CHASSIS: Conformity Meets Online Information Diffusion

Hui Li
School of Computer Sc. and Engg.,
Nanyang Technological University,
Singapore
HLI019@e.ntu.edu.sg

Hui Li
School of Cyber Engineering,
Xidian University, China
hli@xidian.edu.cn

Sourav S Bhowmick
School of Computer Sc. and Engg.,
Nanyang Technological University,
Singapore
assourav@ntu.edu.sg

ABSTRACT

Online information diffusion generates huge volumes of social activities (*e.g.,* tweets, retweets posts, comments, likes) among individuals. Existing information diffusion modeling techniques are oblivious to *conformity* of individuals during the diffusion process, a fundamental human trait according to social psychology theories. Intuitively, conformity captures the extent to which an individual complies with social norms or expectations. In this paper, we present a novel framework called CHASSIS to characterize online information diffusion by bridging classical information diffusion model with conformity from social psychology. To this end, we first *extend* "Hawkes Process", a well-known statistical technique utilized to model information diffusion, to quantitatively capture two flavors of conformity, *informational conformity* and *normative conformity*, hidden in activity sequences. Next, we present a novel *semi-parametric inference approach* to learn the proposed model. Experimental study with real-world datasets demonstrates the superiority of CHASSIS to state-of-the-art conformity-unaware information diffusion models.

ACM Reference Format:
Hui Li, Hui Li, and Sourav S Bhowmick. 2020. CHASSIS: Conformity Meets Online Information Diffusion. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD'20), June 14–19, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3318464.3389780

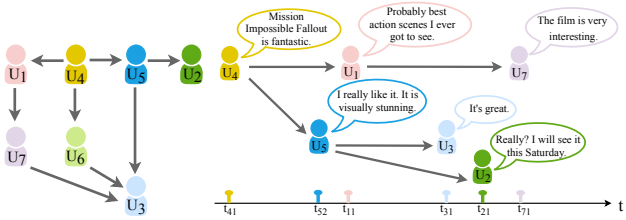
1 INTRODUCTION

Information diffusion is a process by which information and ideas spread over a network, creating a cascade. In particular, information diffusion in social networks continuously generates large-scale activities (*e.g.,* share, post, tweet, retweet, like, comment). These activities can be represented by asynchronous time-stamped sequences wherein each individual gives rise to a sequence of activities over time. In such sequences, there exist abundant *triggering relations* between activities that describe "which activity triggers which activity" [53, 54]. These relations are typically modeled as *diffusion trees* [28, 40]. For example, consider the social network in Figure 1(a) depicting follower-followee relationships. Figure 1(b) depicts a sequence of activities over time, involving some of the users, represented as a diffusion tree. Observe that an activity (*e.g.,* the post of U_4 at time t_{41}) may trigger a succeeding activity (*e.g.,* the comment of U_5 at time t_{52}) represented by a unidirectional link between them. Such unidirectional links between activities lead to diffusion trees. Hence, diffusion trees describe the information cascade (a.k.a informational cascade) generated by the information diffusion process. It is paramount to model this information diffusion process accurately as it underpins a variety of downstream applications such as influence maximization (IM) [26], viral marketing [6, 39], rumour detection [36], user behaviour prediction [12].

Several studies have linked *conformity* [2, 4, 5], a fundamental and well-studied concept in social psychology, to the pivotal role it plays in the generation of information cascade [1, 7]. Intuitively, conformity refers to the inclination to align our attitudes and behaviors with those around us. There are two flavors of conformity, namely *informational conformity* and *normative conformity* [15]. The former occurs when people conform to peer views in an attempt to reach appropriate behaviors and attitudes due to lack of relevant knowledge. The latter occurs because of the desire to be accepted or that keep us from being isolated or rejected by others. For example, reconsider Figure 1. It is indeed possible that although U_3 is unaware of the movie "Mission Impossible Fallout", her response *"It's great"* is same as others because she chose to trust her friend U_5 (*i.e.,* informational conformity). On the other hand, suppose U_3 responds positively because she wants to please her friends even if she dislikes the movie. Then, this is an example of normative conformity.

SIGMOD'20, June 14–19, 2020, Portland, OR, USA

ACM ISBN 978-1-4503-6735-6/20/06…$15.00
https://doi.org/10.1145/3318464.3389780



(a) A social network (b) Diffusion process

Figure 1. Information diffusion.

Since conformity plays a fundamental role in how online users respond to social activities, it naturally influences the information diffusion process. Consequently, it is paramount for information diffusion models to incorporate it.

Example 1.1. The dynamics of spread of information in a network is steered by an information diffusion model. The *Independent Cascade* (IC) is one of the most well-studied information diffusion models. In this model, in the first step a *seed* node is activated and subsequently at any time-step i, each newly activated node U_i gets one independent attempt to activate each of its outgoing neighbors U_j with a probability $p_{i,j}$, which is often set to $\frac{1}{\text{In}(j)}$ where $\text{In}(j)$ refers to the in-degree of U_j [3]. Observe that the IC model is conformity-unaware as the computation of $p_{i,j}$ disregards conformity of users.

