Efficient Dissemination of Emergency Information using a Social Network

Iouliana Litou
Department of Informatics
Athens University of Economics and Business,
Athens, Greece
litou@aueb.gr

Ioannis Boutsis
Department of Informatics
Athens University of Economics and Business,
Athens, Greece
mpoutsis@aueb.gr

Vana Kalogeraki
Department of Informatics
Athens University of Economics and Business,
Athens, Greece
vana@aueb.gr

ABSTRACT
In the recent years social networks have undergone explosive growth. They have been used as major tools for the spread of information, ideas and notifications among the members of the network. In this paper we aim at exploiting this new communication channel for emergency notification, to deliver emergency information to all appropriate recipients. We develop ESCAPE, our system for efficient dissemination of emergency information in social networks. We propose an approach that investigates the interactions and relationships established between the members of the social group, and develops a scalable dissemination mechanism that selects the most efficient routes to maximize the information reach. Our experimental results illustrate the feasibility and performance of our approach.

Categories and Subject Descriptors
C.2.4 [**Distributed Systems**]

Keywords
Distributed Systems, Social Networks, Information Dissemination

1. INTRODUCTION
In the recent years social networks such as Facebook, Twitter and Google+ have undergone an explosive growth, enumerating large numbers of subscribers. For example, Facebook counts over 900 million active users, followed by Twitter with over 550 million users and Google+ with over 170 million users[1]. They have been used as major tools for the spread of information, ideas and notifications among the members of the network. Recent studies reveal that social networks can be used efficiently not only for "viral marketing" [17] to promote new products to targeted sets of users which further propagate the products through the "word-of-mouth" effect to reach a larger audience, but also for discovering emergent topics[1], emergency alerting, management and public safety [16].

Consider for example the emergency event of an earthquake. People in the vicinity of earthquakes are sharing anecdotal information that earthquake alerts lagged behind firsthand notification sent through Twitter, a popular Internet-based service for sending and receiving short text messages[2, 9]. The study reveals that depending on the size and location of the earthquake, scientific alerts can take between two to twenty minutes to publish, while using Twitter's notification capabilities people were notified about the occurrence of the earthquakes shaking within seconds of their occurrences.

Combining geographic coordinates with social networks, enables social networks to interact with users relative to their locations, or connect users with local events, places or groups that match their interests. This is becoming increasingly popular in several geosocial applications such as Facebook Places[2] and Foursquare[3], where users are allowed to share their geographic locations as well as make and receive recommendations for a set of venues. In an emergency scenario, such geosocial networks contribute not only to develop a collective situational awareness about the event, but also allow users to coordinate around geotag information related to hazards and disaster aid activities.

Thus, social networks (*i.e.,* Twitter[4], Facebook[5], LinkedIn[6]) are opening new avenues for massive emergency notification due to their ability to (1) reach millions of social network users, especially family and friends, (2) become alternative communication mediums when wireless and telecommunication networks are congested during emergencies, and (3) provide cost-effective solutions as they have the ability to reach massive amounts of users without added infrastructure costs.

However, adopting social networks as an effective communication medium for emergency alerting raises considerable challenges in the level of availability and responsiveness ex-

[1] http://www.go-gulf.com/social-networking-users.jpg

[2] https://www.facebook.com/about/location
[3] http://www.foursquare.com
[4] https://twitter.com/
[5] https://www.facebook.com/
[6] http://www.linkedin.com/

Published in the Workshop Proceedings of the EDBT/ICDT 2014 Joint Conference (March 28, 2014, Athens, Greece) on CEUR-WS.org (ISSN 1613-0073).

pected from these infrastructures in delivering notifications to reach all recipients interested in receiving this information (these can be people located in the area of the event *i.e.,* students in a campus, as well as their relatives and friends).

In this paper we illustrate the problem of how to leverage the social network for efficient dissemination of emergency information. Our objective is stated as follows: Given a social network comprising a number of users, the social relationships of the users and the set of recipients , our goal is to select an appropriate subset of the users to propagate the emergency information such that (1) the expected spread of information is maximized among interested users, (2) costs are considered. Cost is defined as the amount of messages that need to be exchanged among users. Thus, it could be translated as either monetary cost (for an SMS) or resource allocation cost.

We approach the problem by following discrete procedures where user profiles are built, social relationships are inferred and dissemination paths among the nodes of the social network are computed and during the occurrence of an emergency event, a small number of seed nodes is selected to efficiently disseminate the emergency information to all interested recipients during the event.

Current influence maximization approaches are not adequate to solve these problems. The problem of maximizing the spread of influence in social graphs has been addressed in [8, 19, 24, 13], but none of these works has study the problem in the context of emergency notification. Furthermore, they aim at maximizing the influence in the entire network rather than identifying and informing an appropriate subset of nodes that would be most interested to the event.

