

Seeding Strategies in Online Social Networks for Improving Information Dissemination of Built Environment Disruptions in Disasters

Chao Fan¹; Yucheng Jiang²; and Ali Mostafavi, Ph.D.³

¹Ph.D. Student, Zachry Dept. of Civil Engineering, Texas A&M Univ., College Station, TX 77843-3136. E-mail: chfan@tamu.edu

²B.S. Student, Dept. of Computer Science and Engineering, Texas A&M Univ., College Station, TX 77843-3136. E-mail: jiang50@tamu.edu

³Assistant Professor, Zachry Dept. of Civil Engineering, Texas A&M Univ., College Station, TX 77843-3136. E-mail: amostafavi@civil.tamu.edu

ABSTRACT

The objective of this study is to propose a seed-search algorithm and develop seeding strategies in online social networks for disseminating credible situational information regarding built environment disruptions in disasters. Rapid and extensive dissemination of credible situational information is important for disaster preparedness, response, and recovery in communities. Online social networks such as Twitter have become popular media sources among the public to share information in disasters. Due to the directed relations and fragmentations in networks, however, little is known about the ways of selecting starting nodes (a.k.a., seeds) in order to broadly and rapidly spread information online. To address this gap, this study proposes a computational approach, which is an integration of greedy algorithm and graph analysis, to capture fragmentations, identify critical seeds, and develop seeding strategies to disseminate information in online social networks. A case study of an infrastructure disruption (water release from reservoirs in Houston) during 2017 Hurricane Harvey was used to illustrate the capabilities of the proposed approach. The results indicate that seeding top 10 users is effective for distributing information to more than 80% of nodes in a network. The findings inform about strategies to better report, transmit, and gather situational information on social media, which can further enhance situation awareness and community resilience in disasters.

INTRODUCTION

Built environment disruptions happen abruptly and evolve rapidly in disasters (Zhu & Mostafavi, 2018). Massive and rapid spread of credible situational information during built environment disruptions is significantly important to disaster response and community resilience (Fan et al., 2018). Online social networks such as Twitter, Facebook, and Instagram are emerging media to prompt information dissemination in extreme events. Generally, online social networks are directed networks, which means the direction of sharing information from one node to another is irreversible unless they are following each other (Kim & Hastak, 2018). The irreversibility of users' relations also leads to the fragmentations in online networks where a node from one group cannot distribute its information to a node in another group. Hence, it is impossible to disseminate credible information throughout the whole network by randomly selecting starting nodes (a.k.a., seeds) if fragmentations exist in the network. To maximize the speed and magnitude of spreading credible situational information, thus, identifying the fragmentations and strategizing the seeds in online social networks is essential in built environment disruptions.

To this end, existing studies have attempted to develop computational approaches and

seeding strategies in online social networks (Chin et al., 2018). Some studies focus on one-hop targeting, in which the seeds are selected by selecting the highest in-degree nodes and randomly selecting one of their neighbors (Kim et al., 2015). These studies hypothesized that the online social network is a strongly connected component, in which every node can get access to the rest of the nodes in the network (Shakya et al., 2017). However, as discussed earlier, in some cases, online social networks are sparse, and the connections between different components are irreversible. Hence, the seeds in one component may not be able to deliver the information through the entire network. Another stream of research focuses on the friendship networks and the structural properties by applying social contagion models (Sinan Aral & Dhillon, 2018; Sela et al. 2015). These studies hypothesized that information and influence can spread through the established friendship based on threshold values and did not consider the presence of fragmentations in large-scale social networks. However, network fragmentation can make a drastic impact on the paths of information propagation and the accessibility of online users. Meanwhile, the fragmentation is also a result of the variance of interests among users in networks. On social media, people tend to retweet others as a form of social endorsement and common interests. Therefore, approaches to analyze the retweeting network of localized people who are affected by disasters are needed, and the outcomes can provide reliable evidence to develop seeding strategies for prompt situational information dissemination on social media.

