^2 n}. \tag{1}$$

Let S, W, and U be the set of strong red, weak red, and uncolored super nodes in round t. Let us label the nodes in $\partial(S) \cap U$ from u_1 to u_b, where b is the size of $\partial(S) \cap U$. For each node u_i consider one of its neighbors in S. Let Bernoulli random variable y_i be 1 if and only if u_i is made red by that neighbor in S in the next round (i.e., $t + 1$). For the random variable $Y := \sum_{i=1}^{b} y_i$, we have $\mathbb{E}[Y] \geq bq^* \geq b/(10\log^2 n)$, where we used $\Pr[y_i = 1] = q^*$ and Equation (1). Since y_i's are independent, applying Chernoff bound (cf. [Dubhashi and Panconesi, 2009]) yields

$$\Pr\left[Y \leq \frac{b}{20\log^2 n}\right] \leq \exp\left(-\Theta\left(\frac{b}{\log^2 n}\right)\right). \tag{2}$$

Note that $w_t = \mathcal{O}(s_t/d) = o(s_t)$ implies that $s_t + w_t \leq 1.1 s_t$. Furthermore, $1.1 s_t \leq N/10$ since $s_t \leq C_1 N/D = o(N)$. Thus, we can apply Lemma 2 for $A = S \cup W$ and the graph obtained from contracting each super node to a node. Since $|A| = s_t + w_t \leq 1.1 s_t \leq 1.1 C_1 N/D$, we get $2N/5 \geq |A|C'D$ by selecting C_1 to be sufficiently small. Thus, $|\partial(A)| \geq s_t C'D$. Furthermore, note that $|\partial(A)| = |\partial(S) \cap U| + |\partial(W) \cap U| = b + |\partial(W) \cap U|$ and $|\partial(W) \cap U| \leq w_t D$. Combining the last two statements gives $b \geq s_t C'D - w_t D$. Using $w_t = \mathcal{O}(s_t/d) = o(s_t)$ implies that $b = \Omega(Ds_t)$. Thus, Equation (2) implies that w.p. $1 - \exp(-\Omega(Ds_t/\log^2 n)) = 1 - \exp(-\Omega(ds_t))$, there will be $\Omega(Ds_t/\log^2 n) = \Omega(ds_t)$ nodes in U which become red in the next round. Note that all such nodes are in different super nodes (see Observation 1). If event $\mathcal{E}^*$ holds, then all such super nodes will be strong red in two more rounds. Since $\mathcal{E}^*$ holds w.p. $1 - o(1/\log n)$ (as discussed above), there will be $\Omega(ds_t)$ new strong red super nodes after three rounds w.p. $1 - \exp(\Omega(-ds_t)) - o(1/\log n)$. □

2.4 Experiments and Real-world Networks

The outcome of our experiments in Figure 1-(a) are consistent with our theoretical findings. In particular, the rumor does not spread in the flower graph and ER-low (i.e., $p = 1/(4\sqrt{n})$) while it does for the moderate expander and ER-high (i.e., $p = 4/\sqrt{n}$). Note that in this set-up, a node in the moderate expander is of degree $d + \log^2 n - 1 \approx 100$ (actually, we observe in the experiments that for $D = 64$ rather than $D = d \cdot 100 = 4 \cdot 100$ the rumor already spreads), which indicates the rumor spreads even in very sparse graphs if they possess some level of expansion and community structure. Furthermore, we observe that the process on the moderate expander ends in around 50 rounds, which indeed appears to be logarithmic rather than linear in $n = 16000$ (this is aligned with the bound $\mathcal{O}(\log_d n)$ proven in Theorem 3).

Figure 1-(b) depicts the extent to which the rumor spreads in Twitter and Facebook graph and random graph model HRG with comparable parameters. (Please refer to Section 1.1 for more details.) The plots for the other three studied SNs are given in the extended version. We observe that the rumor spreads to a large part of the graph very quickly. This can be explained by the fact that all these graphs have a decent level of expansion and community-like structure, which are the necessary properties for a fast and wide spread according to our theoretical results. As a by-product, our experiments also support that HRG is a suitable choice for modeling real-world SNs.