Reconsider Figure 1(a). Based on the IC model, $p_{5,3} = \frac{1}{3}$ and $p_{5,2} = 1$. Thus, U_2 is more likely to be activated by U_5 than U_3. However, if we consider their responses in Figure 1(b), U_3 may exhibit higher degree of conformity to U_5 than U_2. Consequently, when conformity of individuals are taken into account, U_3 is more likely to be activated instead of U_2. Hence, a *conformity-aware* information diffusion model may potentially provide more accurate information on the diffusion process. As several existing information diffusion models (*e.g.,* IC) are based only on the network structure, they fail to exploit information related to conformity of individuals. ∎

Despite the crucial role of conformity in online information diffusion, research in this arena is scarce [43, 52]. It is challenging to detect and quantify the two flavors of conformity from social activities. First, the private beliefs of individuals may not be exposed explicitly in the activities. For instance, U_3 may not explicitly mention in her post that she wants to please her friends or have not watched the movie. Hence, it may not be possible to determine whether an individual is conforming to another by simply searching for one's beliefs in the posts. Second, conformity of an individual may vary with the context. One may show high degree of conformity for one topic of discussion (*e.g.,* movies) but not another (*e.g.,* politics). Hence, any conformity computation technique needs to be *context-sensitive*. Third, the knowledge of topology of a social network is insufficient to address this problem as connectivities between individuals do not necessarily indicate manifestation of social activities among them. For instance, some followers may rarely or never interact with some of their followees, and some individuals may respond to some other unconnected individuals in online discussions. For example, in Figure 1, U_5 may respond to a comment by U_1 although they are not connected.

In this paper, we present a novel framework for information diffusion called ***C**onformity-aware **HA**wkes proce**SS**-based **I**nformation Diffu**S**ion* (CHASSIS) to characterize the underlying dynamics of diffusion in the presence of conformity. Specifically, we investigate how the aforementioned two flavors of conformity can be captured in individuals' interactions by exploiting diffusion trees constructed from the observed activity sequences.

Since social activities represent asynchronous time-stamped sequences, we deploy a well-known statistical technique called "Hawkes process" [22, 23], which is a type of point process* [47] that has been utilized recently to model information diffusion [49, 53]. Specifically, we *extend* classical Hawkes processes for information diffusion to capture time-varying conformity of individuals in our model (Section 4). We design a practical *semi-parametric approach* to learn the model components from observed data (Section 7) as well as *infer* the diffusion trees (Section 6) in an alternating fashion (*i.e.,* an instance of the Expectation-maximization method). To this end, as detailed in Section 3, we represent the diffusion trees by utilizing the *branching structure*, an equivalent representation of Hawkes processes, which isolates the events (*e.g.,* activities) in a Hawkes process into *immigrants* (*i.e.,* events that arrive independently) and *offsprings* (*i.e.,* events triggered by existing events). Then we utilize the parent-child pairs of events (*e.g.,* activities) in the branching structure (*i.e.,* diffusion trees), to quantify the two types of conformity in Section 5 and use them in our model. Extensive experiments with real-world datasets show superior performance of CHASSIS in modeling information diffusion compared to several state-of-the-art conformity-oblivious techniques. CHASSIS can be utilized to predict individuals' future behavior with considerable confidence, illustrating the powerful effects of an individual's inclination to align one's attitudes and behaviours with others [32].

In summary, this paper makes the following key contributions: (a) We propose a novel conformity-aware Hawkes process-based framework called CHASSIS to characterize online information diffusion. Our work *bridges the classical online information diffusion problem in data analytics with conformity from the domain of social psychology.* (b) We quantitatively capture two flavors of conformity, informational

*Point processes are stochastic processes that are used to model events that occur at random intervals relative to the time or space axis, and provide the statistical language to describe the timing and properties of events.

conformity and normative conformity, hidden in activity sequences by utilizing diffusion trees (*i.e.,* branching structure) constructed from the activity sequences. In this context, we propose a novel *diffusion tree inference* technique when explicit information about links between activities are unavailable to a downstream application. (c) We present a novel and efficient *semi-parametric inference approach* that leverages on the diffusion trees to learn the conformity-aware information diffusion model competently from observed data. (d) We conduct an experimental study with real social datasets to demonstrate the superiority and effectiveness of CHASSIS.

2 RELATED WORK

Conformity in online social networks. A rich line of work in social psychology [2, 4, 5, 8] has demonstrated the existence and importance of conformity in social interactions. However, there is scant research on investigating conformity in online social networks. The seminal work of Li *et al.* [30] studied the interplay between influence and conformity of each individual in online social networks by utilizing the positive and negative relationships between individuals. Subsequently, they modeled conformity in the context of IM problem [31]. Recently, [34] adopted group profiling in conformity-aware IM problem. Tang *et al.* [43] proposed a probabilistic factor graph model that predicts user behavior by exploiting the effect of conformity. The work in [52] assigns hidden roles to users and then learns the correlation between roles and conformity. None of these work model the interplay of informational and normative conformity, which is a more realistic way to capture the role conformity plays in social networks. Importantly, we focus on *inferring* the conformity-aware information diffusion model from the data, which is orthogonal to these efforts.

