Privacy-Conscious Information Diffusion in Social Networks

George Giakkoupis, Rachid Guerraoui, Arnaud Jégou, Anne-Marie Kermarrec, Nupur Mittal

Privacy-Conscious Information Diffusion in Social Networks

George Giakkoupis[1], Rachid Guerraoui[2], Arnaud Jégou[1], Anne-Marie Kermarrec[1], and Nupur Mittal[1]

[1] INRIA, Rennes, France
{george.giakkoupis, arnaud.jegou, anne-marie.kermarrec, nupur.mittal}@inria.fr
[2] EPFL, Lausanne, Switzerland
rachid.guerraoui@eplf.ch

Abstract. We present RIPOSTE, a distributed algorithm for disseminating information (ideas, news, opinions, or trends) in a social network. RIPOSTE ensures that information spreads widely if and only if a large fraction of users find it interesting, and this is done in a "privacy-conscious" manner, namely without revealing the opinion of any individual user. Whenever an information item is received by a user, RIPOSTE decides to either forward the item to all the user's neighbors, or not to forward it to anyone. The decision is randomized and is based on the user's (private) opinion on the item, as well as on an upper bound s on the number of user's neighbors that have not received the item yet. In short, if the user likes the item, RIPOSTE forwards it with probability slightly larger than $1/s$, and if not, the item is forwarded with probability slightly smaller than $1/s$. Using a comparison to branching processes, we show for a general family of random directed graphs with arbitrary out-degree sequences, that if the information item appeals to a sufficiently large (constant) fraction of users, then the item spreads to a constant fraction of the network; while if fewer users like it, the dissemination process dies out quickly. In addition, we provide extensive experimental evaluation of RIPOSTE on topologies taken from online social networks, including Twitter and Facebook.

1 Introduction

Social networking websites have become an important medium for communicating and disseminating news, ideas, political opinions, trends, and behaviors. Such online networks typically provide a *reposting* functionality, e.g., *sharing* in Facebook or *retweeting* in Twitter, which allows users to share other's posts with their own friends and followers. As information is reposted from user to user, large cascades of reposts can develop, and an information item can potentially reach a large number of people, much larger than the number of users exposed to the information initially (e.g., the users who witness a news event, or learned about it from some local media). Since people tend to propagate information which they find interesting and worth sharing (rather than random content) [21], an information item may spread widely only if sufficiently many users find it interesting. Ideally, the opposite direction should also hold: content that a

sufficiently large fraction of users would find interesting and would propagate (if they knew about it), should be likely to spread widely. This is, however, not always the case.

In countries with authoritarian regimes, users may not propagate anti-government ideas for the fear of being prosecuted. There are in fact several examples of political activists (and others) that have been convicted for posting or just reposting anti-government opinions on social media [24,26]. But even in democratic regimes, users may refrain from openly supporting their opinion on certain sensitive issues, from politics and religion to sexuality and criminal activity. For example, a user may not propagate a post supporting recreational drug use for the fear that it may have a negative impact on his career—as it is a common practice of employers to use social media for screening prospective employees [14]. Or more generally, users may refrain from reposting a (political or other) opinion when they believe it is not widely shared by their cycle—a well known principle in sociology known as the "spiral of silence" [23]. In all these cases, the dissemination of an idea in the social network is impeded by privacy considerations; even if many users support the idea, they may choose not to contribute to its propagation because they do not wish to reveal their own opinion (as reposting the idea would suggest the user is in favor of it).

Our Contribution: Privacy-Conscious Diffusion. We investigate a dissemination algorithm that has, roughly speaking, the following properties: (1) information that a sufficiently large fraction of the population finds interesting is likely to spread widely; (2) information that not sufficiently many people find interesting does not spread far beyond the set of users exposed to it initially; and (3) by observing the spreading process (in particular, the users' reposts), one cannot determine (with sufficient confidence) the opinion of any single user on the information that is disseminated.

More specifically, we propose the following simple, local dissemination algorithm, which we call RIPOSTE. Let G denote the (directed) graph modeling the social network, and n be the total number of users, and suppose that some (small) initial set of users learn an information item t. For each user u that learns t, RIPOSTE decides to either repost t, to all u's outgoing neighbors in G, or to not repost t, to anyone. The decision is randomized and depends on the user's (private) opinion on the information, and the number of the user's neighbors that have not received the information yet. Precisely, if u likes t, then t is reposted with probability λ/s_u, and if u does not like t, then t is reposted with a (smaller) probability δ/s_u, where $0 < \delta < 1 < \lambda$ are global parameters of the dissemination mechanism, and s_u is an *upper bound* on the number of u's outgoing neighbors that have not received t yet. If the algorithm cannot have access to information about whether u's neighbors already know the information, then the *total* number of u's outgoing neighbors can be used as the upper bound s_u.[3]

We argue that RIPOSTE achieves the property of *plausible deniability*: A user u can claim that, with reasonable probability, the act of reposting (or not) some information, does not reflect u's truthful opinion on the information, and is a result of the random-

[3] RIPOSTE can be viewed as a set of distributed pieces of software running at each user's machine connected to the social network. It is not the user who has the control of whether the item will be eventually reposted or not but this piece of software. It solicits the user's opinion on the item, and then flips a coin to determine whether the information will be reposted.

ness in the decision mechanism. Intuitively, the closer the parameters λ and δ are to each other, the better the privacy. In the extreme case of $\lambda = \delta$, we have perfect privacy, but then the dissemination is independent of u's opinion (and thus of how interesting the information is). In the other extreme, if λ is the maximum degree and $\delta = 0$ (i.e., the user reposts the information iff it likes it), the act of reposting (or not) the information reveals with certainty u's opinion. We formally quantify the privacy properties of RIPOSTE in terms of *ϵ-differential privacy*. In particular, we argue that RIPOSTE is $\ln(\lambda/\delta)$-differentially private.