3 How to Stop the Rumor Spreading?

We present six countermeasures (the first four are inspired by prior work as explained in Section 1.3, but the last two are completely novel) and then compare them. The outcome of our experiments on the countermeasures for Twitter and Facebook graphs and moderate expander are given in Figure 1 and for the other three SNs (T-GE, T-FR, and G+) in the extended version.

CM1: Blocking Nodes. We assume that the 5% highest degree nodes and 20% randomly chosen nodes are blocked (i.e., do not receive/spread the rumor). As Figure 1-(c) demonstrates, this countermeasure is not very effective. We believe blocking nodes according to the highest betweenness, closeness, or eigencentrality (instead of highest degree) would not improve the countermeasure significantly since in real-world SNs there is a large overlap between the highest degree nodes and nodes chosen by the mentioned parameters due to certain properties such as the power-law degree distribution.

CM2: Blocking Edges. The graph is partitioned into communities using the Louvain algorithm [Blondel *et al.*, 2008]. In each round of the process, if the fraction of red nodes is above a global threshold τ_g, then we block all the edges which are on the boundary of the "spreader" communities. A community is a spreader if its fraction of red nodes is larger than a local threshold τ_c. The blocked edges remain blocked until the community is not a spreader anymore. (Both threshold are set to 0.05 in our set-up.) Figure 1-(d) demonstrates while this countermeasure slows down the spread, the rumor still spreads to a large part of the graph. It is worth to mention that around $20 - 30\%$ of edges were blocked during the process in our experiments. (Unlike other experiments, this was executed only 10 times due to its high computational cost.)

CM3: Accuracy Flags. Assume that every time a node is supposed to become red, it rejects the rumor with some *reject* probability p_r, and becomes orange directly. In practice, this countermeasure corresponds to for example accuracy flags in online social platforms, in which posts containing certain keywords (say hot controversial or polarizing topics) are automatically accompanied by a banner warning the user about the trustworthiness of the content. The outcome of our experiments for $p_r = 0.3$, depicted in Figure 1-(e), demonstrates that the rumor still continues to spread to a significant portion of the community.

CM4: Let's Spread the Truth. Let the *truth spreading* process be the same as the rumor spreading with the following two differences: (i) green and light green are used instead of red and orange, respectively (ii) the probability a node becomes green is one half of the probability of becoming red in the rumor spreading process (this is to account for the observation that rumors spread faster than facts, cf. [Vosoughi *et al.*, 2018]). After τ rounds into the rumor spreading process, we color an uncolored node green and the truth starts spreading simultaneously. (We assume that the rumor and truth spread only to uncolored nodes, that is, a red/orange node does not become green and vice versa.) The outcome of experiments, depicted in Figure 1-(f), indicates that this countermeasure cannot stop the rumor effectively even when $\tau = 4$ (which implies that there is a strong rumor detection algorithm in place) and the node which starts the truth is the node with the highest degree among the uncolored nodes. We depict the influence of the delay τ on the final fraction of orange nodes in a figure in the extended version.

CM5: Fact Checkers. Consider a set of *fact checker* nodes, who starts spreading the truth (i.e., anti-rumor) once exposed to the rumor, as the truth spreading process in CM4. These correspond to "good citizens" (e.g., credible news outlets or scientists on the topic) who are educated or incentivized to verify the received information and spread the truth if necessary. (In our experiments, we assume they include 10% of the network and are distributed randomly.) This has some similarities to CM4, but instead of starting the spread of the truth by implementing a green node in the graph (which needs to be executed by a third entity), the fact checkers become green and trigger the spread of the truth once contacted by a rumor spreader. Furthermore, the fact checker spread the truth more aggressively: (i) the forgetting parameter k is much larger for the fact checker nodes (say 20 rather than 5) (ii) fact checkers can make their red neighbors green as well (iii) the fact checkers are three times more active in spreading (you can think of each round as three sub-rounds, where all nodes (red/green) spread in the first sub-round while the green fact checker nodes continue to spread in the second and third sub-round too). Note that green nodes which are not fact checker behave as in the original truth spreading process. Our experiments (see Figure 1-(g)) demonstrate that this countermeasure is very effective.