Diffusion models. According to a previous study [21], information diffusion models can be categorized into *predictive* and *explanatory* models. Predictive models aim to uncover and predict how a specific diffusion process would unfold in a given network. These works consider the diffusion as a *discrete* random process happened among network nodes and can be further classified into *non-progressive* and *progressive* models. In the former model, a node affected by a piece of information cannot switch to unaffected status subsequently. This includes the independent cascade (IC) [39, 44] and linear threshold (LT) [17] models, which are widely adopted in the IM problem. IC/LT model has also been augmented with topic [11, 33], economic theory [6], and spatial-temporal features [29] in estimating the diffusion spread. In comparison, the *progressive model* (*e.g.,* SIR and SIS [24, 38] for virus propagation) allows an affected node to be unaffected again. All these predictive models are used for estimating the diffusion scope. They simplify the diffusion process to happen at discrete steps instead of continuous time.

Explanatory models are used to infer the diffusion path in order to retrace and understand how a piece of information is propagated, and can benefit a series of applications including fake news detection [48], user behavior prediction [12], etc. For instance, [19] models the diffusion process as a spatially discrete network of *continuous*, conditionally independent temporal processes occurring at different rates. They presented NETRATE algorithm to infer pairwise transmission rates and the graph of diffusion. Recently, Hawkes process has been employed in modeling the information diffusion process. ADM4 [53] uses the mutually-exciting linear Hawkes model to capture the temporal patterns of user behaviors, and infer the social influence. MMEL [54] captures the temporal dynamics of the observed activities by utilizing multi-dimensional linear Hawkes processes, and learns the triggering kernels nonparametrically. Although these models are able to uncover the diffusion as a continuous temporal process, they fail to take into account conformity of individuals.

Lastly, there are also several efforts in the literature to predict the information cascade [12, 20, 50]. However, all these efforts are conformity-unaware.

3 BACKGROUND

In this section, we provide the necessary background knowledge to understand the paper.

3.1 Hawkes Processes

Many applications may need to deal with timestamped events in continuous time. *Point process* [47] is a principled framework to model such event data. Specifically, a point process on a time line is a random process for realization of the *event times* $t_1, t_2, \ldots$ where t_i is the time of occurrence of the ith event (*e.g.,* a tweet, like). Point process can be equivalently represented as a counting process $N = \{N(t)|t \in [0, T]\}$ over the time interval $[0, T]$ where $N(t)$ records the number of events up to time t. Let $\mathcal{H}_t$ be the history of events before time t. Then dynamics of the point process could be characterized by a *conditional intensity function* $\lambda(t)$ as follows:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[N(t + \Delta t) - N(t)|\mathcal{H}_t]}{\Delta t} \tag{3.1}$$

where two events coincide with probability 0, *i.e.,* $N(t+\Delta t) - N(t) \in \{0, 1\}$. Intuitively, larger the intensity $\lambda(t)$, greater the likelihood of observing an event in the time window $[t, t + \Delta t]$.

In some applications, the arrival of an event increases the likelihood of observing events in the near future. To model these applications, there exists a type of point processes in which the event arrival rate explicitly depends on past events. These processes are referred to as *self-exciting processes*. *Hawkes processes* [22] is the most well-known self

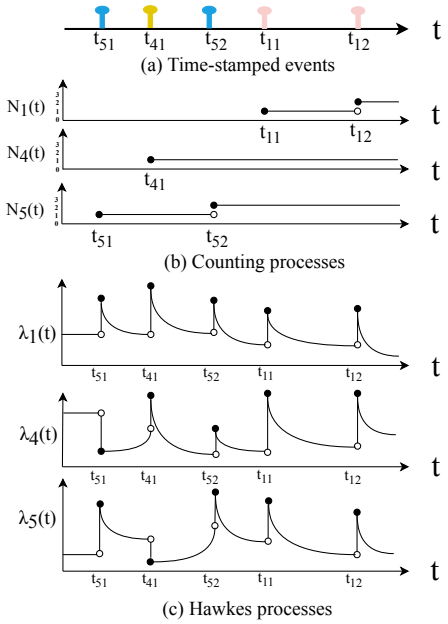


Figure 2. 3-dimensional Hawkes processes: (a) Five social activities during a time interval. (b) Counting process over time for each individual. $N(t)$ increases by one when an activity happens. (c) Intensity functions.

exciting process and have been extensively used in many domains (*e.g.,* finance, seismology, social media).