For the dissemination of information, we prove the following threshold behavior. Suppose that each user likes a given item t with probability p_t, independently of the other users (p_t is the same for all users and depends only on t). Thus p_t is a measure of how interesting item t is, and is equal to the expected fraction of users that like t. Let S denote the set of users who receive item t initially (e.g., these users receive the information from a news channel). We show that if $p_t < p^*$, for $p^* = (1-\delta)/(\lambda-\delta)$, then the expected number of users that learn the information is $O(|S|)$, i.e., at most a constant factor larger than the users exposed to the information initially. This is true for any graph G. On the other hand, we show that the following statement holds for a G from a family of random directed graphs with arbitrary out-degree distribution [6]. (Such a graph could, for example, model the Twitter network). If $p_t > p^*$, then for a random initial set S, information t spreads to $\Theta(n)$ users (i.e., at least some constant fraction of the network), with probability $1 - e^{\Omega(|S|/d)}$, where d denotes the average degree of G. In particular, this result says that information spreads to $\Theta(n)$ users with constant probability when $|S|$ is close to the average degree, and with high probability if $|S|$ is $\log n$ times larger than the average degree. The analysis draws from the theory of branching processes [3], and the intuition is simple: Basic computations yield that the expected number of users that a given user passes the information to, is less than 1 when $p_t < p^*$, and greater than 1 if $p_t > p^*$. The threshold phenomenon we observe follows then from a similar phenomenon in branching processes. We note that the result for $p_t > p^*$ does not hold for arbitrary graphs. However, we expect that it should hold for many graph families, of sufficiently high *expansion*.

We complement our analysis with extensive experimental results. We use a complete snapshot of the Twitter graph of roughly 40 million users from 2009, and smaller samples from other social networks, including Facebook and LiveJournal. The experiments demonstrate clearly the predicted threshold phenomenon, with very limited spread below the p^* threshold, and substantial spread above p^*. The latter suggests that our result for $p_t > p^*$ should qualitatively hold for a larger family of networks than the stylized model analysed formally. We also experiment with non-uniform distributions of user opinions, where users closer to the source users are more likely to like the information, obtaining qualitatively similar results. Experiments suggest that reasonable values for RIPOSTE's parameters in the networks considered are $\delta = 0.75$ and $\lambda = 3$. For these values, the plausible deniability achieved ensures that, for example, if the prior probability for a user to like the information is 0.01 or 0.1, and the user reposts the information, then the probability increases to 0.04 and 0.3 respectively (see Sect. 2.1 for details).

We view the results of this paper as potentially useful for addressing some of the increasing concerns about users' privacy in social networking services. In particular, we think that RIPOSTE could be of interest as a tool for spreading information and petitions in Internet-based activism, a topic of considerable current interest [12,10]. More generally, it is a tool that could be used for widespread dissemination of sensitive information, which people would care to be exposed to, but are not willing to disseminate themselves for the fear of being charged, stigmatized, or isolated. We believe that such a tool could be incorporated in existing social network services. Also our technique could find applications to other distributed problems, such as distributed polling algorithms.

Related Work. RIPOSTE uses a technique that is conceptually similar to the *randomized response technique (RRT)*. RRT was first introduced in 1965 [27] for survey interviews, to increase the validity of responses to sensitive questions. Roughly, the idea is to tell responder to *lie* with some fixed, predetermined, probability p_{lie} (e.g., roll a die and lie whenever the die shows one or two, in which case $p_{\text{lie}} = 1/3$).[4] Since p_{lie} is known to the interviewer, the distribution of responders' truthful answer can be easily estimated, and thus, accurate estimations of aggregate results can be extracted—but an individual's answer is always plausibly deniable. (See [5] for other variations of RRT, and [2] for a variant using cryptography to guarantee that the responder follows the RRT.) In our diffusion mechanism, the same probability of reposting could be achieved using the following RRT-like approach: User u is asked if she likes the post, but is instructed to lie with probability $p_{\text{lie}} = \delta/(\delta + \lambda)$; and if the answer is 'yes' then the post is reposted with probability $(\delta + \lambda)/s_u$.

We are not aware of other works that use randomized responses in a way similar to ours: to achieve dissemination that reflects user's aggregate opinion, while preserving the privacy of individual users' opinion. In a more standard use of RRT, Quercia et al. [25] proposed a method to aggregate location data of mobile phone users by having each user report several erroneous locations in addition to the correct one. Recently, Erlingsson et al. [11] presented an RRT-based algorithm for crowdsourcing of population statistics from end-user client software, deployed on Google's Chrome Web browser.

Another mechanism provided by social networking services, besides reposting, which has the potential to make interesting posts widely visible is that of *liking*. This mechanism has similar privacy issues as reposting. In [1], Alves et al. proposed a scheme to anonymize user's likes, which keeps the actual like count of a post without revealing the names of the users who like it. Unlike our approach, the scheme employs cryptographic techniques to achieve privacy, and requires a centralized server (but the server does not know the users' opinion).