This paper presents a framework built upon retweeting networks related to built environment disruptions to identify the seeds by taking the percentage of reachable users, neighbors' overlapping, and depth of cascades into account. Through the proposed framework, we identified critical seeds who play a primary role in spreading situational information and examined the robustness of the proposed approach in a six-day disaster and investigated the evolution of the seed sets.

SEED-SEARCH ALGORITHM

The development of seeding strategies relies on the characteristics of the seeds and their social networks. There are two characteristics that are defined by existing studies to evaluate the capability of seeding strategies: (1) the number of reachable users; and (2) the size of the seed sets (Sinan Aral & Dhillon, 2018). As such, we can convert the seeding strategy problem to a constrained extreme value problem, which can be defined as follows:

Given a directed online social network $G = (V, E)$ where V is the set of active users, and E is the set of edges, which represent the retweeting behaviors among the active users. Each node $u \in V$ has posted or retweeted at least one tweet about the disruption, water release from reservoirs. The problem asks to find a seed set $S \subset V$ for a certain network G with $\min|N|$ and $\max|N|$, where N is a set of reachable users by the seeds in S . The problem of searching seeds in a directed network is NP-hard (Ahmad & Ali, 2017). The computational cost is exponentially proportional to the size of the network. To reduce the computational complexity, this study is only interested in top seeds which can reach more than 80% of the active users.

To address this problem in online social networks, we adopted a greedy algorithm (Das, 2017) integrated with pre-defined criteria (see Algorithm 1). The searching approach starts with a user who has the highest in-degree. To take the dissemination efficiency into account, meanwhile, the algorithm computes the depth of cascades that a seed can access. Imagining that two seeds can reach the same number of users, the deeper of the cascades that the seed has, the lower the dissemination efficiency will be. Then, we iterate the computational process to identify the seeds and cumulate the number of reachable users. The iteration criterion is that the next seed

should maximize the cumulative number of reachable users in $|N|$, and the $v \in N$ cannot be the seeds. This criterion can ensure that the reachable nodes of seeds $s \in S$ do not overlap, and subsequently the dissemination process can converge in an efficient way. Based on this criterion, we do the iterations until the cumulative number of reachable nodes is greater than 80% of $|V|$.

Algorithm 1: Seed-search algorithm (SSA)

Input: $G = (V, E)$
Output: A set $S \subseteq V$ so nodes in S can reach all the nodes in G .

```

1 visit_list  $\leftarrow []$ 
2 Max_visit  $\leftarrow 0$ 
3 Min_step  $\leftarrow \text{INT\_MAX}$ 
4  $S \leftarrow \text{set}()$ 
5 for each  $u \in V$ 
6    $visit, u \leftarrow \text{Find\_reachable}(G, u)$ 
7    $visit\_list.add([visit, u])$ 
8  $visit \leftarrow \text{set}()$ 
9 while  $|visit| < |V|$  do
10    $node \leftarrow \text{Find\_next}(visit, visit\_list)$ 
11    $S \leftarrow S \cup [node]$ 
12 return  $S$ 

```

```

13 Function Find_reachable( $G, u$ )
14  $visit \leftarrow \text{set}()$ 
15  $Q \leftarrow \emptyset$ 
16  $\text{ENQUEUE}(Q, u)$ 
17 while  $|Q| \neq 0$  do
18    $adj \leftarrow \text{adjacent nodes of } \text{DEQUEUE}(Q)$ 
19    $visit \leftarrow visit \cup \text{DEQUEUE}(Q)$ 
20   for each  $a$  in  $adj$ 
21     if not in  $visit$  then  $\text{ENQUEUE}(Q, a)$ 
22 return  $visit, u$ 
23 Function Find_next( $visit, visit\_list$ )
24  $gain\_margin \leftarrow 0$ 
25 for each  $t$  in  $visit\_list$ 
26    $gain \leftarrow |visit \cup t.visit| - |visit|$ 
27   if  $gain > gain\_margin$  then  $gain\_margin \leftarrow gain$ 
28 return  $t.node$ 