Emergency response outside social networks has also been studied. The use of geographical notification systems has been considered in [14]. The system described is meant primary for constructing overlays that support location-based regional multicasting where they also consider issues of providing reliable storage of social information events under extreme regional conditions. Traditional approaches such as multicast [21, 11] and publish/subscribe systems [18] are not appropriate for our setting since they will inform only subscribed users, while we need to alert all users associated with the emergency event. In our system, we consider users are already subscribed to the network. Moreover, the set of users to be informed by our system is not determined based only on locational criteria, but also on relationship criteria, so its not considered to be a strictly location-based approach.

On the other hand, approaches like flooding and gossiping [4, 6, 7] will inform most of the users interested in the event, but they will also produce a large amount of spamming to the other users, thus adding extra cost to the network.

Our paper makes the following contributions:

- We present ESCAPE (Efficient diSsemination using soCiAl graPh for Emergency response), a system that solves the problem of efficient dissemination of emergency information in social networks. We show that the problem of selecting an appropriate influential set of individuals to maximize the spread of information is NP-hard and provide a greedy algorithm to solve it. We do not aim at spread maximization as previous works, but rather at reachability maximization of a subset of users with the least cost.
- We perform experiments to validate our approach. Our experimental results illustrate that our approach is practical and effectively addresses the problem of informing the maximum amount of users with the least messages when an emergency event occurs, and outperforms its competitors.

2. BACKGROUND

Kempe et al.[13] were the first to propose cascade models in Social Networks. They define two models describing the way influence is propagated in Social networks, namely the Independent Cascade Model and the Linear Threshold Model.

Both models require that a weighted graph representing the social network is given. In the graph, nodes represent users, edges represent influence flow between users and weights represent the probability that the influence propagation is successful among nodes connected with that edge.

In the Linear Threshold Model, aside from the weights to the edges, a threshold is also associated with each user. That threshold expresses the susceptibility of a user to the influence. Nodes that are already influenced are referred as active nodes and the remaining as inactive. A node in the Linear Threshold Model in considered to be activated when the sum of the edges of its currently active neighbors reaches the threshold.

Unlike the Linear Threshold Model, in the Independent Cascade Model no thresholds are considered. The cascade is progressing in steps and at each step, currently active nodes have the chance to influence their neighbors. Nodes have only a single chance of activating their neighbors, and the probability that they succeed is defined by the weight of the edge.

In this work we consider the Independent Cascade Model to be more appropriate of describing the way the information is spread.

3. PROBLEM DEFINITION

We now provide a formulation of the problem and prove its NP-completeness. Consider a social graph $G = (V, E, W)$, where each vertex $u \in V$ represents a user, each edge $e_{uv} \in E$ denotes a social relation between a pair of users (u, v), and w_{uv} corresponds to the strength of the relationship between the users. The relations between the users are assumed to take place over the lifespan of the network (an edge occurs between a pair of users if and only if the two users are connected socially in some manner). Given a social graph $G = (V, E, W)$, a subset of the vertices $S \subset V$, and a positive integer $k < |V|$, our goal is to find a seed set $M \subset V$, such that the expected number of nodes in S informed by M is maximized, and $|M| \leq k$.

Thus, the problem to be solved is how to maximize the amount of nodes $n \in S$ that will be informed given a maximum amount of k seeds that can be used. Note however that not all nodes are constantly connected to the system. Thus, we aim at maximizing the information spread by selecting at most k vertices from V to efficiently disseminate the messages, under the condition that these nodes are connected.

Our problem differs from traditional influence maximization problems, such as the Independent Cascade (IC) Model [13]. The difference is that our goal is to inform a subset of nodes, $S \subset V$, referred as *interested nodes*, which are closely affected by the event. Thus, we aim at maximizing the number of nodes $n \in S$ that will be informed, rather than informing all the nodes in the graph. The key challenge here is that the reachability of the nodes, in terms of physical connectivity, introduces constraints on the availability of the nodes in the graph, since there may only be a subset R of nodes that can be reached. Thus, not all nodes of the network are candidate seeds.

Similarly to the IC model, whenever a node is informed, there exists only one chance that this node forwards the message to its neighbors. The problem as stated above is NP-complete. The reduction from Hitting-Set to this problem is quite trivial. The Hitting-Set problem is defined as described below.

Hitting-Set: Given a set $A = a_1, ..., a_n$ and a collection $B_1, ..., B_m$ of subsets of A and a number k. There exists a Hitting-Set $H \subseteq A$ of size k such that $H \bigcap Bi \neq \emptyset$, $1 \leq i \leq n$.

Proof: Reduction from Hitting-Set problem. If we consider the original set V of nodes and subsets of V B_1, B_2,.., B_m constructed in a way that if one node in B_i is informed then all others in B_i are informed too, which is an assumption that makes the problem easier, then we need to find a set $H \subseteq V$ of size k such that $H \bigcap Bi \neq \emptyset$, $1 \leq i \leq m$. Without considering the assumption $H \subseteq R$ the problem is NP-Complete. The restriction just makes it harder.