We have said that our diffusion scheme could provide a tool for Internet-based activism [12,10]. The use of pseudonyms, combined with methods for hiding the user's IP, has also been a common practice used by activists to hide their identity while spreading sensitive information [29]. Our scheme protects the users who contribute to the dissemination of information, but not the sources of the information. This is not a problem in some settings, for example, if we assume that anti-government information originates

[4] The closer is p_{lie} to $1/2$ the better the privacy.

from a news channel (say, WikiLeaks) located in a different country. If this is not the case, then pseudonyms could be used to protect the privacy of the source.

We measure the privacy properties of our diffusion scheme in terms of *differential privacy* [8,9]. Differential privacy was introduced in the context of privacy-preserving analysis of statistical databases. Roughly speaking, differential privacy ensures that (almost, and quantifiably) no risk is incurred by joining a statistical database. More relevant to our setting is the *local model* of differential privacy [17], also known as *fully distributed model*. In this model, there is no central database administrator of private data; each individual maintains their own data element (a database of size 1), and answers questions about it only in a differentially private manner. The local privacy model was first introduced in the context of learning, where it was shown that private learning in the local model is equivalent to non-private learning in the statistical query model [17,15].

2 The Diffusion Algorithm

In this section, we describe our diffusion mechanism for disseminating information in an online social network, and provide an analysis of its properties.

We model the social network as a directed graph $G = (V, E)$ with $|V| = n$ nodes. Each node $u \in V$ represents a user (from now on we will use the terms node and user interchangeably), and a directed edge from node u to v denotes that user u can send information to v. For example, for the case of the Twitter social network, an edge $(u, v) \in E$ in the underlying graph G denotes that user v is "following" u. Borrowing Twitter's parlance, in this paper, we will say that v is a *follower* of u if $(u, v) \in E$. The number of u's followers is thus the same as u's out-degree.

We assume that initially a set of users $S \subseteq V$ learns an information item (from a source external to the network). From each user that learns the information, this information can be *reposted* to all its followers. (So, information can either be sent to all followers of the user, or to none.) We propose a randomized distributed algorithm, running locally at each user (i.e., at the user's device connected to the social network service), which decides whether or not to repost the received information; we call this algorithm RIPOSTE.

RIPOSTE takes as input the opinion of the user on the information item, i.e., if the user *likes* or *does not like* the information, and the algorithm's effect is to either repost the information or not. RIPOSTE' decision depends on: (1) the user's opinion, (2) an upper bound on the number of the user's followers that have not received the information yet, and (3) two global parameters of the protocol (the same for all users), denoted δ and λ; both parameters are non-negative real numbers satisfying $\delta < 1$ and $\lambda > 1$. As explained later, these parameters control the privacy properties of the protocol, and influence the dissemination.

RIPOSTE Algorithm: For each new information item received by user u, if u has k followers and $s \leq k$ is an estimate bounding from above the number of u's followers that have not received the item yet, then:
if u likes the item, the algorithm reposts the item with probability

$$r_{\text{like}}(s) := \begin{cases} \lambda/s, & \text{if } s \geq \lambda + \delta, \\ 1 - \frac{\delta(s-\delta)}{\lambda s}, & \text{if } 0 < s < \lambda + \delta; \end{cases}$$

if u does not like the item, it is reposted with probability $r_{\mathrm{dis}}(s) := \delta/s$ (if $s > 0$).

It is easy to verify that $r_{\mathrm{dis}}(s) \leq r_{\mathrm{like}}(s)$, for all s, i.e., the probability of reposting is larger when u likes the item. Also, the closer are δ and λ to each other, the closer are the two probabilities r_{dis} and r_{like}.

The definition of $r_{\mathrm{like}}(s)$ for the case of $s < \lambda + \delta$ will be justified when we analyse the privacy of the protocol. Until then we can assume the following simpler definition for *all* $s > 0$: $r_{\mathrm{like}}(s) := \min\{\lambda/s,\, 1\}$.

RIPOSTE needs to know an upper bound on the number of the user's followers who have not yet received the item. This information is readily available in some existing social network services, including Twitter, where the default setting is that a user can access the list of items each of its followers has received. If this information is not available, then the total number of followers k of the user can be used as the upper bound s. For the analysis and the experimental evaluation, we will make use also of that special variant of RIPOSTE, where $s = k$.

DB-RIPOSTE Algorithm (Degree-Based-Riposte): This algorithm is a special instance of RIPOSTE, where the total number of followers k of user u is used as the upper bound s on the number of u's followers who have not already received the information.

An attractive analytical property of DB-RIPOSTE is that the outcome of the dissemination does not depend on the order in which the algorithm is executed at different users, unlike in the general RIPOSTE algorithm. For our analysis of RIPOSTE we assume that the order can be arbitrary.

We stress that RIPOSTE does not reveal any information on the value of its input (the user's private opinion), other than the statistical information inferred by the outcome of the algorithm, to repost or not. Also, the user cannot *prevent* the algorithm from reposting the information, even if she does not like the information. In particular, if the user refuses to answer whether she likes an item or not, this is interpreted as a negative answer by the algorithm (the user has an incentive to answer positively if she likes the item, as this would potentially result in larger spread).

We now analyze the properties of RIPOSTE, regarding privacy and the spread of information.

2.1 Privacy

RIPOSTE achieves the property of *plausible deniability*: A user can claim that, with reasonable probability, the act of reposting (or not) an information, does not reflect the user's truthful opinion on the information, and is a result of the randomness in the algorithm.