```

CASE STUDY

To illustrate the capabilities of this algorithm for strategizing seeds in online social networks, we conducted a case study of information dissemination on Twitter during water release from reservoirs in the west side of Houston in 2017. In this study, we investigated the event associated with Barker and Addicks reservoirs during Hurricane Harvey and their effects on nearby neighborhoods and other localized users in Houston. Water release from reservoirs is a built environment event that was induced by Hurricane Harvey. Due to the sudden heavy rainfall, the water levels in both reservoirs reached their maximum capacity, which led to water release and flooded some nearby neighborhoods. This event followed the damages of Hurricane Harvey and lasted a couple of days (August 27th to September 1st) (Olsen, 2018).

To investigate the dissemination of situational information on Twitter, we collected all of the tweets, which was around 21 million, over the Houston area from August 22nd to September 30th (Fan et al., 2019). The rules for filtering the tweets included multiple bounding boxes in Houston and users' profiles with a location of Houston. Based on those rules, the tweets we collected were all posted by localized users so that they were reliable for the analysis on understanding the impacts of built environment disruptions on residents. In this study, we extracted 6991 tweets posted or shared during the water release and relevant to this event. As shown in Figure 1, the directed social networks mapped the posting and retweeting behaviors of online users in each day. Due to the rapid changes in the built environment and its influence on the public, the retweeting networks are varied. For example, the networks on August 29th and August 30th are much denser than the other networks. Much more users were active during these two days (2417 users on August 29th and 1617 users on August 30th). That is because the reservoirs reached their maximum capacities on these two days. More water was released from those

reservoirs and it led to severe damages in the nearby neighborhoods. After that, the hurricane past and the rain stopped. Floods gradually receded from the neighborhoods on August 31st and September 1st. In addition, August 27th was the first day when Harris County started to release water from reservoirs. Therefore, there are some users with a large number of in-degrees due to their distribution of the first-hand information on August 27th.

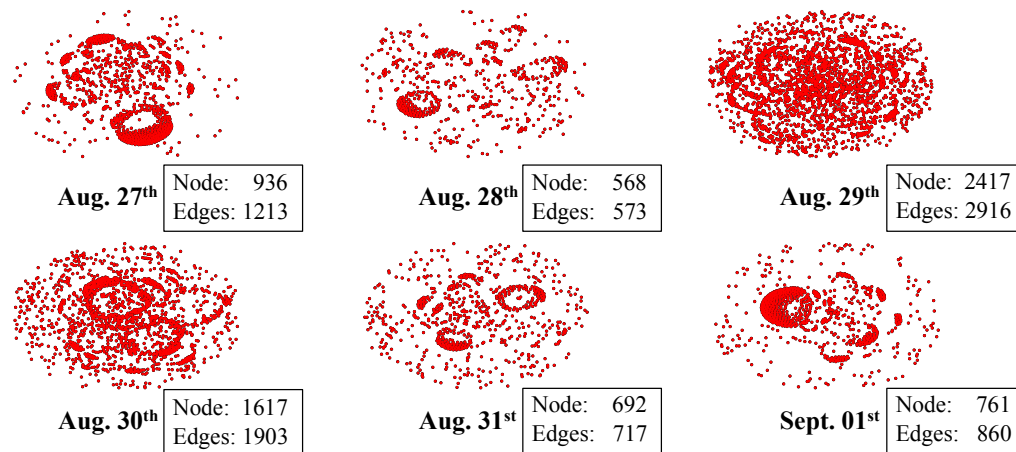


Figure 1. Retweeting networks among localized online users.