Since the problem is NP-Complete we develop an approximation algorithm to solve the problem efficiently.

3.1 Overview of the ESCAPE System

In this section we provide an overview of our ESCAPE (Efficient diSsemination using soCiAl graPh for Emergency response) system. In order to achieve maximum reachability of interested users, the system implements the following procedures: i)Profiling of users based on past actions, ii) Social Strength assignment, denoted as the weights of the edges among users in the social graph and iii) a Dynamic Notification of a subset of users (referred as seeds) to initially receive the information when an event occurs, and further propagate it.

As users connect to the social network, it is possible to extract user information by exploiting the networks created and the messages exchanged among the users. The Profiling procedure (further discussed in 4.1) is responsible for building user profiles and maintain user statistics. It uses the raw data of the user interactions to generate a list I_u for each user $u \in V$, that contains the interactions of user u with any other user in the network. The weights of the social graph $G(V, E, W)$ are inferred based on the information extracted by the Profiling procedure and it characterizes the social relationships among the users based on past interactions, thus it denotes the Social Strength among a pair of user in the graph.

Finally, when an event occurs the Dynamic Notification procedure is triggered. Information computed and maintained by the Social Graph that is formed among users is utilized to identify the initial receivers of the message so as to maximize the spread of information to the interested users, while using the least amount of messages.

The overall architecture and the interaction between different procedures is depicted in Figure 1 and the corresponding functionalities are described in the following sections.

4. EFFICIENT DISSEMINATION OF EMERGENCY INFORMATION

In this section we start by describing the metrics we use to identify and characterize the social relationships among the users in the social graph and then we present our dynamic notification algorithm that aims at selecting a subset of the users to propagate the information.

4.1 Profiling

To construct the social graph, the first step is to identify and characterize the social relationships through the messages exchanged among the users. We build user profiles for each user $u \in V$. Each user u is associated with a unique id, this for example might be the user's id in the social network, and is used each time the user logs into the system.

Whenever user u sends a message to a user v, the list of interactions I_u of user u is updated. The form of the tuples in the I_u list is: $< v, m_v, timestamp_v >$, where v is the id of the receiver of the message m_v, $timestamp_v$ denotes the unix timestamp when the message was sent.

The data retrieved from the user profiles is used for the Social Graph construction $G(V, E, W)$. This is represented as a directed weighted graph. The graph is not required to be updated in a continuous manner, but in discrete time intervals. Each user u in the Social Network forms a node in the graph. For each node v in the I_u list of user u, an edge is instantiated in the graph between u and v. We associate a weight with each edge in the graph to represent the "strength of the relationship" among the users in the social network. The weight takes values in the range between $(0, 1]$, where a value of 1 denotes a strong relationship between the users. The weights w_{uv} are assigned according to one of the metrics described below. We note that there are several mechanisms for assigning weights to edges as stated in the related work section [10, 22, 25]. Although in our work we focus on the metrics described below, the model is generic enough and can be extended so that several metrics can be considered.

To compute the weights of the edges we consider a simple metric, that represent the "social strength" of the relationship between two users:

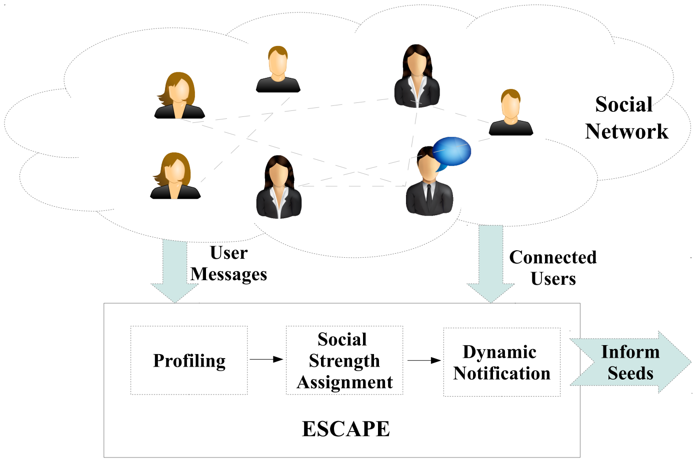

Figure 1: The ESCAPE Architecture.

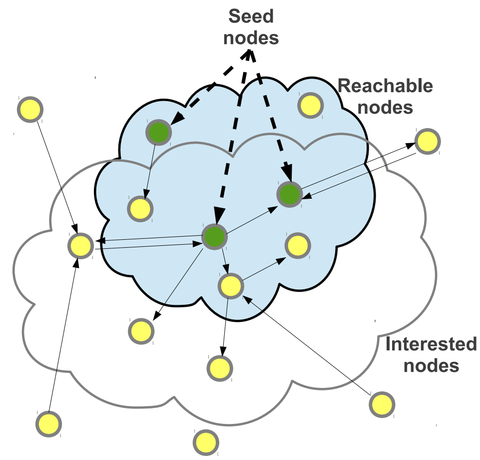

Figure 2: The Node Types.