The standard notion used to quantify plausible deniability is that of *differential privacy* [9]. We recall now the definition of an ε-differentially private algorithm. Let A be a randomized algorithm with input a collection of values, $x_1, \ldots, x_m$, that returns a value from some domain R. Since the algorithm is randomized, for a fixed input $x_1, \ldots, x_m$, its output $A(x_1, \ldots, x_m)$ is a random variable, with some distribution over R. Suppose that the input to A is not known to us (is private), and by observing the output of A we

want to find out the value of some of the inputs. More generally, we may have some information about the input, i.e., a distribution over the possible combinations of input values, and we want, by observing A's output, to improve this information, i.e., obtain a distribution closer to the true input values. We can quantify the extent to which this is possible in terms of ε-differential privacy: algorithm A is ε-differentially private if changing exactly one of it inputs $x_1, \ldots, x_m$ changes the distribution of the output by at most an e^{ε} factor.

Definition 1 (ε-differential privacy). *A randomized algorithm A with inputs $x_1, \ldots, x_m$ from some finite domain and output $A(x_1, \ldots, x_m)$ on some domain R, is ε-differentially private if for any two sets of inputs $x_1, \ldots, x_m$ and $x'_1, \ldots, x'_m$ that differ in exactly one value, and for any set of outputs $Q \subseteq R$,*

$$\Pr\big(A(x_1, \ldots, x_m) \in Q\big) \leq e^{\varepsilon} \cdot \Pr\big(A(x'_1, \ldots x'_m) \in Q\big).$$

In our setting, algorithm A is RIPOSTE, which takes a single binary input: the opinion of the user, and has a binary output: repost or not-repost.

Theorem 2. RIPOSTE *is ε-differentially private for $\varepsilon = \ln(\lambda/\delta)$.*

The proof is a straightforward application of the definitions, and can be found in the full version of the paper [13].

Theorem 2 implies that the closer is the ratio λ/δ to 1, the better the achieved privacy. In particular, if $\delta = \lambda$ we have perfect privacy, as the probability of reposting does not depend on the user's opinion—but this is not desirable from a dissemination point of view.

We discuss now what Theorem 2 implies about the information one can gain for the opinion of a user on some information item it receives, by observing whether or not the item was reposted from that user.

Let q be the (prior) probability that the user likes the information, capturing the knowledge of an observer about the user's opinion *before* the observer sees whether or not this information is reposted from the user. Then from Theorem 2 it follows that the probability $\hat{q}$ with which the observer believes that the user likes the information, after the observer learns whether or not there was a repost, satisfies the inequalities

$$\frac{q}{q + (1-q)(\lambda/\delta)} \leq \hat{q} \leq \frac{q}{q + (1-q)(\delta/\lambda)}. \tag{1}$$

(The proof is by Bayes' Rule.) For the typical parameter values $\delta = 3/4$ and $\lambda = 3$ we use later in the experimental evaluation, Ineq. (1) yield, e.g., that if $q = 0.01$ then $0.0025 < \hat{q} < 0.039$; if $q = 0.1$ then $0.027 < \hat{q} < 0.31$; and if $q = 0.9$ then $0.69 < \hat{q} < 0.97$.

Above we have considered the amount of information leaked when observing the cascade of a single information item. However, if one can observe the cascades of a *sufficiently large* number of *sufficiently similar* items, possibly over a long period, then more information can potentially be leaked about the opinion of a user on this type of information. We leave as a future work the study of such correlation attacks.

2.2 Dissemination

In terms of dissemination, the goal of RIPOSTE is that the fraction of users receiving an information item should reflect the users' overall opinion on the item. In particular, information that a large fraction of users like should, typically, be received by a lot of users, while less interesting information should not be received by many users. In the following, we quantify the notions of interesting/not-interesting information by defining a *popularity* threshold, and we provide bounds on the spread of *popular* items (with popularity above this threshold) and *unpopular* items (with popularity below the threshold).

For the analysis, we make the assumption that all users are equally likely to like a given item, independently of their position in the network and the opinion of other users.

Definition 3 (Uniform opinion model & popularity). *Each item t is associated with a probability p_t, called the* popularity *of t, and for each user u, the probability that u likes t is equal to p_t and independent of the other users' opinion about t.*

We note that popularity p_t is also equal to the expected fraction of users that like t. An item's popularity is not known in advance by the diffusion protocol.

We define the popularity threshold p^* as follows. Suppose that user u receives an item with popularity p. Since u has probability p of liking the item in the uniform model, the probability that RIPOSTE reposts the item, if $s > 0$, is $p \cdot r_{\text{like}}(s) + (1-p) \cdot r_{\text{dis}}(s)$. If $s \geq \lambda + \delta$, this probability is $p \cdot (\lambda/s) + (1-p) \cdot (\delta/s)$. Moreover, if s is the exact number of u's followers that have not received the item yet, then the expected number of new users that learn the item from u is s times that, i.e., $p\lambda + (1-p)\delta$. The popularity threshold p^* is then the probability p for which this expectation is equal to 1.

Definition 4 (Popular/Unpopular items). *For given λ and δ, we define the popularity threshold $p^* := \frac{1-\delta}{\lambda-\delta}$, and we call an information item t* popular *if its popularity is $p_t > p^*$, and* unpopular *if $p_t < p^*$.*[5]

Next we establish an upper bound on the spread of unpopular items, and a lower bound on the spread of popular items.