Once the retweeting networks were constructed, we applied the proposed seed-search algorithm to those networks and obtained six sets of top seeds which can reach at least 80% of active users in these networks. Figure 2 shows the cumulative percentage of reachable users in six online social networks by expanding the seed sets. Due to the directions of retweeting behaviors, the information starting from one user cannot be disseminated through the entire network. As a result, the contributions of different seeds to disseminating information through a network vary. As shown in Figure 2, the slopes of the cumulative reachable nodes in these six days are skew in the first ten seeds and then become flat when continuously adding more seeds. Therefore, to minimize the efforts on seeding online users, in each daily online social network, seeding 10 top users to spread the situational information can reach most of the online users. In addition, at the beginning and the end of the disruption, the top one seed can reach more than 50% of the users in the networks, while they can only reach around 30% of the users during the disruption. During the built environment disruption, the top one seed is not the only user that can gather and report real time situational information. Many other users are active in posting and sharing their observations on social media. Therefore, the top one seed can reach fewer users but still important for credible information dissemination in online social networks during the built environment disruption. For example, the local news media with most reachable residential users on Twitter can be the top one seed. Other accounts such as a meteorological department, police department and journalists who have their own massive audiences also play the role of top seeds in distributing situational information. Capitalizing on this finding, the public who want to obtain real-time information in disasters only need to pay close attention to such ten seeds rather than all possible online users. This practical suggestion can save users much effort on gathering information in a time-sensitive context, and subsequently improve their response efficiency in disasters.

To better distinguish the contributions of the seeds, we analyzed the relationship between the cumulative percentage of reachable nodes and the number of reachable nodes of a seed. As

shown in Figure 3, the rank of the seeds is created based on their contribution to the cumulative percentage of reachable nodes in networks. Note that, in some networks such as August 27th, August 29th and August 30th, some seeds with a large number of reachable nodes are out of the top ten. In particular, there are three users who can reach more than 600 users in the social network on August 29th. However, except the top one seed, the others are at 15th and 28th places respectively. That is because there is a significant overlap of reachable users among these three seeds. Therefore, when one seed can reach the overlapped group of users, the other seeds would have less contribution to increase the total number of reachable users in the network. In the context of built environment disruptions, the number of users who can get access to the situational information is important to effective response actions. Identifying the overlap of the reachable users among the seeds can better seed the users to spread information broadly in online social networks. For example, a residential user who wants to report a road damage only needs to mention or send direct messages to a few top seeds who do not have much overlap in their reachable users. This practical way of reporting situation will benefit the efficiency of spreading information and also reduce the information redundancy in only a few groups of users.

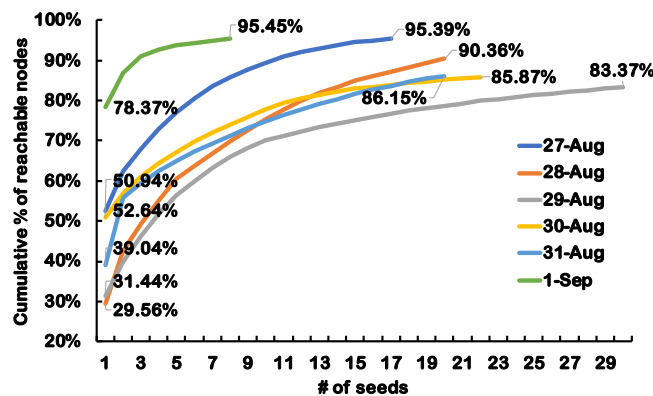


Figure 2. Cumulative percentage of reachable users in online social networks

In addition, Figure 3 also indicates that the seeds such as seed 1, 15, and 28 on August 29th, who have a great number of reachable users do not connect to each other. Even though these seeds have overlaps of reachable users, they did not share or retweet the information among each other. This is a common social problem, “Network Fragmentation” (Chami et al., 2017), in most of the retweeting networks. The presence of fragmentations is a cause of disparities in situation awareness and inconsistency in response actions. For example, one group of users can get access to the real-time weather information, while another group of users has the information of inundation level in reservoirs. Due to lack of connections between these two groups, the users in different groups may have different understandings of the coming risks to their lives and properties, and consequently, they would take different actions such as evacuation or staying at home. Such disparities will induce severe losses to the people who take wrong actions. The seed-search algorithm presented in this paper can benefit the identification of fragmentations in online social networks and provide evidence to address this issue. For example, we can inform the seeds who have a similar number of reachable nodes in networks to follow each other and share situational information during built environment disruptions. Adding connections among these seeds cannot only reduce the fragmentations in networks, but also improve the efficiency of information distribution.