- *Frequency of communication (FC).* Usually people that are strongly connected to each other, communicate more often. Thus, the frequency of the communication captures the strength of the relationship between the users. The frequency is not defined as an absolute value for all users (i.e. twice a day) as some users are more sociable than others, and that should be taken into account. So, the frequency of communication between a pair of users (u, v) is defined as:

$$f_{uv} = |m_v| \in I_u / \sum_{i \in I_u} m_i \tag{1}$$

  Thus, f_{uv}, for user u, denotes the amount of messages exchanged with user(v), denoted as m_v, out of the total messages that user u has sent to every other user i, m_i. In the case of the Twitter social network, the strength of the user relationships is perceived through the tweets exchanged. It is important to note that f_{uv} differs from f_{vu}, since Twitter presents large asymmetry in the relations due to broadcasters and miscreants [15].

- *Regularity of communication (RC).* It is known that people having strong social bonds may not necessarily communicate very often, but they may be communicating at regular time intervals. So regularity of communication is also considered as an important factor in calculating the strength of a relationship. The regularity metric is defined as:

$$f_{uv} = 1/\log\left(d_{uv} + 1\right) \tag{2}$$

  where d_{uv} represents an average time lapse (e.g. days) between user communication. For instance, if user u interacts at each time lapse (e.g. daily) with user v then d_{uv} equals 1. The regularity is time-window based and considers the regularity of communication within this time window.

We compute the normalized weight so that it is insensitive to the user's special characteristics and does not depend on the set of data measured for a specific user as:

$$w_{uv} = f_{uv}/max\{f_{uv'} : v' \in neighbor(u)\} \tag{3}$$

Note, though, that the above mechanisms aim at deriving the strength of a relationship among users automatically, without user involvement. However, in some cases, it is desirable that users are given the opportunity to define their own set of individuals to be informed in cases of emergency. In this case, the user defines the list of emergency contacts and the corresponding weights of those edges equal 1, regardless of the metric.

For our experiments we use the FC metric for characterizing the weights among users in the network, since it is more appropriate for our dataset, as RC may require data collected for a longer time to appropriately and stably characterize the strength.

4.2 Dynamic Notification

Whenever an event occurs we identify the following roles among the users (shown in Figure 2): (a) **Interested nodes** are all nodes that are interested in the occurrence of the event. These are nodes that are more important in information spreading process and the ones we aim to reach. They are subset of the original graph and we represent them as the nodes belonging to the white cloud in the figure. They represent users within the area of the event or users related to them or to the area in someway, but are not physically there. (b) **Reachable nodes** are the nodes that can be reached after the occurrence of the event, *i.e.* information can be directly sent to them and they can be either interested or not. These nodes are the ones that belong to the blue cloud. (c) **Seed nodes** are the nodes to which information is initially sent, so they are the nodes that we aim at identifying to initiate the propagation process. These nodes are subset of the reachable nodes and we illustrate them as green nodes.

The basic functions of the Dynamic Notification process is to determine the users interested in the event, and select an initial set of nodes (*seeds*) that have will be informed for the event and initialize the propagation process to reach interested users. That is accomplished using a greedy approximation algorithm in order to maximize the finally reached nodes given the size of seed nodes. It is noted that seeds are nodes that can be accessed during the occurrence of the event, thus consists of the reachable nodes in figure 2.

4.2.1 Seed Selection

When an event occurs, the seed selection process is triggered. The first step of the process is to determine the interested nodes. These can be determined in various ways based on geosocial criteria. They could be: (i) users who are physically located to the place of the event, (ii) related to users (*i.e.* relatives), (iii) related to the place of the event (*i.e.* students). The next step is to determine the reachable nodes in the network *i.e.* nodes that are connected and are accessible, that is they are connected in the social network immediately after the event (*e.g.* these can be obtained when a user logs into the network). Let R be the set of reachable nodes and S be the set of interested to the event users. For nodes in R we determine the paths that start at node $u \in R$ and terminate at node $v \in S$. For those paths the probability p(u,v) that a message initialized be u traverses the path and reaches v is calculated. The probability that the path is traversed can be calculated in various ways. For our experiments we define the probability as the product of the weights of all edges that must be traversed from u in order to reach v. After probabilities are calculated the greedy step for the seed selection process follows.