CM6: Let's Hear It Twice. We require a node to hear a rumor from at least two of its neighbors before accepting and spreading it (i.e., becoming red), instead of once as in the original process. Figure 1-(h) demonstrates that this countermeasure is immensely effective, where in our experiments, initially two randomly chosen nodes are red. We formalize this observation in Theorem 4, whose proof is given in the extended version.

Theorem 4. *Consider the (n,d)-moderate expander $\mathcal{M}_{n,d}$ with $d \leq n^{\frac{1}{2}-\epsilon}$ for a constant $\epsilon > 0$. If initially two super nodes x and y, chosen uniformly at random, have red node(s) (and the rest is uncolored) and CM6 is in place, then w.h.p. the rumor does not spread.*

Comparison of Countermeasures. We consider four fundamental criteria that a good countermeasure should possess. To the best of our knowledge, this is the first attempt to formalize such a list of criteria.

C1: Effective. A good countermeasure substantially reduces the extent that a rumor spreads.

C2: Easy To Apply. An acceptable countermeasure should be feasible and easily executable. If implemented by the agents of the network, it should not require full knowledge of the whole network or the complete history of the process. If it is administrated by a third entity, such as the government, it should not postulate a perfect rumor detection strategy or running algorithms which are computationally very costly.

C3: Not Against Freedom of Expression. A countermeasure ideally should not take away the freedom of expression

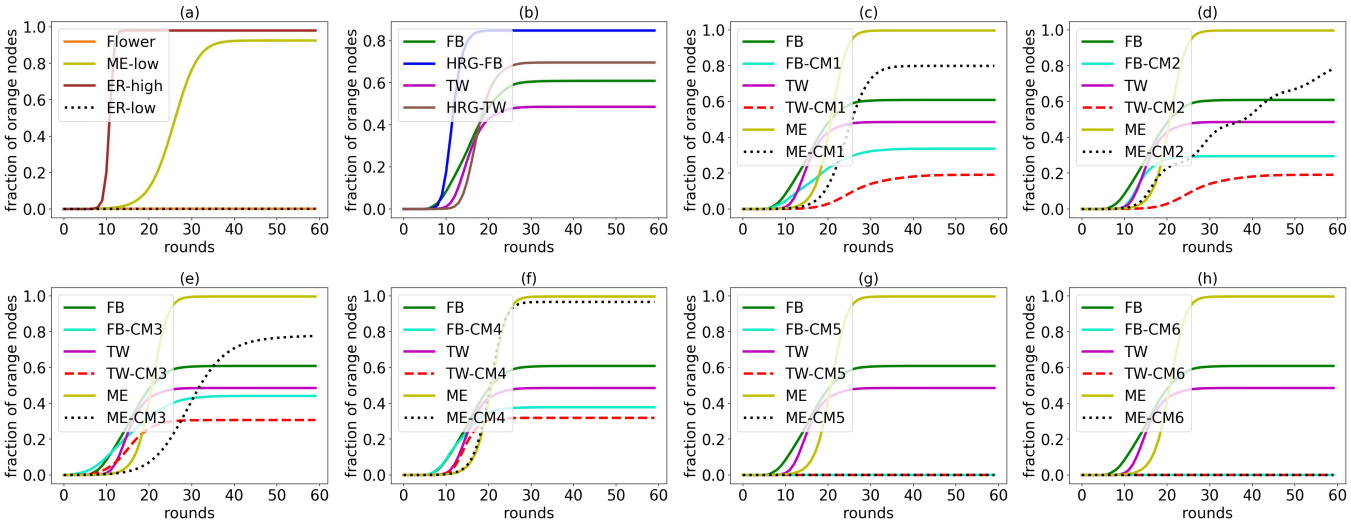

Figure 1: Fraction of orange nodes starting from one randomly chosen red node in (a) n-flower, (n, d)-moderate expander (ME-low) with $d = 4$ and $D = 64$ and super nodes of size 16, ER with $p = 4/\sqrt{n}$ (ER-high) and $p = 1/(4\sqrt{n})$ (ER-low), where $n = 16000$ (b) FB and TW graphs and HRG with comparable parameters (c-h) moderate expander (ME) for $n = 22000$ and $d = 12$ and TW and FB graphs before and after the implementation of countermeasures CM1 to CM6.

and liberties of the agents.

