Chapter 1

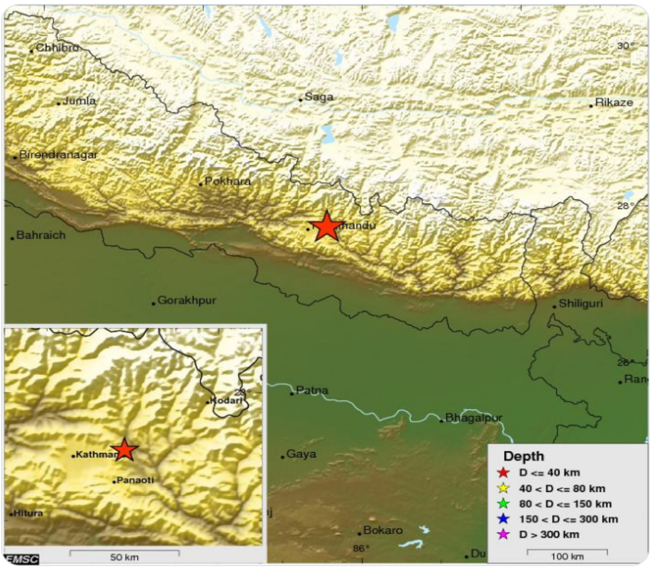
Introduction

1.1 Background and Motivation

With the characteristics of rapid dissemination and instant communication, social network platforms have become a popular tool to spread real-world events [46]. These events may hold critical information, but do not receive enough public attention. In such cases, the propagation of these events has significant social value. For example, disaster events involve identifying injured and missing people, assessing medical and nutritional needs, and their dissemination facilitates relief coordination, which is vital to reducing injuries, loss of life, and economic damage in affected areas [5]. Public health events provide information on the symptoms and high-risk populations of diseases. Promotion of them increases participation in preventive programs and medical treatments, ultimately saving lives, lowering healthcare costs, and strengthening community resilience [38]. Similarly, cybersecurity events educate people about various types of cyber threats and effective response practices. Their broadcast raises public awareness of sensitive data protection and helps mitigate financial losses at the personal, corporate, and national levels [4].

All the events mentioned above are inherently complex as each contains a series of sub-events. A sub-event describes a specific aspect of an event, and as the event evolves, new sub-events continue to emerge from the social stream. Twitter (now X) as an important communication medium during Nepal Earthquake 2015 has captured the event's complexity in the tweet stream. For example, Figure 1.1 presents two tweets from different stages of the earthquake. The tweet in Figure 1.1a reported that an aftershock of magnitude 5.3 struck 24 kilometers east of Kathmandu, Nepal, 10 minutes before tweeting. The tweet in Figure 1.1b depicted that a user was distributing food to earthquake victims in Nepal. These two tweets can be regarded as two sub-events, each with its own distinct content and context. They are causally linked, as a major earthquake triggers aftershocks, leading to resource shortages that, in turn, prompt relief efforts. Their temporal sequence reflects the event's evolution, transitioning from real-time monitoring to relief efforts.

Felt #earthquake M5.3 strikes 24 km E of #Kathmandu (#Nepal) 10 min ago. Please report to: emsc-csem.org/Earthquake/ear...



7:06 PM · Apr 25, 2015 from Nepal

(a) A tweet reporting an aftershock's occurrence (source by: https://x.com/LastQuake/status/591890923920949248)

Random Act of G,distributing foods, Earthquake victims Nepal@brucepoontip @G_JulieFitzG @Manishsinghvns @shiva_nepal

5:42 PM · May 4, 2015 from Nepal

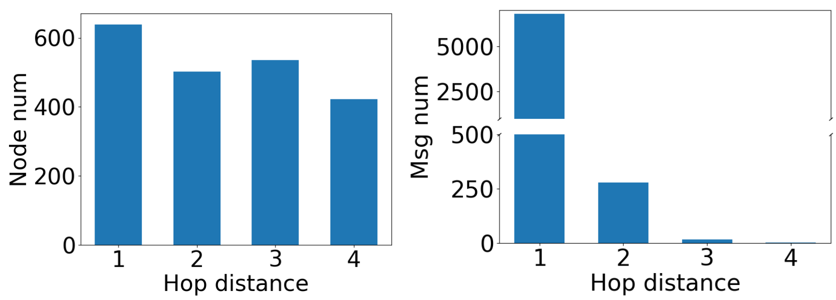
(b) A tweet showing food distribution for victims (source by: https://x.com/PuskarGAdv/status/595131367668613120)

Figure 1.1: Two tweets about Nepal Earthquake 2015

Although much attention has been paid to the analysis of complex events [21, 20, 72, 71], strategies for their effective dissemination have not been widely studied. The spreading of a complex event fosters user engagement, extending information propagation in cyberspace to prompting real-world actions. For example, engagement with the “aftershock” sub-event enables individuals to make informed decisions to avoid damage from aftershock, while engagement with the “food distribution” sub-event encourages people to donate food or assist in distribution efforts. As such, maximizing user engagement with an event is significant to make a real-world impact. Considering each engaged user is active for the event, the number of engaged users in the social network aligns with the concept ”influence”. Consequently, our research interest lies in influence maximization (IM), an area extensively studied in viral marketing [25], where companies leverage social connections to spread the adoption of a new product from initially selected adopters. However, these studies are not adaptive to maximize the influence of a complex event. They model information diffusion as the dissemination of a piece of information through interactions among users in a network and evaluate its influence by using black-box-style Monte Carlo simulations. This simplified diffusion pattern and evaluation method encounters numerous challenges when applied to a complex event.