We first argue that the expected number of users who receive a given unpopular item is by at most a constant factor larger that the number of user $|S|$ who receive the item initially (e.g., from a source external to the network). The constant factor depends on the popularity of the item and parameters δ and λ. This bound holds for any network G, assuming the uniform opinion model. Recall that an item is unpopular if its popularity is smaller than $p^* = (1-\delta)/(\lambda-\delta)$.

Theorem 5 (Spread of unpopular items). *For any G, and under the uniform opinion model,* RIPOSTE *guarantees that an item with popularity $p < p^*$ starting from any set S of users is received by an expected total number of at most $|S|/\beta$ users, where $\beta = (p^* - p)(\lambda - \delta)$.*

[5] For the asymptotic bounds we show later, we assume for a popular item t that $p_t > p^* + \epsilon$, and for an unpopular item t that $p_t < p^* - \epsilon$, for some arbitrary small constant $\epsilon > 0$.

The proof of Theorem 5, which can be found in the full version of the paper [13], is based on the fact that the expected number of new users that learn the item from a given user that knows the item is smaller than one.

Observe that as p approaches the popularity threshold p^*, factor β decreases, and thus the bound on the expected spread increases. Further, substituting the definition of p^* gives $\beta = 1 - \delta - p(\lambda - \delta)$, which implies that increasing either λ or δ increases the expected spread. These observations are consistent with the intuition.

Next we consider the spread of popular items. We focus on a particular family of random directed graphs which is convenient for our analysis, but is also a reasonable model of some social network graphs, such as the Twitter graph, characterized by large variation in the nodes' out-degree (i.e., the number of followers) and small variation in the nodes' in-degree. This model is a simplification of one considered in [6], and has a single parameter, a distribution ϕ on the nodes' out-degree.

Definition 6 (Random graph G_ϕ). *For any probability distribution ϕ on the set $\{0, \ldots, n-1\}$, G_ϕ is an n-node random directed graph such that the out-degrees of nodes are independent random variables with the same distribution ϕ, and for each node u, if u has out-degree k, then the set of u's outgoing neighbors is a uniformly random set among all k-sets of nodes not containing u.*

We establish a lower bound on the probability of a popular item to be received by a constant fraction of users in G_ϕ, for an arbitrary distribution ϕ (under a mild constraint on the min out-degree). The above probability and the fraction size grow respectively with the number $\sigma = |S|$ of source nodes, and the popularity p of the item. In particular, the probability converges to 1 for σ larger than the average node degree μ.

Theorem 7 (Spread of popular items). *Let ϕ be any probability distribution on the set $\{\lceil \lambda + \delta \rceil, \ldots, n-1\}$, let $\epsilon, \epsilon' > 0$ be arbitrary small constants, and $1 \leq \sigma \leq n$ be an integer. Any information item with popularity $p \geq p^* + \epsilon$, that starts from a random initial set of σ nodes and spreads in G_ϕ using* RIPOSTE*, is received by at least $(1 - \epsilon') \cdot \frac{\beta n}{\beta + 1}$ users, with probability at least $1 - e^{-\Omega(\sigma/\mu)}$, where $\beta = (p - p^*)(\lambda - \delta)$ and μ is the mean of distribution ϕ.*

Observe that the same constant $\beta = |p - p^*| \cdot (\lambda - \delta)$ appears in both Theorems 5 and 7. Unlike the bound of Theorem 5, the bound of $(1 - \epsilon') \cdot \frac{\beta n}{\beta + 1}$ in Theorem 7 is independent of the number $\sigma = |S|$ of source nodes; substituting the definitions of β and p^*, yields $\frac{\beta}{\beta+1} = 1 - \frac{1}{p\lambda + (1-p)\delta}$, thus the bound above increases when any of λ, δ, or p increases. The independence from σ is intuitively justified, because as long as the item reaches a "critical mass" of users, it will almost surely spread to a constant fraction of the network. However, the probability with which such a critical mass will be reached does depend on σ. For σ close to the average degree μ, this probability is at least a constant, and quickly converges to 1 as σ/μ increases above 1.

The proof of Theorem 7 uses a coupling between the dissemination process and an appropriate branching process, to show that the probability of the event we are interested in, that at least a certain fraction of users receive the item, is lower-bounded by the survival probability of the branching process. Then we bound this survival probability using a basic result for branching processes.

Proof of Theorem 7. It suffices to prove the claim for DB-RIPOSTE. The reasons is that the reposting probabilities $r_{\mathrm{like}}(s)$ and $r_{\mathrm{dis}}(s)$ are minimized when s equals the number k of the user's followers, and thus a standard coupling argument shows that the number of users that receive the item if DB-RIPOSTE is used is dominated stochastically by the same quantity when RIPOSTE is used.

We couple the diffusion process in G_ϕ with an appropriate branching process. Recall that a (Galton-Watson) branching process is a random process starting with one or more individuals, and in each step of the process a single individual produces zero of more offsprings and then dies. The number of offsprings of an individual follows a fixed probability distribution, the same for all individuals. The process either finishes after a finite number of steps, when no individuals are left, or continues forever. The probabilities of these two complementary events are called *extinction* and *survival probability*, respectively.