In this paper, we focus on *multi-dimensional Hawkes process* [53], which is defined by an M-dimensional point process where M Hawkes processes are *integrated* with each other. That is, it is an M-dimensional counting process where an arrival in one dimension can affect the arrival rates of all dimensions. In information diffusion, each dimension i represents an individual U_i in a social network and an event represents a social activity. Hence, each Hawkes process corresponds to an individual U_i and the influence between them is modeled by utilizing the *mutually-exciting* property of the M-dimensional Hawkes process. Formally, the intensity function of the ith dimension takes the following form [53]:

$$\lambda_i(t) = \mathcal{F}_i\big(\mu_i + \sum_{j\in[M]} \sum_{t_{jl}<t} \alpha_{ij}\phi_{ij}(t - t_{jl})\big) \qquad (3.2)$$

Wherein the constant $\mu_i > 0$ is the *base intensity* of the ith Hawkes process, describing the arrival of events (*e.g.,* social activities) triggered by external sources. It is also referred to as *exogenous intensity*, and their arrivals are independent of the previous events. The strength of influence between dimensions (*i.e.,* individuals) is parameterized by a sparse *excitation matrix* $\boldsymbol{A} = [\alpha_{ij}]_{i,j\in[M]}$. In particular, the coefficient $\alpha_{ij} \geq 0$ captures the *mutually-exciting* property between the ith and jth processes. Larger value of α_{ij} indicates that events (activities) in the ith dimension are more likely to trigger an event in the jth dimension in the future. The *triggering kernel* $\phi_{ij}(t - t_{jl})$ quantifies the change in the rate

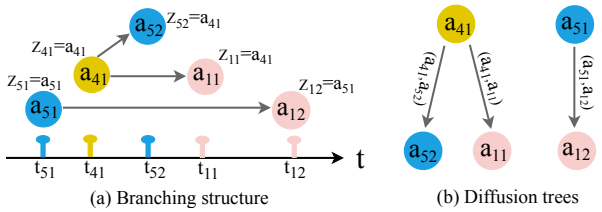


Figure 3. (a) The branching structure of the 3-dimensional Hawkes processes in Figure 2. (b) The corresponding diffusion trees.

of occurrence caused by the historical realization t_{jl}. The second item $\sum_{j\in[M]} \sum_{t_{jl}<t} \alpha_{ij}\phi_{ij}(t - t_{jl})$ is referred to as *endogenous intensity* and captures the mutually-exciting nature of the point processes. In our context, it captures the interactions between individuals in a social network – each event occurred to an individual U_j may increase (*i.e.,* mutual excitation) or decrease (*i.e.,* mutual inhibition) the arrival rate of occurrence in U_i by a certain amount which itself decays over time. Figure 2 depicts a 3-dimensional Hawkes processes involving U_1, U_4, and U_5 in Figure 1(a). Specifically, Figure 2(a) shows activity sequences at times t_{41}, t_{52}, and t_{11} in Figure 1(b) along with two additional activities by users U_5 and U_1 at times t_{51} and t_{12}, respectively. Figure 2(b) shows the corresponding counting process of each dimension. Figure 2(c) illustrates $\lambda_i(t)$ of the three individuals, provoking different changes due to these activities (*i.e.,* events). Observe that each occurred activity causes a jump (up or down) in the intensity function. Each jump is followed by a rapid decay guided by the kernel function.

Essentially, various combinations of kernel functions could recognize various temporal characteristics. When $\mathcal{F}_i(x) = x$, such processes are referred to as *linear* Hawkes processes [22] where the intensity is a linear accumulation of a series of kernel functions. Unfortunately, such linearity may not capture some real-world applications including information diffusion [37]. Consequently, *nonlinear* Hawkes processes [9] are proposed to address this limitation. In this paper, we *integrate conformity with linear or nonlinear M-dimensional Hawkes processes for modeling information diffusion.*

3.2 Branching Structure

An equivalent view of the Hawkes process refers to the Poisson cluster process interpretation [23], which isolates the events in a Hawkes process into two categories: *immigrants* and *offsprings*. The offspring events are triggered by existing events in the process whereas the immigrants arrive independently and hence do not have an existing parent event. That is, we call an event an *immigrant* if it is generated due to the exogenous intensity $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M)^T$ spontaneously, otherwise, it is an *offspring*. The offsprings are structured into *clusters*, associated with each immigrant event. This is referred to as the *branching structure* [22, 23].

It provides a way to capture the parent-child triggering relations between events as follows: (a) an immigrant event starts generating offsprings; (b) each offspring starts generating other offsprings immediately after birth.

To facilitate exposition in the context of social activity sequences, we denote a collection of activities (*i.e.,* events) as $\{a_{ik} = (t_{ik}, C_{ik})\}_{k=1}^{N_i(T)}$ in the time window $[0, T]$, where t_{ik} denotes the occurrence time of the k^{th} activity of an individual U_i, and C_{ik} records the activity's content. W.l.o.g, we assume that an arbitrary activity can be triggered by at most one activity. We introduce a set of auxiliary variables, denoted as $\{\{Z_{ik}\}_{k=1}^{N_i(t)}\}_{i=1}^{M}$, to represent the branching structure as following:

- $Z_{ik} = a_{ik}$ if activity a_{ik} is an immigrant; and
- $Z_{ik} = a_{jl}$ if the parent of activity a_{ik} is activity a_{jl}.

If activity a_{jl} triggers a_{ik}, (a_{jl}, a_{ik}) is referred to as a *parent-child* pair of activities (*i.e.,* event a_{ik} is an offspring of event a_{jl}). Given one offspring activity a_{ik}, we denote its parent activity as Z_{ik} accordingly, and the corresponding parent-child pair of activities as (Z_{ik}, a_{ik}).