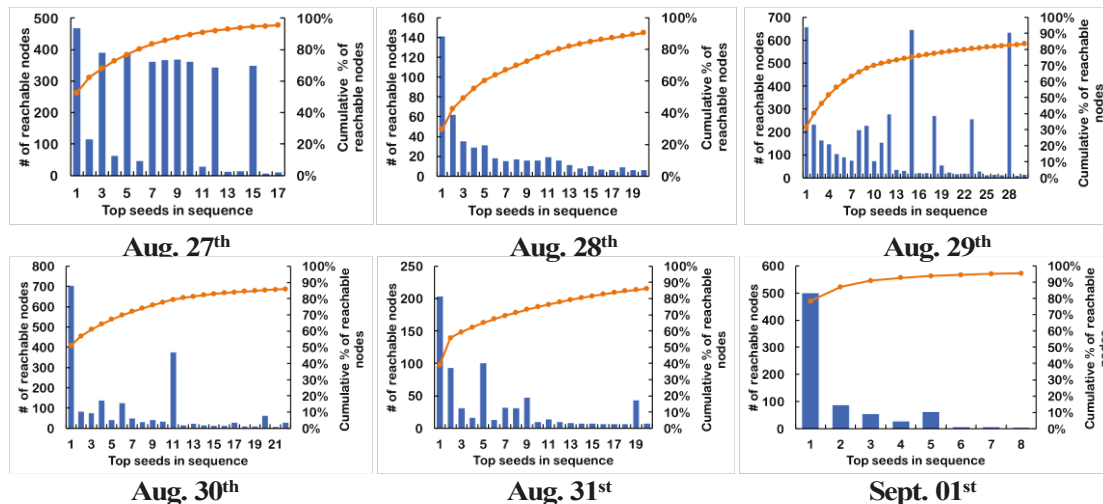


Figure 3. Top seeds and their number of reachable users in online social networks

The depth of cascades is an important indicator to quantify the capability of a seed to reach other users in networks. Figure 4 shows the depths of the cascades of the top eight seeds in each network. Obviously, the top-one seeds have deeper cascades than the others. The cascade depth of other seeds is only one in most cases. However, some seeds such as seed 3, 5, 7, and 8 on August 27th also have more than 4 cascades. A seed with a depth of 2 means that information from the seed can spread to a cascade in two steps, and a seed with a depth of 4 means the information from the seed can spread to a cascade in four steps. The more steps the seed have, the more influential the seed is. Therefore, the seed with a depth of more than 4 are potential disseminators who can broadly spread information in networks. These results support the implications in Figure 4. Thus, the findings here can validate the reliability of the proposed algorithm in identifying and strategizing seeds in online social networks for disseminating credible situational information in disasters.

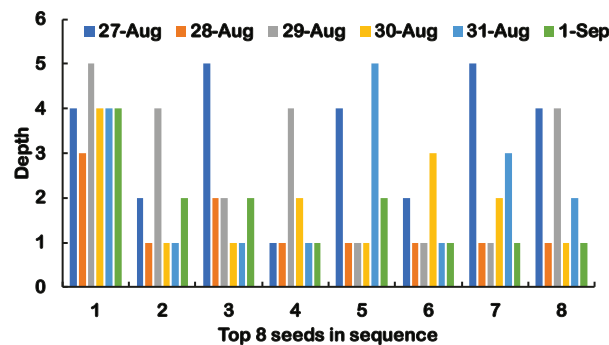


Figure 4. Depth of cascades that the top seeds can reach

CONCLUDING REMARKS

This paper presents a computational approach to identify seeds and develop seeding strategies in online social networks regarding built environment disruptions in disasters. The results of a case study of water release from reservoirs during the 2017 Hurricane Harvey implies three important seeding strategies: first, seeding 10 top users to spread credible information is efficient for reaching more than 80% of users in a network; second, identifying overlaps of

reachable nodes among the seeds can help users spread the information more extensively; third, the proposed approach can capture the fragmentations in social networks and help users to reduce the fragmentation and spread information more efficiently.