First we compute the distribution of the number of new users that learn the item from a user u, at a point in time when *fewer than ℓ users in total have received the item*—we will fix the value of ℓ later. The probability that u has exactly i followers is $\phi(i)$, for $i \in \{\lceil \lambda + \delta \rceil, \ldots, n-1\}$ (and 0 for other i). Given that u has i followers, the probability that DB-RIPOSTE reposts the item from u is $(p\lambda + (1-p)\delta)/i = (\beta+1)/i$. Further, by the principle of deferred decision, we can assume that if the item is reposted from u, only then are u's i followers chosen. We can also assume that they are chosen sequentially, one after the other, and the item is sent to a follower before the next follower is chosen (this does not change the overall outcome of the dissemination). Then the probability that the j-th follower of u has not already received the item is at least $1 - \ell/n$, provided that at most ℓ users already know the item (including the first $j-1$ followers of u).

Consider now the branching process in which σ individuals exist initially, and the number X of offsprings of an individual is determined as follows. First an integer i is drawn from distribution ϕ; then with probability $1-(\beta+1)/i$ we have $X = 0$ offspring, and with the remaining probability, $(\beta+1)/i$, we draw X's value from the binomial distribution $B(i, q)$, for $q := 1 - \ell/n$ (this is the distribution of the number of successes among i independent identical trials with success probability q).

We use a simple coupling of the diffusion process with the branching process above, until the point when ℓ users have received the item or the dissemination has finished (whichever occurs first). We assume that the diffusion process evolves in steps, and each step involves the execution of the DB-RIPOSTE algorithm at a single node. Similarly a step in the branching process is that a single individual reproduces and then dies. Let N_t denote the number of new users that learn the item in step t of the diffusion process, and let X_t be the number of offsprings born in step t of the branching process. From our discussion above on the distribution of N_t and from the definition of the distribution of $X_t \sim X$, it follows that we can couple N_t and X_t such that $N_t \geq X_t$ if no more than ℓ users in total have received the item in the first t steps.

From this coupling, it is immediate that the probability at least ℓ users receive the item in total, is lower-bounded by the probability that the total progeny of the branching process (i.e., the total number of individuals that ever existed) is at least ℓ. Further, the latter probability is lower bounded by the *survival probability* of the branching process;

we denote this survival probability by ζ_σ. Thus to prove the theorem it suffices to show

$$\zeta_\sigma = 1 - e^{-\Omega(\sigma/\mu)},$$

for $\ell := (1-\epsilon') \cdot \beta n/(\beta+1)$. The remainder of the proof is devoted to that.

By the definition of the branching process, the expected number of offsprings of an individual is

$$\begin{aligned}\mathbf{E}[X] &= \sum_i \phi(i) \cdot \frac{\beta+1}{i} \cdot \mathbf{E}[B(i,q)] \\ &= \sum_i \phi(i) \cdot \frac{\beta+1}{i} \cdot iq = \sum_i \phi(i) \cdot (\beta+1) \cdot q = (\beta+1) \cdot q.\end{aligned}$$

We observe that $\mathbf{E}[X] > 1$, as

$$(\beta+1) \cdot q = (\beta+1) \cdot \left(1 - \frac{(1-\epsilon')\beta}{\beta+1}\right) = 1 + \epsilon'\beta. \tag{2}$$

Further,

$$\begin{aligned}\mathbf{E}[X^2] &= \sum_i \phi(i) \cdot \frac{\beta+1}{i} \cdot \mathbf{E}[(B(i,q))^2] = \sum_i \phi(i) \cdot \frac{\beta+1}{i} \cdot (i^2q^2 + iq(1-q)) \\ &= \sum_i \phi(i) \cdot (\beta+1) \cdot (iq^2 + q(1-q)) = (\beta+1) \cdot (\mu q^2 + q(1-q)),\end{aligned}$$

where $\mu = \sum_i \phi(i) \cdot i$ is the mean of ϕ. We will use the following standard lower bound on the survival probability ζ_1, when there is just one individual initially (see, e.g., [16, Sect. 5.6.1]),

$$\zeta_1 \geq \frac{2(\mathbf{E}[X]-1)}{\mathbf{E}[X^2] - \mathbf{E}[X]}.$$

Substituting the values for $\mathbf{E}[X]$ and $\mathbf{E}[X^2]$ computed above yields

$$\begin{aligned}\zeta_1 &\geq \frac{2(q(\beta+1)-1)}{(\beta+1)(\mu q^2 + q(1-q)) - q(\beta+1)} = \frac{2(q(\beta+1)-1)}{q^2(\beta+1)(\mu-1)} \\ &= \frac{2(q(\beta+1)-1)(\beta+1)}{q^2(\beta+1)^2(\mu-1)} \overset{(2)}{=} \frac{2\epsilon'\beta(\beta+1)}{(1+\epsilon'\beta)^2(\mu-1)} = \Omega(1/\mu),\end{aligned}$$

where the final equation holds because $\beta = (p-p^*)(\lambda-\delta) \geq \epsilon(\lambda-\delta) = \Omega(1)$.

We can now express ζ_σ in terms of ζ_1, by observing that a branching process starting with σ individuals can be viewed as σ independent copies of the branching process starting with a single individual each.[6] The former branching process survives if and only if at least one of the latter ones survives, thus,

$$\zeta_\sigma = 1 - (1-\zeta_1)^\sigma \geq 1 - e^{-\zeta_1\sigma} = 1 - e^{-\Omega(\sigma/\mu)}.$$

This completes the proof of Theorem 7. □

[6] This is true for any branching process, and does not relate to the original diffusion process.