The first challenge is that previous methods do not recognize a sub-event as a component subordinate to a specific event. These methods are based on diffusion simulations of a piece of information to evaluate the influence spread of this information. As for a complex event, they only work for maximizing each sub-event’s influence independently, which is impractical for spreading the influence of the whole event. Given that an event may comprise thousands of sub-events, independent propagation of each sub-event incurs significant computational overhead. Besides, this approach fails to address the overlap in sub-events’ influence, potentially leading to multiple sub-events targeting the same large group of users. This neglects the possibility that a user’s willingness to engage with an event diminishes as the number of sub-events with which they have already interacted increases.

The second challenge is that previous methods do not account for users’ dynamic sensitivity to sub-events as the event evolves. Users’ sensitivity to sub-events involves the relationships between them across dimensions of text, location, and time. Among previous methods, topic-aware IM [14], location-aware IM [44] and time-aware IM [33] have each attempted to exploit users’ relationship with propagating information at the levels of text, location, and time, respectively. However, none of them considers scenarios in which the text or location of the propagating information evolves over time. For event propagation, this evolution results in the change of target users for an event. For example, during the response stage of Nepal Earthquake 2015, the representative sub-event is the aftershock tweet illustrated in Fig. 1.1a and users located in Nepal are more responsive to the event. During the recovery stage, the representative sub-event is the food distribution tweet presented in Fig.



(a) Node distribution. (b) Message distribution.

Figure 1.2: Simulation of users' information access in Nepal Earthquake 2015. The social network is modeled as a directed graph, where the initial information publishers are named as seeds, and the remaining users are referred to as nodes. "Hop distance" is the minimum distance among all possible hop distances from each seed to a node in the social graph; "Node num" is the number of nodes at a specific hop distance from these seeds; "Msg num" is the number of messages received by nodes at a given hop distance from the seeds.

1.1b and users with donation intentions are more likely to be encouraged by the event, while the location of them becomes less significant. Capturing this evolution enables the update of target users for an event within a specific time period, thereby facilitating broader influence spread.

The third challenge is the inequality in individual message reception during event propagation. Treating each piece of information relayed between users as a message, and given the diverse information within a complex event, the spread of such events across social networks leads to significant inequity in individual message reception. On one hand, users with good information access face message bombardment, which is offensive and might pressure people to engage with the event. On the other hand, users who lack information access encounter message isolation, meaning they can barely receive messages. This phenomenon is demonstrated by the simulation of event propagation in Fig. 1.2. The social graph is constructed using the data from 15-24 April, with the propagation probabilities generated by the commonly used trivalency model [17]. 10 sub-events about Nepal Earthquake 2015 are disseminated, each independently published by a set of 50 users randomly selected from the top 1% nodes with the highest out-degrees in the graph. The propagation of each sub-event is simulated using the Independent Cascade (IC) model [40] via 1000 Monte Carlo simulations. Fig. 1.2a illustrates the distribution of node numbers at different hop distances from the seed set to a node. Fig. 1.2b presents the distribution of message numbers received by nodes at different hop distances from the seed set. "Hop distance" is employed to quantify the difficulty of a user accessing information, where a shorter distance indicates closer social relationships with the information publishers, implying easier information access. As shown in Fig. 1.2b, as the hop distance increases, the number of messages reduces drastically, leading to highly unbalanced message distribution. This message distribution is not only unfair, but also hinders influence spread. As depicted in Fig. 1.2a, the node numbers at hop distances of 2, 3 and 4 are 79%, 84% and 66%, respectively, of the node number at the hop distance of 1. However, the message numbers at hop distances of 2, 3 and 4 are just 5%, 0% and 0%, respectively, of the message number at the hop distance of 1. The disproportion between node numbers and message numbers reveals significant potential of increased influence among nodes not directly connected to the seeds, as insufficient messages have been delivered to them. There are already a series of fair IM works addressing the disproportionate exclusion of influence for individuals from racial minorities or those identifying as LGBTQ. Such exclusion occurs because conventional IM techniques frequently overlook smaller groups that contribute less to overall influence. As a result, these works prioritize a balance between achieved influence and fair influence allocation for user groups, rather than solely maximizing influence [64, 3, 55]. However, the division of users based on sensitive information is unnecessary for the spread of complex events. Especially considering the propagated events in our scenario contain critical information, users should receive relevant sub-events to help themselves or others in critical situations. This relevance depends on user's situational needs or capabilities rather than inherent sensitive features. For example, for the spread of earthquake event, users' age or gender information is not important compared with their geolocation. Additionally, Fig. 1.2 has illustrated unequal information access among users, which theoretically hinders broader influence spread. Therefore, we focus on addressing the disparity of influence received by individual users caused by unequal message reception.

The last challenge is the visualization of event propagation paths. Existing IM works typically present their influence spread results as the expected number of influenced users, which is approximated by inputting a seed set and a diffusion network into a diffusion model and performing Monte Carlo simulations within that model. In each simulation, the diffusion model outputs the number of influenced users, and the influence spread is calculated as the average number of influenced users for these simulations. This kind of influence score overlooks details about the diffusion process, which are essential to reveal the differences between event propagation and other information propagation patterns. Existing fair IM works evaluate fairness using a predefined equation to convert influence spread results into a fairness score. However, given the abstract nature of fairness concepts, it is difficult for individuals without specialized knowledge of fairness to comprehend the implications of a fairness score. To deliver an accurate and comprehensible description of the information diffusion process and fairness, it's worthwhile to investigate how to visualize fair event-aware influence spread.