C4: Not Too Intrusive. A countermeasure which demands fundamental changes in the mechanism of information spreading or the network structure is not desirable.

	C1	C2	C3	C4	Decentralized
CM1	no	jein	no	no	no
CM2	no	jein	no	no	no
CM3	no	jein	yes	yes	no
CM4	no	jein	yes	yes	jein
CM5	yes	jein	yes	yes	yes
CM6	yes	yes	yes	yes	yes

Table 1: Determining which criteria are satisfied by each countermeasure, where "jein" means both yes and no.

Table 1 indicates which criteria each of the proposed countermeasure satisfies. Note that it is inherently difficult to measure the above criteria in a strict quantitative manner. Thus, the entries in the table are relative and up to interpretation. The choices for C1 are according to the results depicted in Figure 1. The entries for C2 are mostly set to jein since while they are not extremely difficult to implement, they need a smart rumor detection strategy or the full knowledge of the network. Furthermore, CM1 and CM2 violate C3 since they clearly intrude the freedom of expression and do not satisfy C4 since they change the network structure radically. The other countermeasures, arguably, satisfy the last two criteria. A more comprehensive discussion on the entries of Table 1 is provided in the extended version of the paper.

We say a countermeasure is *decentralized* if it is executed by the members of the network rather than being enforced by a third party such as the government or an online social platform management team. Summarizing the entries of Table 1 implies that, interestingly, the decentralized countermeasures, namely CM5 and CM6 (and CM4, up to some degree), satisfy most of the desired criteria while the centralized ones do not. Hence, instead of developing centralized countermeasures which need to be imposed by a forceful third entity, the focus should be devoted to the design and implementation of decentralized countermeasures which can be obtained through educating the members of the network. In short, educating is preferred over regulating.

4 Conclusion

We introduced a rich rumor spreading model and building on our theoretical and experimental findings, we argued that the abundance of community structures and good expansion properties are two of the main driving forces behind the spread of rumors. A potential avenue for future research is to determine other graph parameters which govern the spread of rumors. We also investigated several countermeasures. We observed that the decentralized countermeasures (which do not require a direct and forceful interference of a third entity but rather the education of the network's members) outperform the centralized ones vigorously. Therefore, a natural suggestion for the future studies is the shift of focus from centralized countermeasures to decentralized ones, which have been scarcely investigated by the prior work.

References

[Blondel *et al.*, 2008] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[Bredereck *et al.*, 2021] Robert Bredereck, Lilian Jacobs, and Leon Kellerhals. Maximizing the spread of an opinion in few steps: opinion diffusion in non-binary networks. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 1622–1628, 2021.

[Chen *et al.*, 2011] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *Proceedings of the 2011 siam international conference on data mining*, pages 379–390. SIAM, 2011.

[Coro *et al.*, 2020] Federico Coro, Emilio Cruciani, Gianlorenzo D'Angelo, and Stefano Ponziani. Exploiting social influence to control elections based on scoring rules. In *Proceedings of the twenty-Eighth international conference on international joint conferences on artificial intelligence (IJCAI)*, 2020.

[Daley and Kendall, 1965] Daryl J Daley and David G Kendall. Stochastic rumours. *IMA Journal of Applied Mathematics*, 1(1):42–55, 1965.

[Ding *et al.*, 2020] Li Ding, Ping Hu, Zhi-Hong Guan, and Tao Li. An efficient hybrid control strategy for restraining rumor spreading. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(11):6779–6791, 2020.

[Dubhashi and Panconesi, 2009] Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[Feige *et al.*, 1990] Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Randomized broadcast in networks. *Random Structures & Algorithms*, 1(4):447–460, 1990.

[Figeac and Favre, 2021] Julien Figeac and Guillaume Favre. How behavioral homophily on social media influences the perception of tie-strengthening within young adults' personal networks. *New Media & Society*, page 14614448211020691, 2021.

[Friedman, 2003] Joel Friedman. A proof of alon's second eigenvalue conjecture. *STOC*, pages 720–724, 2003.