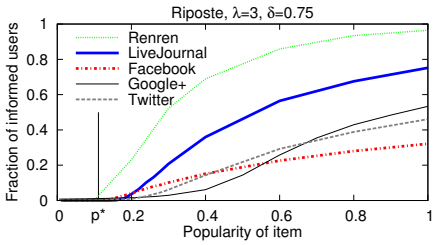


Fig. 1: Dissemination for RIPOSTE as a function of the item popularity.

Table 1: Network topologies used in the experiments. By avg-deg we denote the average degree of the network.

Network	Nodes	Edges	Avg-deg	Source
Twitter	41.65M	1468M	35.2	[18]
LiveJournal	4.847M	68.99M	14.2	[4,19]
Facebook	3.097M	23.66M	15.3	[28]
Renren	965.3K	57.56M	59.6	[7]
Google+	107.6K	13.67M	127	[20]

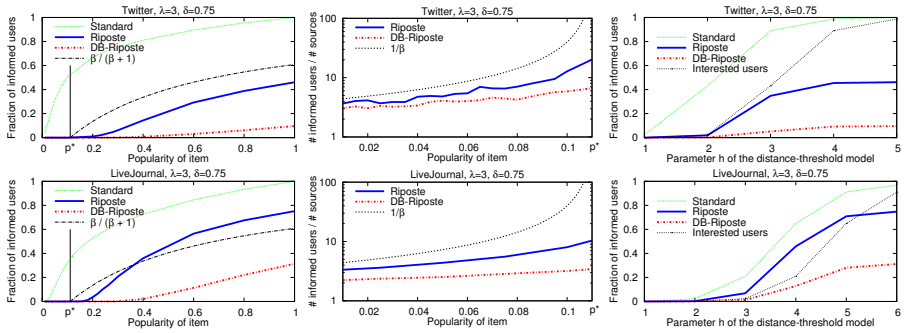
3 Experiments

In this section we provide experimental evaluation of the dissemination achieved by RIPOSTE on some real topologies of online social networks. The results are surprisingly consistent with our analysis, even though some of the analytical results were proven only for an ideal random graph model.

Datasets. We use the network topologies listed in Table 1. The Twitter dataset is a complete snapshot of the network from 2009 [18], while the other datasets are partial network samples. Twitter is a micro-blogging network service, LiveJournal is a blogging network site, while Facebook, Renren, and Google+ are online social networking sites. In each of these networks, every user maintains a list of friends, and/or a list of users she follows. The friendship relation is typically reciprocal, whereas the follower relation is not; the former is represented as an undirected edge, and the latter as a directed. In Twitter, LiveJournal and Google+ edges are directed, while in Renren and Facebook undirected.

Setup. We consider the following protocols: (1) RIPOSTE, with exact information on the number of non-informed followers, i.e., s is the *actual* number of the user's followers that do not know the item yet—not just an upper bound; (2) DB-RIPOSTE, where no information about the followers status is available, and thus s is the total number of followers; (3) the basic non privacy-conscious protocol where a user reposts an item if she likes it and does not repost it if she does not like it; we refer to this as the STANDARD protocol.

While datasets on social network topologies are publicly available, access to user's activity, including the list of items they post, receive, like or repost, is severely restricted. Therefore, for our evaluation we rely on two synthetic models to generate users' opinions: (i) the *uniform opinion model*, where every item is assigned a popularity $p \in [0, 1]$, and each user likes the item independently with probability p—this is the same model used in the analysis (see Definition 3); and (ii) the *distance-threshold opinion model*, where a user likes the item precisely if the (shortest-path) distance from a source to the user is at most some threshold h. The latter model is motivated by the principle that users close to each other tend to have similar opinions [22].



(a) (b) (c)

Fig. 2: Dissemination in Twitter (top) and LiveJournal (bottom). (a) Comparison with STANDARD and the $\beta/(\beta+1)$ lower bound of Theorem 7. (b) Comparison with the $1/\beta$ upper bound of Theorem 5 for unpopular items. (c) Distance-threshold model (all users within distance h from the source, and only them, like the item).

In all experiments, we choose the set S of users who know the item initially to be the followers of a random user, among all users with at least μ followers, where μ is the average degree. We think that this is more realistic than choosing an arbitrary or random set S: It is often the case that the source of the information (e.g., a news channel) is itself a node in the online social network; then the followers of that node constitute the set S of nodes exposed to the information initially. For each point in the plots we present, we average the results over 10,000 random independent experiments, with a new random set S each time. For the RIPOSTE algorithm, where the dissemination may depend on the order in which the protocol is executed at different users, we experimented with both breadth-first and depth-first orders, obtaining very similar results.

Results. Fig. 1 shows the average number of users that receive the item when using RIPOSTE, as a function of the item popularity, for all networks (for parameters $\lambda = 3$ and $\delta = 0.75$). In all cases, unpopular items (with popularity p below the threshold p^* identified by our analysis) have very limited spread, while popular items (with $p > p^*$) spread to a fraction of the networks that grows quickly with p. Due to space limitations, in the following we present results only for Twitter and LiveJournal; the results for the other three datasets are qualitatively similar and can be found in the full version [13].