Figure 3(a) depicts the branching structure representation of the Hawkes processes in Figure 2. Each circle represents an event (*i.e.,* activity) and a directed link represents the parent-child relationship between two events. For instance, $Z_{51} = a_{51}$ indicates that the activity a_{51} is an immigrant, and $Z_{12} = a_{51}$ denotes that the activity a_{51} generates a_{12}. Hence, (a_{51}, a_{12}) expresses a parent-child pair of events. Observe that *each connected component* in the branching structure represents a tree structure.

3.3 Diffusion Tree

A popular approach to represent a sequence of user activities over a time period is by using a collection of *diffusion trees* [28, 40], denoted by $\mathcal{D}$. A *diffusion tree*, $D_t = (V, E)$, consists of a set of user activities as its nodes V, and a set of unidirectional edges $E = \{(a_{ik}, a_{jl})\}$ denoting that the activity a_{ik} triggers the activity a_{jl} w.r.t the temporal precedence $t_{ik} < t_{jl}$. For example, in *Twitter*, the root node of D_t is an original tweet. If the original tweet triggers a series of response, it generates a series of child activities (*e.g.,* retweet, comment, like), referred to as *first generation* descendants. Following this, first generation descendants subsequently generate their own child activities (*i.e., second generation* descendants), and so on. Figure 1(b) depicts a diffusion tree.

Observe that an original activity and its descendants in a diffusion tree represent an immigrant and its offsprings, respectively, in the branching structure. Hence, there is a direct correspondence between a set of diffusion trees and the branching structure of Hawkes processes. Each diffusion tree D_t is a connected component in the branching structure where a node and an edge in D_t are an event and a parent-child pair of events (*e.g.,* activities) in the latter, respectively. For example, the two diffusion trees in Figure 3(b) represent the branching structure in Figure 3(a). In the sequel, we shall use these two concepts interchangeably.

4 CONFORMITY-AWARE INFORMATION DIFFUSION MODEL

Equation 3.2 is used to model information diffusion by recent works [53, 54]. Specifically, its components are estimated from the observed social activities. Then, we can simulate the diffusion process beyond time T and predict various properties of the cascade. Observe that the strength of influence from an individual U_j to an individual U_i in Equation 3.2 (*i.e.,* α_{ij}) solely determines their degree of interaction in these classical Hawkes-based models. Intuitively, the stronger the influence α_{ij}, the more likely U_i responds to U_j during information diffusion. However, as remarked earlier, interactions between individuals are also likely to be impacted by conformity of users. Hence, we need to augment classical information diffusion models to capture this phenomenon by incorporating informational conformity and normative conformity [15].

Although in some scenarios conformity may be purely informational or purely normative, in most cases these two occur concurrently [25]. Furthermore, contributions of these two types of conformity are likely to vary between different instances of conformity and between individuals [13]. Consequently, we decompose the time-varying *influence strength* $\alpha_{ij}(t)$ into two additive parts, *informational influence* $\alpha_{ij}^I(t)$ and *normative influence* $\alpha_{ij}^N(t)$, to quantify the presence of informational conformity and normative conformity, respectively. That is,

$$\alpha_{ij}(t) = \gamma_{ij}^I(t)\alpha_{ij}^I(t) + \gamma_{ij}^N(t)\alpha_{ij}^N(t) \tag{4.1}$$

In the above equation, the time-dependent *informational coefficient* $\gamma_{ij}^I(t)$ and *normative coefficient* $\gamma_{ij}^N(t)$ are parameterized to weigh informational conformity against normative conformity at time t. Observe that if $\alpha_{ij}(t) > 0$, then we know that conformity plays a role when U_j is influencing U_i. Substituting it into Eq. 3.2 gives us the model for conformity-aware Hawkes process-based information diffusion:

$$\lambda_i(t) = \mathcal{F}_i\left(\mu_i + \sum_{j\in[M]}\sum_{t_{jl}<t}(\gamma_{ij}^I(t)\alpha_{ij}^I(t) + \gamma_{ij}^N(t)\alpha_{ij}^N(t))\phi_{ij}(t - t_{jl})\right) \tag{4.2}$$

We elaborate on how to quantify $\alpha_{ij}^I(t)$ and $\alpha_{ij}^N(t)$ in the next section. In Section 7, we describe the inference of remaining components.

5 COMPUTATION OF CONFORMITY

In this section, we delineate how to quantify the two types of conformity using diffusion trees (*i.e.,* branching structure). The formal algorithms are given in [32].