[Gausen *et al.*, 2021] Anna Gausen, Wayne Luk, and Ce Guo. Can we stop fake news? using agent-based modelling to evaluate countermeasures for misinformation on social media. *Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media*, 2021.

[Giakkoupis and Sauerwald, 2012] George Giakkoupis and Thomas Sauerwald. Rumor spreading and vertex expansion. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, page 1623–1641, USA, 2012. Society for Industrial and Applied Mathematics.

[Giakkoupis, 2011] George Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *Symposium on Theoretical Aspects of Computer Science (STACS2011)*, volume 9, pages 57–68, 2011.

[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.

[Goyal *et al.*, 2010] Amit Goyal, Francesco Bonchi, and Laks VS Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 241–250, 2010.

[Granovetter, 1973] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.

[He *et al.*, 2015] Zaobo He, Zhipeng Cai, and Xiaoming Wang. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In *2015 IEEE 35Th international conference on distributed computing systems*, pages 205–214. IEEE, 2015.

[Kempe *et al.*, 2003] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[Kimura *et al.*, 2008] Masahiro Kimura, Kazumi Saito, and Hiroshi Motoda. Minimizing the spread of contamination by blocking links in a network. In *AAAI*, volume 8, pages 1175–1180, 2008.

[Le *et al.*, 2017] Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.

[Leskovec and Krevl, 2014] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

[Lewandowsky and Van Der Linden, 2021] Stephan Lewandowsky and Sander Van Der Linden. Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2):348–384, 2021.

[Ma *et al.*, 2016] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 3818–3824. AAAI Press, 2016.

[Mossel and Roch, 2007] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 128–134, 2007.

[Onnela *et al.*, 2007] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie

strengths in mobile communication networks. *Proceedings of the national academy of sciences*, 104(18):7332–7336, 2007.

[Qian *et al.*, 2018] Feng Qian, Chengyue Gong, Karishma Sharma, and Yan Liu. Neural user response generator: Fake news detection with collective user intelligence. In *Proceedings of the twenty-seventh international conference on international joint conferences on artificial intelligence (IJCAI)*, volume 18, pages 3834–3840, 2018.

[Staudt *et al.*, 2016] Christian L Staudt, Aleksejs Sazonovs, and Henning Meyerhenke. Networkit: A tool suite for large-scale complex network analysis. *Network Science*, 4(4):508–530, 2016.

[Tripathy *et al.*, 2010] Rudra M Tripathy, Amitabha Bagchi, and Sameep Mehta. A study of rumor control strategies on social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1817–1820, 2010.

[Vosoughi *et al.*, 2018] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *science*, 359(6380):1146–1151, 2018.

[Wang *et al.*, 2015] Zhefeng Wang, Enhong Chen, Qi Liu, Yu Yang, Yong Ge, and Biao Chang. Maximizing the coverage of information propagation in social networks. In *Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[Yan *et al.*, 2019] Ruidong Yan, Yi Li, Weili Wu, Deying Li, and Yongcai Wang. Rumor blocking through online link deletion on social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(2):1–26, 2019.

[Yu *et al.*, 2008] Yintao Yu, Tanya Y Berger-Wolf, Jared Saia, et al. Finding spread blockers in dynamic networks. In *International Workshop on Social Network Mining and Analysis*, pages 55–76. Springer, 2008.

[Zhao *et al.*, 2012] Laijun Zhao, Jiajia Wang, Yucheng Chen, Qin Wang, Jingjing Cheng, and Hongxin Cui. Sihr rumor spreading model in social networks. *Physica A: Statistical Mechanics and its Applications*, 391(7):2444–2453, 2012.

[Zhao *et al.*, 2013] Laijun Zhao, Wanlin Xie, H Oliver Gao, Xiaoyan Qiu, Xiaoli Wang, and Shuhai Zhang. A rumor spreading model with variable forgetting rate. *Physica A: Statistical Mechanics and its Applications*, 392(23):6146–6154, 2013.

[Zheng *et al.*, 2022] Jiaqi Zheng, Xi Zhang, Sanchuan Guo, Quan Wang, Wenyu Zang, and Yongdong Zhang. Rumor detection on social media with graph structured adversarial learning. In *Proceedings of the thirty-first international conference on international joint conferences on artificial intelligence (IJCAI)*, 2022.