Greedy Node Selection: Consider a set of seeds M, a set of nodes A that we expect that will be informed by M, thus $M \subseteq A$, the set of interested nodes S, the set of reachable nodes R, and a candidate seed $u \in R$ to be added in M. The greedy step for the selection of u is:

$$M \cup u \; s.t. \; \sigma(u) = max\{\sigma(v) : \forall v \in R \setminus A\} \tag{4}$$

$$where \quad \sigma(u) = \frac{(\sum_{v \in S \setminus A} p(u,v))^2}{|S \setminus A|}$$

Intuitively, $\sigma(u)$ computes the number of interested nodes $v \in S$ informed when u is selected as the root of the dissemination process, while taking into account the average probability that those nodes are effectively informed. Nodes that are possibly informed by previous seeds selected are not computed to the efficiency of u. Thus, the algorithm may produce a seed set that has multiple seeds reaching the same node in S, i.e. consider nodes in $u_1, u_2 \in M$ that both have paths to node $v \in S$. This case is not undesirable since the chance of v being informed is increased. Possibly informed nodes are computed using Monte-Carlo simulation while considering as seed only the latest node added in the seed set.

The above function for seed selection can be proved to be monotonous and submodular, and thus approximates the best solution.

4.2.2 Dissemination of Information

Assuming a set of seed nodes M, our model propagates the emergency information in a number of steps. Let A_t be the set of nodes informed at step t. At the beginning, the seed nodes are the only ones informed and thus $A_0 = M$. At step $t+1$, every node $u \in A_t$ is able to inform each of its currently uninformed neighbors v and the probability that u informs v is given by the probability denoted by the weight of the edge (u, v), w_{uv}. Each node has a single chance of informing its currently uninformed neighbors about the event.

The above described procedures are summarized in algorithm 1.

Algorithm 1 Efficient Dissemination

```
S_1 → Directly connected to the event users
S_2 → Compute Indirectly Connected to the Event (S_1)
S → S_1 ⋃ S_2
R → get reachable nodes(G)
Seeds → ∅ //nodes that will act as propagation starters (Subset of R)
A → ∅ //nodes possibly informed by previously selected seeds
while (|Seeds| < k) do
   newseed → get most effective node(G, R, S, A)
   A → add possibly informed users(newseed)
return Seeds
```

5. EXPERIMENTAL EVALUATION

5.1 Experimental Setup

We have implemented our ESCAPE system and tested it with a real-world dataset.

Our Twitter dataset is composed of 513.449 tweets posted by 175.974 unique users in the city of Dublin for a four-month period (Dec 2012 to Mar 2013). We collected tweets using the Streaming API [7] of Twitter, where we applied a filter so that only tweets geographically located in Dublin are extracted. The filtering is based on the "Location" field set by the user and is either expressed as GPS coordinates or as part of the text of the tweet[23]. Tweets that were posted and had GPS location(Latitude, Longitude) provided by the devices' GPS sensor are also extracted. After tweets related to Dublin are extracted, users that posted tweets are gathered and by using the REST API [8] and the user ids, all tweets for each user are requested. A tweet has the following structure `<Tweet ID, User ID, UTC/GMT timestamp, Latitude, Longitude, ID of tweet replying, ID of user replying, Number of retweets, Source (iPad, Android), Text>`. From this structure, the User ID is used to obtain the screen name of the user, the UTC/GMT timestamps are used to compute the regularity metric and the Text is used to extract the users mentions with the '@' symbol. No further information related to the users is used. We do not consider any anonymization issue [5] since we assume that the authority that executes the system can be trusted (*e.g.*, the case of a campus social network). Twitter users can be classified into three major clusters as previously shown in [15]. These clusters are the broadcasters, the acquaintances and miscreants. Due to this asymmetry presented between users the graph that is formed is a directed weighted graph.

The experimental evaluation focuses on the following parameters: (i) Number of Informed Nodes from the users that are interested to be alerted for the event and (ii) Performance under different amount of Reachable users.

[7]https://dev.twitter.com/docs/streaming-apis
[8]https://dev.twitter.com/docs/api