The proposed seed-search algorithm provides a new computing approach to promote information spreading by seeding influential users. Considering the overlaps of reachable users among the seeds, the strategies presented in this paper can enable infrastructure managers, disaster managers, and public officials to effectively identify users to maximize the diffusion of information on social media regarding built environment disruption events that affect residents. The use of these findings in educating online users can enhance the adoption of situational information, the response to risks, and the resilience of communities in ever-growing disasters. This work can be further extended in: (1) examine the relationship between the priority of the seeds and their features, such as number of followers, timeliness of the posted information, and verification of their profiles; and (2) interpret the temporal topology of retweeting networks as an indicator of community resilience to infrastructure disruptions.

ACKNOWLEDGEMENT

This material is based in part upon work supported by the National Science Foundation under Grant Number IIS-1759537. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- Ahmad, W., & Ali, R. (2017). "A framework for seed user identification across multiple online social networks." *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017, 2017-January* 708–713.
- Aral, S., & Dhillon, P. S. (2018). "Social influence maximization under empirical influence models." *Nature Human Behaviour*, 2(June), 1–8.
- Chami, G. F., Ahnert, S. E., Kabatereine, N. B., & Tukahebwa, E. M. (2017). "Social network fragmentation and community health." *Proceedings of the National Academy of Sciences*, 114(36), E7425–E7431.
- Chin, A., Eckles, D., & Ugander, J. (2018). "Evaluating stochastic seeding strategies in networks." 1–42. Arxiv
- Das, S. (2017). "Seed Node Distribution for Influence Maximization in Multiple Online Social Networks." *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing*, 1260–1264.
- Fan, C., Mostafavi, A., Gupta, A., & Zhang, C. (2018). "A System Analytics Framework for Detecting Infrastructure-Related Topics in Disasters Using Social Sensing." In I. F. C. Smith & B. Domer (Eds.), *Advanced Computing Strategies for Engineering* (pp. 74–91). Cham: Springer International Publishing.
- Fan, C., Mostafavi, A., Yao, W., & Huang, R. (2019). "A Graph-based Approach for Detecting Critical Infrastructure Disruptions on Social Media in Disasters." In *2019 52th Hawaii International Conference on System Sciences (HICSS)* (pp. 1975–1984).
- Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H., & Christakis, N. A. (2015). "Social network targeting to maximise population behaviour change: A cluster randomised controlled trial." *The Lancet*, 386(9989), 145–153.
- Kim, J., & Hastak, M. (2018). "Social network analysis: Characteristics of online social

- networks after a disaster.” *International Journal of Information Management*, 38(1), 86–96.
- Olsen, L. (2018). “Record reservoir flooding was predicted even before Harvey hit Houston.” *Houston Chronicle*. Retrieved June 5, 2018, from <https://www.houstonchronicle.com>
- Sela, A., Ben-Gal, I., Pentland, A. S., & Shmueli, E. (2015). “Improving Information Spread through a Scheduled Seeding Approach.” In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.
- Shakya, H. B., Stafford, D., Hughes, D. A., Keegan, T., Negron, R., Broome, J., ... Christakis, N. A. (2017). “Exploiting social influence to magnify population-level behavior change in maternal and child health: Study protocol for a randomized controlled trial of network targeting algorithms in rural Honduras.” *BMJ Open*, 7(3), 1–13.
- Zhu, J., & Mostafavi, A. (2018). “Enhancing Resilience in Disaster Response: A Meta-Network Analysis Approach.” *Construction Research Congress 2018*, (1), 2250–2259.