Fig. 2a compares the dissemination using RIPOSTE to that of DB-RIPOSTE and STANDARD, and also to the lower bound for the spread of popular items predicted by Theorem 7. As expected, DB-RIPOSTE informs fewer users than RIPOSTE but has overall qualitatively similar behaviour. STANDARD achieves significantly wider dissemination, even for items with very low popularity, which may be undesirable. The $\beta/(1-\beta)$ bound of Theorem 7 is relatively close to the curve for RIPOSTE (slightly above it in the case of Twitter and intersecting it in the case of LiveJournal). This lower bound was derived for an idealized random graph model, so it is reasonable that it does not apply exactly to the real topologies considered. On the other hand, the $1/\beta$ upper

bound for unpopular items of Theorem 5 holds for *any* graph, and Fig. 2b shows that it indeed bounds the dissemination with RIPOSTE in both Twitter and LiveJournal. Finally, Fig. 2c presents the same results as Fig. 2a but for the distance-threshold opinion model. We observe that RIPOSTE achieves spread to a fraction of users that is relatively close to the fraction of users that like the item. As before, STANDARD may spread the item to a fraction significantly larger than the fraction that likes the item, in particular, for items that not many users like. Additional experimental results can be found in the full version of the paper [13].

4 Conclusion

We have presented a simple and local diffusion mechanism for social networks, which guarantees widespread dissemination of interesting but possibly sensitive information, in a privacy-conscious manner. The mechanism randomizes the user's action of reposting (or not) the information, in a way reminiscent of the randomized response technique, and chooses the probabilities so that a branching-process-like phenomenon takes place: if more than a certain fraction of people like the information then a large cascade of reposts is formed, and if fewer people like it then the diffusion process dies quickly. We believe this mechanism to be relevant as a tool for internet-based activism, and more generally for promoting free speech. We also think that our techniques could find applications to other distributed problems, such as distributed polling.

References

1. P. Alves and P. Ferreira. AnonyLikes: Anonymous quantitative feedback on social networks. In *14th International Middleware Conference (Middleware)*, pages 466–484, 2013.
2. A. Ambainis, M. Jakobsson, and H. Lipmaa. Cryptographic randomized response techniques. In *7th International Workshop on Theory and Practice in Public Key Cryptography (PKC)*, pages 425–438, 2004.
3. K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, 1972.
4. L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 44–54, 2006.
5. A. Chaudhuri. *Randomized response and indirect questioning techniques in surveys*. CRC Press, 2010.
6. N. Chen and M. Olvera-Cravioto. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.
7. C. Ding, Y. Chen, and X. Fu. Crowd crawling: Towards collaborative data collection for large-scale online social networks. In *1st ACM Conference on Online Social Networks (COSN)*, pages 183–188, 2013.
8. C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *3rd Theory of Cryptography Conference (TCC)*, pages 265–284, 2006.
9. C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.
10. J. Earl. The dynamics of protest-related diffusion on the web. *Information, Communication & Society*, 13(2):209–225, 2010.

11. Ú. Erlingsson, V. Pihur, and A. Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *ACM Conference on Computer and Communications Security (CCS)*, pages 1054–1067, 2014.
12. K. Garrett. Protest in an information society: A review of literature on social movements and new icts. *Information, Communication & Society*, 9(02):202–224, 2006.
13. G. Giakkoupis, R. Guerraoui, A. Jégou, A.-M. Kermarrec, and N. Mittal. Privacy-conscious information diffusion in social networks. Technical report, INRIA Rennes - Bretagne Atlantique, Aug. 2015. `https://hal.archives-ouvertes.fr/hal-01184246`.
14. J. Grasz. Forty-five percent of employers use social networking sites to research job candidates, CareerBuilder survey finds. *CareerBuilder Press Releases*, Aug. 2009. `http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?id=pr519&sd=8%2f19%2f2009&ed=12%2f31%2f2009&siteid=cbpr&sc_cmp1=cb_pr519_`.
15. A. Gupta, M. Hardt, A. Roth, and J. Ullman. Privately releasing conjunctions and the statistical query barrier. *SIAM Journal on Computing*, 42(4):1494–1520, 2013.
16. P. Haccou, P. Jagers, and V. A. Vatutin. *Branching processes: Variation, growth, and extinction of populations*. Cambridge Univ. Press, 2005.
17. S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal of Computing*, 40(3):793–826, 2011.
18. H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *19th International Conference on World Wide Web (WWW)*, pages 591–600, 2010.
19. J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.
20. J. Leskovec and J. J. Mcauley. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 539–547, 2012.
21. S. A. Macskassy and M. Michelson. Why do people retweet? Anti-homophily wins the day! In *5th International Conference on Weblogs and Social Media (ICWSM)*, 2011.
22. M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.
23. E. Noelle-Neumann. The spiral of silence a theory of public opinion. *Journal of Communication*, 24(2):43–51, 1974.
24. NPR news. In South Korea, old law leads to new crackdown. `http://www.npr.org/2011/12/01/142998183/in-south-korea-old-law-leads-to-new-crackdown`, Dec. 2011.
25. D. Quercia, I. Leontiadis, L. McNamara, C. Mascolo, and J. Crowcroft. SpotME if you can: Randomized responses for location obfuscation on mobile phones. In *31st IEEE International Conference on Distributed Computing Systems (ICDCS)*, pages 363–372, 2011.
26. TIME magazine. Indian women arrested over facebook post. `http://newsfeed.time.com/2012/11/19/indian-woman-arrested-over-facebook-like/`, Nov. 2012.
27. S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.
28. C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *EuroSys*, pages 205–218, 2009.
29. V. Wulf, K. Misaki, M. Atam, D. Randall, and M. Rohde. 'On the ground' in Sidi Bouzid: Investigating social media use during the tunisian revolution. In *16th ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 1409–1418, 2013.