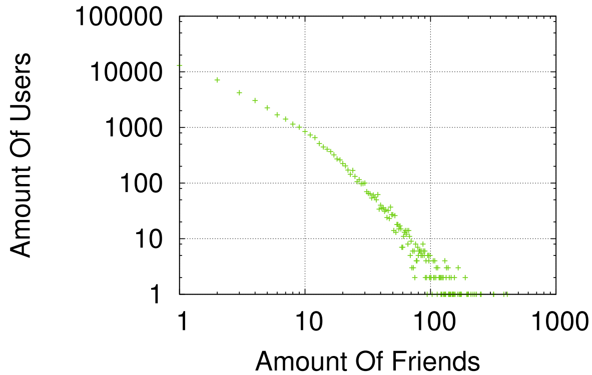

Figure 3: Friends of Users

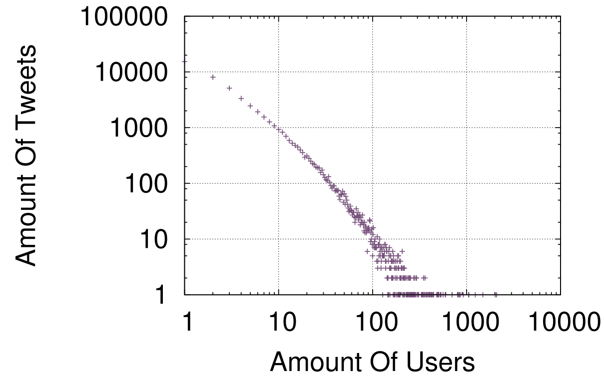

Figure 4: Tweets of Users

5.1.1 IRIE Overview

Jung et al. proposed IRIE[12], an algorithm that incorporates Influence Ranking and Influence Estimation for the influence maximization problem in the IC and IC-N (negative opinion propagation) model, which is proved to run faster than previous algorithms aiming to solve the problem of Influence Maximization (IM). They use a greedy algorithm for selecting the most influential nodes as in previous works on IM, but the process of estimating the influence integrates a system of linear equations whose solutions are computed iteratively. They compute influence rank of nodes, and for each node added to the seed set they compute the influence estimation of the seed set, using Monte-Carlo simulations for their experiments, though they note that other techniques for influence estimation can be deployed. For the weighting of the graph's edges the trivalency and the weighted cascade (WC) models are used.

5.2 Experimental Results

5.2.1 Twitter Data

We present Figures 3,4 to better understand the form of our Dataset in terms of number of users connected in the network and social relationships between users. In figure 3 we show the sets of users and their corresponding amount of friends. As can be observed from the figure, the majority of the users have friends that range from 1 to 100, while the amount of users with higher number of friends is small. In figure 4 the amount of tweets for the corresponding size of users is presented. This figure illustrates that the amount of tweets decreases as the amount of users increase and that only a few users have a great amount of tweets.

5.2.2 ESCAPE Evaluation

In this section we present the performance of our approach and we compare it with the state-of-the-art algorithm IRIE, which is the fastest algorithm in the literature that we know of and is able to perform influence maximization equally effectively to its competitors.

In the first set of experiments we present our approach when we set that 10% of the users are reachable (2371) out of the interested set (this corresponds to a set of 23710 nodes). In figure 5 we present the percentage of interested nodes that each approach is able to inform, relative to the users informed if all reachable nodes were selected as seeds. ESCAPE manages to inform more users than IRIE at all times, with a percentage that ranges from 8% to 13%.

In the second set of experiments we have the same setting but we consider that 20% of the users is reachable (this corresponds to 4742 users). Figure 6 presents the percentage of interested nodes that were informed and we see that the percentage has decreased. That is because when sending to all 20%, more interested nodes would be informed, compare to 10% since the size of reachable users set is doubled. The above results are relative to the users that would be informed if all reachable nodes were used as seeds. However, the percentage of users that ESCAPE manages to inform is more than 9 units percent over IRIE at all times. As can be observed, as the number of seeds goes up the gab between ESCAPE and IRIE becomes slightly smaller. That is due to the fact that the intersection of seeds sets for both algorithms contains more nodes. We expect that they will converge later when all reachable nodes are added as seeds. We add 50 seeds per iteration, so that is why the angle is presented in figures 5 and 6. When no seeds are selected, no users are informed. IRIE requires about 4427 seconds for determining 500 seeds while ESCAPE requires only 385 seconds when reachable modes are set to 20%, and 5158 seconds against 292 seconds respectively when reachable nodes are set to 10%. That makes ESCAPE more appropriate for emergency response. Due to the sparsity presented in the social graph, we notice a stabilized performance for both algorithms when more seeds are added.

6. RELATED WORK

Using social networks as dissemination tools has attracted interest in recent years in various application domains, including viral marketing campaigns and voting systems. In the majority of the applications the key point is influence maximization, *i.e.,* spreading the information to as many people as possible[13] or maximizing the likelihood that someone is being informed for a particular issue[10]. In all related works, the social network was represented as a weighted graph, with users as nodes and relationships between them as edges.