5.1 Informational Conformity

We often look to people around us who are better informed and more knowledgeable, and then use their opinions as a guide to our own behaviour and response. Such phenomenon (*i.e.,* the desire to be correct) not only occurs between friends but also individuals who have never known one another. This is known as informational conformity [10, 15]. Intuitively, informational conformity in social networks is not *symmetrical*. That is, informational conformity from an individual U_i to an individual U_j (*i.e.,* U_i conforms to U_j) may differ from that of U_j to U_i. According to social psychology theories [14, 41], the higher the influence of U_i, the higher the informational conformity of U_j to U_i. Following this, if U_j interacts with U_i frequently, then we should boost their *informational influence*. At the same time, during such interactions, if U_i almost always agrees with U_j, we can say that U_i is likely to conform to U_j. Consequently, we utilize the notions of *influence degree* (*i.e.,* measure of interaction frequency) and *context stance* (*i.e.,* opinion polarity *w.r.t* a topic) to quantify the pairwise informational conformity. Intuitively, the product of these two items describes how likely U_i's attitudes and behaviors are infected by another individual U_j in the presence of informational conformity. That is, informational influence from U_j to U_i (denoted as $\alpha^I_{ij}(t)$) is computed as follows: $\alpha^I_{ij}(t) = \Phi_{ij}(t) \times \Psi_{ij}(t)$ wherein the first item $\Phi_{ij}(t)$ is referred to as *influence degree*, and the second item $\Psi_{ij}(t)$ aims to compute the *context stance*. Evidently, both of them are derived from the historical interactions between individuals. Put simply, the higher the $\alpha^I_{ij}(t)$, the higher is the informational conformity and vice versa.

Influence Degree. Frequent interactions between individuals demonstrate their closeness, and lead to high pairwise influence degree [51]. Furthermore, such influence degree of one individual to another evolves over time. However, the effect of previous interactions may decrease with time, namely the time decaying effects [35]. For simplicity, we assume each response (*i.e.,* one offspring activity in the branching structure) provokes one interaction, followed by an exponential decay [42]. Hence, we measure the influence degree from individual U_j to individual U_i as:

$$\Phi_{ij}(t) = \frac{\sum_{k=1}^{N_i(t)} \mathbb{I}_{N_{ij}(t)}(Z_{ik}, t_{ik}) \exp\{-\beta_{ij}(t - t_{ik})\}}{\mathbb{N}_i(t)} \tag{5.1}$$

Wherein $N_{ij}(t)$ records the collection of parent-child activity pairs [†], $\{(t_{jl}, t_{ik})\}$, up to time t. $\mathbb{I}_{N_{ij}(t)}(Z_{ik}, t_{ik})$ is an indicator function, which equals to one when $(Z_{ik}, t_{ik}) \in N_{ij}(t)$ and zero otherwise. We use $\boldsymbol{\beta} = \{\beta_{ij}\}$ to capture the decay rate of previous interactions between individuals. Different from $N_i(t)$, $\mathbb{N}_i(t)$ denotes the total number of offspring activities occurring to individual U_i until time t (*i.e.,* $\mathbb{N}_i(t) \leq N_i(t)$), which could be calculated by leveraging the diffusion trees of the activity collection $\mathbf{X}_t$. Obviously, the domain of influence degree from individual U_j to individual U_i is $[0, 1]$. Observe that $\Phi_{ij}(t)$ does not assume any connection between U_i and U_j (*i.e.,* U_i and U_j may or may not be friends/followers).

Context Stance. We glean insights on respondents' opinion polarity with respect to a topic in social interactions and apply stance detection [16] to obtain the dissemination of individuals' beliefs. Generally, such opinion polarity is often expressed in the form of discrete class labels, *e.g.,* positive or favor, negative or against, and neutral or none [16], either *explicitly* or *implicitly*. Explicit stances are direct expressions of opinion toward target concepts, such as "like" or "angry" given to a particular post and the corresponding polarity is 1 or 0, respectively. Implicit stances can be extracted from social media posts using *NLTK* (www.nltk.org), which is a popular sentiment analysis package.

Given each parent-child pair of activities (t_{jl}, t_{ik}) considered in $N_{ij}(t)$, we calculate the polarity p_{jl}, p_{ik} of activity t_{jl} and t_{ik}, and then append them into two vectors: $\overrightarrow{\boldsymbol{p}}^I_{\boldsymbol{j}}(\boldsymbol{t}) = (p_{jl})_{a_{jl} \in \mathbf{X}_t}$ and $\overrightarrow{\boldsymbol{p}}^I_{\boldsymbol{i}}(\boldsymbol{t}) = (p_{ik})_{a_{ik} \in \mathbf{X}_t}$, respectively. Next, we evaluate the Pearson correlation coefficient (Pcc) of the vectors, denoted as $\Psi_{ij}(t) = \text{Pcc}(\overrightarrow{\boldsymbol{p}}^I_{\boldsymbol{j}}(\boldsymbol{t}), \overrightarrow{\boldsymbol{p}}^I_{\boldsymbol{i}}(\boldsymbol{t})) \in [-1, 1]$, to quantify the *context stance* over time. Intuitively, the higher the value of context stance, the higher is the informational conformity from individual U_i to individual U_j.