Several efforts have focused on inferring edge weights in so-

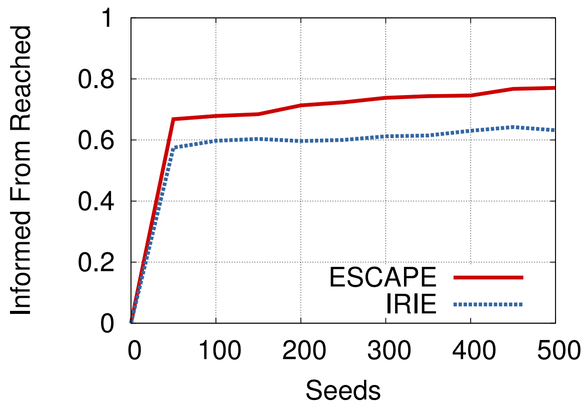

Figure 5: Informed From Reached - 10% Reachable

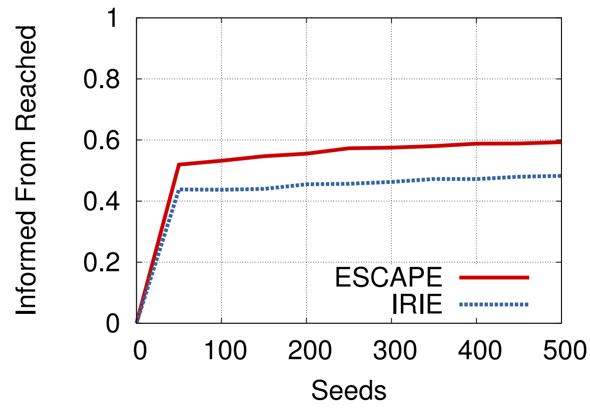

Figure 6: Informed From Reached - 20% Reachable

cial graphs. These take into account various information crawled by the social network related to users or the types of actions among the users. Recently, Facebook announced a metric called Edgerank, used to determine the items to populate at a users news feed. Edgerank considers the affinity of users, the actions made between users' profiles and time elapsed since the last interaction to determine the strength of the bond between different users and order news feeds according to the Edgerank metric[22]. Goyal *et al* [8] consider a number of models, including a Weighted Cascade (WC) model where the probability on an edge (v, u) is defined as $1/in - degree(u)$, a Trivalency model where probabilities are randomly and uniformly selected from the set {0.1, 0.01, 0.001}, and the model where probabilities learned from a training set using the Expectation Maximization (EM) based method suggested by Saito *et al* [20]. Unlike the Edgerank affinity, the number of users is not considered in this metric and different types of actions are not defined. Hangal *et al* in [10] suggest metrics for weighting the edges of a Social Graph so that the asymmetry presented in the network is captured. They also argue the fact that the shortest paths are eventually the strongest ones. They prove that utilizing edge weights may well improve global social search. Perhaps a more sophisticated metric that aims in predicting the strength of a relationship based on interaction and user similarity is established by Xiang et al. [25]. A latent variable model is considered in which the relationship strength forms the latent variable. An unsupervised model to estimate relationship strength from interaction activity and user similarity is developed. However, the model proposed requires data that may not be publicly available due to users privacy settings, thus, it is not applicable to all types of social networks.

Maximization of spread, given the weight of the edges already computed, is studied in [13]. The two models proposed are the Independent Cascade Model and the Linear Threshold model. The problem of influence maximization with the least effort (*i.e.*, using k or less nodes as starters), is stated as an NP-Hard problem, so provable approximation guarantees are obtained. In both models suggested by Kempe, nodes have two states, either active or inactive. When a node is active it means that the information has reached the node and the node is in position of forwarding it to its neighbors. They explore the above described models when a set of k initial nodes is being activated (targeted). It is proved that a greedy hill-climbing algorithm for both models gives a good approximation as long as the influence function has certain properties which, as proved, is true. Influence probability and propagation is also a case of study in [3].

In [24] they propose variations for the independent cascade model so as to minimize computational cost. Paths whose probabilities is below a given threshold θ are excluded, with θ being tunable. Contrary to our approach, these efforts (1) assume that the influence graph is known and aim at maximizing the influence in the entire network rather than identifying and informing an appropriate subset of nodes most interested to the event, and (2) they focus on cases of viral marketing campaigns or voting systems, rather than emergency response situations.

Gomez-Rodriquez and Schölkopf [19] study the efficient influence maximization in time diffusing networks. They consider the influence maximization problem where information or influence can spread at different rates across different edges and analytically compute and efficiently optimize the influence avoiding the use of heuristics. The greedy algorithm in combination with the Lazy Evaluation, Localized Source Nodes and Limited Transmission Paths are proposed as speed-ups for the computations.

Kyungabaek *et al* [14] define two types of correlated to an event users. The ones located to the area of the event and those that have social ties with them. The notification system proposed for alerting users in case of an emergency event is aware of the geographies the message needs and social ties. Their system has a prior knowledge of Global Target Geography (GTG). Nodes are classified as physical nodes (PNodes) and trusted physical nodes (T-PNodes) that represent some sort of authority or public figure. After the occurrence of an event Possibly Affected Region (PAR) is defined as a sub-region of GTG and Possibly Damaged Region (PDR) as sub-region of PAR. They define two types of overlays. The Delivery Overlay which aims in reaching PNodes and the Information Overlay which is responsible for maintaining information about social entities. They customize the social diffusion process so that good initiators are selected.