5.2 Normative Conformity

Without loss of generality, a new post (*e.g.,* tweet) may lead to a chain of interactions (*e.g.,* retweet, comment, reply, like) in a social network. In practice, such an immigrant activity (*e.g.,* a tweet) generates its offspring activities (*e.g.,* a series of retweets, comments, replies, likes) possibly involving multiple individuals (*i.e.,* dimensions). In the sequel, we refer to an immigrant activity together with all its offspring activities as *one informational cascade*. Hence, based on Section 3.3, a diffusion tree D_t represents one informational cascade. For example, all activities in Figure 1(b) construct one informational cascade. Figure 4 depicts another example.

An informational cascade occurs when it is optimal for an individual, having observed the actions of individuals ahead of him, to follow their behavior without regard to his own information [7]. In other words, a cascade can occur when people observe and follow "the crowd", even when the group consensus conflicts with their own private information [1]. Such phenomenon of following the crowd is known as normative conformity [10, 15].

Once one informational cascade starts around a particular topic, a few early individuals commit their actions (*i.e.,* adopt or reject) through a sequence of activities (*e.g.,* retweet, comment, reply, like), and then subsequent individuals may refer

[†] In the sequel, for simplicity, we sometimes use the occurrence time to denote one activity, *e.g.,* t_{jl} represents a_{jl}.

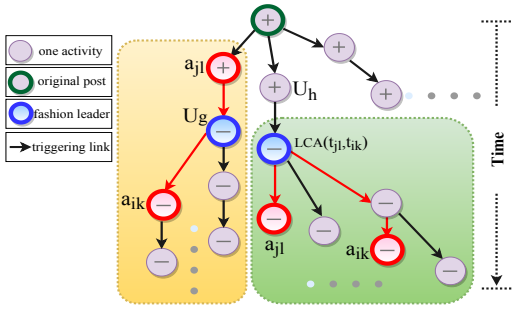


Figure 4. Diffusion tree of one informational cascade: "+" and "-" denote the opinion polarity (*i.e.,* adoption and rejection, respectively) of one activity *w.r.t* a topic.

to them. Hence, triggering links between the preceding and following activities give us the opportunity to extract normative conformity by analysing the context stance hidden in the diffusion trees.

Specifically, the normative conformity of U_i to U_j depends on the *aggregated* adoptions of U_i to behaviors and attitudes of U_j. In reality, in one cascade, even though U_j's activity does not immediately precede U_i's, U_j's decision may convey information to U_i, and then U_i may act according to the information conveyed by the actions of preceding individuals (including U_j) [7]. Hence, in order to quantify the *normative influence* $\alpha_{ij}^N(t)$, we need to ensure the followings in the diffusion trees: (a) both individuals U_i and U_j are involved within one cascade; and (b) the corresponding activity a_{jl} happens before a_{ik}.

Furthermore, informational cascades can be fragile, with abrupt shifts or reversals in direction [7, 18]. Specifically, either small shocks (*i.e.,* when new information becomes available) can easily shift the behavior of many individuals, or higher-precision individuals can shift a cascade because they are more inclined to use their own information than those that precede them. For example, consider the activities of individuals highlighted in blue (also known as *fashion leader* [7]) in Figure 4. Observe that there is a shift in opinion polarity at this point, which is adopted by subsequent actions from individuals. Hence, the normative conformity of U_i to U_j may vary due to such sudden shifts.

We formulate the *normative influence* $\alpha_{ij}^N(t)$ as follows to capture how likely individual U_i's attitudes and behaviors are infected by another individual U_j:

$$\alpha_{ij}^N(t) = \text{Pcc}(\vec{\boldsymbol{p}}_j^N(t), \vec{\boldsymbol{p}}_i^N(t)) \tag{5.2}$$

In order to compute the context stance for this type of conformity (right side of the above equation), we consider the following two scenarios.

Scenario 1: Given an informational cascade, the two activities a_{jl} and a_{ik} *lie on the same path* of the diffusion tree in chronological order. No matter a_{jl} immediately precedes a_{ik} or not, U_j's action a_{jl} impacts U_i's response a_{ik} to some extent. In this case, we could directly capture their normative influence from the two activities in two steps: (a) append the polarity scores p_{jl}, p_{ik} into vectors $\vec{\boldsymbol{p}}_j^N(t), \vec{\boldsymbol{p}}_i^N(t)$, respectively; (b) recalibrate $\text{Pcc}(\vec{\boldsymbol{p}}_j^N(t), \vec{\boldsymbol{p}}_i^N(t))$. For example, consider the yellow panel of Figure 4. Observe that U_j gives a positive response a_{jl} to the original post, and then U_g replied U_j with an opposing view (*i.e.,* "-" denotes U_g's negative opinion polarity) due to some reason. Afterwards, U_i agrees with U_g. Obviously, U_g has a greater normative influence on U_i than U_j.