7. CONCLUSIONS

In this paper, we have presented our ESCAPE system that investigates the relationships and interactions among the members of a social group, and develops a dissemination mechanism to maximize the information reach to a target set of users, when an emergency event occurs. As we illustrate in our experimental evaluation ESCAPE with its intelligent seed node selection process, manages to inform more interested users than the state-of-the-art technique IRIE, that aims in Influence Maximization.

Acknowledgment

This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Programs: Aristeia-MMD Aristeia-Inception Investing in knowledge society through the European Social Fund, the FP7 INSIGHT project and the ERC IDEAS NGHCS project.

8. REFERENCES

[1] F. Alvanaki, S. Michel, K. Ramamritham, and G. Weikum. See what's enblogue: real-time emergent topic identification in social media. In *EDBT*, pages 336–347, Berlin, Germany, March 2012.

[2] J. P. Bagrow, D. Wang, and A.-L. Barabasi. Collective response of human populations to large-scale emergencies. *CoRR*, abs/1106.0560, 2011.

[3] F. Bonchi. Influence propagation in social networks: A data mining perspective. *IEEE Intelligent Informatics Bulletin*, 12(1):8–16, 2011.

[4] A. Boutet, D. Frey, R. Guerraoui, A. Jégou, and A.-M. Kermarrec. Whatsup decentralized instant news recommender. In *IPDPS*, Boston, MA, May 2013.

[5] S. Das, Ö. Egecioglu, and A. El Abbadi. Anónimos: An lp-based approach for anonymizing weighted social network graphs. *IEEE Trans. Knowl. Data Eng.*, 24(4):590–604, 2012.

[6] P. Eugster, P. Felber, and F. Le Fessant. The "art" of programming gossip-based systems. *SIGOPS Oper. Syst. Rev.*, 41(5):37–42, Oct. 2007.

[7] R. Friedman, D. Gavidia, L. Rodrigues, A. C. Viana, and S. Voulgaris. Gossiping on manets: the beauty and the beast. *SIGOPS Oper. Syst. Rev.*, 41(5), Oct 2007.

[8] A. Goyal, F. Bonchi, and L. V. Lakshmanan. A data-based approach to social influence maximization. volume 5, pages 73–84, Seattle, WA, August 2011.

[9] M. Guy, P. Earle, C. Ostrum, K. Gruchalla, and S. Horvath. Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. In *IDA*, pages 42–53, Tucson, AZ, 2010.

[10] S. Hangal, D. MacLean, M. S. Lam, and J. Heer. All friends are not equal: Using weights in social graphs to improve search. In *SNA-KDD*, Washington, DC, July 2010.

[11] M. Iqbal, X. Wang, D. Wertheim, and X. Zhou. Swanmesh: A multicast enabled dual-radio wireless mesh network for emergency and disaster recovery services. *JCM*, 4(5):298–306, 2009.

[12] K. Jung, W. Heo, and W. Chen. Irie: Scalable and robust influence maximization in social networks. In *ICDM*, pages 918–923, Brussels, Belgium, December 2012.

[13] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *SIGKDD*, pages 137–146, Washington, DC, USA, August 2003.

[14] K. Kim, Y. Zhao, and N. Venkatasubramanian. Gsford: Towards a reliable geo-social notification system. In *SRDS*, pages 267–272, Irvine, CA, October 2012.

[15] B. Krishnamurthy, P. Gill, and M. Arlitt. A few chirps about twitter. In *WOSN*, pages 19–24, Seattle, WA, USA, August 2008.

[16] A. M. Lesperance, M. A. Godinez, and J. R. Olson. *Social networking for emergency management and public safety.* Pacific Northwest National Laboratory Richland, WA, 2010.

[17] B. Liu, G. Cong, D. Xu, and Y. Zeng. Time constrained influence maximization in social networks. In *ICDM*, pages 439–448, Brussels, Belgium, December 2012.

[18] J. J. Ordille, P. Tendick, and Q. Yang. Publish-subscribe services for urgent and emergency response. In *COMSWARE*, pages 8:1–8:10, Dublin, Ireland, June 2009.

[19] M. G. Rodriguez and B. Schölkopf. Influence maximization in continuous time diffusion networks. In *ICML*, Edinburgh, Scotland, June 2012.

[20] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *KES*, pages 67–75, Zagreb, Croatia, Sep 2008.

[21] J. Singh and J. Bacon. Event-based data control in healthcare. In *Middleware*, pages 84–86, Leuven, Belgium, December 2008.

[22] D. Taylor. Everything you need to know about facebookâĂŹs edgerank. *The Next Web*, 9, May 2011.

[23] G. Valkanas and D. Gunopulos. Location extraction from social networks with commodity software and online data. In *ICDMW*, Brussels, Belgium, December 2012.

[24] C. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012.

[25] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *WWW*, pages 981–990, Raleigh, NC, USA, April 2010.