Scenario 2: Consider the two activities a_{jl} and a_{ik} located in the green panel of Figure 4. Even though they are triggered by different parent activities and are located in different paths of the diffusion tree, they are both impacted by the highlighted activity in blue (*i.e.,* the lowest common ancestor of a_{jl} and a_{ik}, denoting as $\text{LCA}(a_{jl}, a_{ik})$). Furthermore, if $\text{LCA}(a_{jl}, a_{ik})$ happens to be a fashion leader (*i.e.,* U_h gives positive response to the original post, afterwards, $\text{LCA}(a_{jl}, a_{ik})$ suddenly shifts his opinion), it definitely would have some effect on subsequent activities. Consequently, in such scenario, we quantify the normative conformity of U_i to U_j as follows. We first append the polarity pair p_{jl} (resp. p_{ik}) and $p_{\text{LCA}(t_{jl}, t_{ik})}$ into vectors $\vec{\boldsymbol{q}}_j^N(t)$ (resp. $\vec{\boldsymbol{q}}_i^N(t)$) and $\vec{\boldsymbol{q}}_{\text{LCA}_{ij}}^N(t)$, and then recalculate their Pearson correlations coefficient $\text{Pcc}(\vec{\boldsymbol{q}}_j^N(t), \vec{\boldsymbol{q}}_{\text{LCA}_{ij}}^N(t))$ (resp. $\text{Pcc}(\vec{\boldsymbol{q}}_i^N(t), \vec{\boldsymbol{q}}_{\text{LCA}_{ij}}^N(t))$) before appending to $\vec{\boldsymbol{p}}_j^N(t)$ (resp. $\vec{\boldsymbol{p}}_i^N(t)$).

Scanning all information cascades satisfying the aforementioned conditions up to time t, we calculate the Pearson correlation coefficient of the vectors $(\text{Pcc}(\vec{\boldsymbol{p}}_j^N(t), \vec{\boldsymbol{p}}_i^N(t)))$ to quantify $\alpha_{ij}^N(t)$ between individuals in the presence of normative conformity.

Note the difference in the computation of context stance for normative conformity compared to informational conformity. For the latter, an individual may refer to surrounding people who are better informed and more knowledgeable, and then use their opinions as a guide for his/her own behaviours. Hence, computation of the context stance for informational conformity focuses on the parent-child activity pairs (*i.e.,* U_j precedes U_i immediately). For the former, an individual follows the behaviour of the preceding individuals during an informational cascade. Consequently, context stance of U_i and U_j is computed by considering the aggregated activities of U_i to the activities of U_j even though U_j's activity may not immediately precede U_i's activity (*i.e.,* they are not parent-child activity pairs).

6 CONSTRUCTION OF DIFFUSION TREES

In the preceding section, the informational and normative influence (*i.e.,* $\alpha_{ij}^I(t)$, $\alpha_{ij}^N(t)$) are computed by utilizing the

diffusion trees. In this section, we elaborate on how the diffusion trees are construction.

Connectivity-aware construction. If an online social network explicitly exposes connectivity information (*i.e.,* parent-child link) of activity sequences to an application then it is straightforward to construct the diffusion trees. Activities with no parents form the immigrants and those with parents form the offsprings.

Diffusion tree inference. The construction of diffusion trees becomes challenging when parent-child link information is unavailable from the social activities exposed to an application (*e.g.,* links in Figure 3 are missing). For example, the *Twitter* API returns the following fields: *tweet_id*, *created_time*, *text*, and *user_id*. That is, it does not provide connectivity information (*e.g., reply_id*) of the activities. Hence, we need to *infer* the latent diffusion trees (*i.e.,* the branching structure).

Branching structure or diffusion tree inference for information diffusion has been addressed in several prior work [53, 54]. Unfortunately, these existing methods are only suitable for *linear* Hawkes processes. Hence, it is desirable to devise an inference strategy that can handle both linear and nonlinear Hawkes processes by relaxing the requirement of linearity.

We propose an expectation-maximization (EM) iterative learning scheme to infer the diffusion trees. To initialize the EM procedure, we firstly sample the auxiliary variables $\{\{Z_{ik}\}_{k=1}^{N_i(t)}\}_{i=1}^{M}$. Afterwards, we update the probability of branching structure (*i.e.,* infer the diffusion trees) in the E-step given the CHASSIS model learned from the previous iteration. Thus, the inference procedure of CHASSIS can be embedded into the M-step naturally. In this section, we briefly describe the inference of diffusion trees. We defer the details of the inference procedure of CHASSIS in Section 7.

We present an overview of our inference strategy here. Detailed discussion of the steps are given in [32]. The greater the influence of the preceding activity to the following activity, the more likely there is a triggering link between them (*i.e.,* they are a parent-child pair of activities). From this perspective, we first deduce the *Papangelous conditional intensity* [46] of Hawkes processes to weigh the extent to which removing one activity will affect the subsequent activities in chronological order. Then, we utilize it to reflect the probability that a preceding activity a_{ik} triggers on a succeeding activity a_{jl}. After that, we obtain the parent-child pairs of activities probabilistically.

7 INFERENCING CHASSIS MODEL

Once the diffusion trees are updated, we optimize the CHASSIS model to best explain the information diffusion process. Consequently, in this section we propose a novel semi-parametric inference algorithm regardless of whether Hawkes processes are linear or nonlinear, wherein exogenous intensity $\{\mu_i\}_{i\in[M]}$, decay rate of previous interactions $\{\beta_{ij}\}_{i,j\in[M]}$, informational and normative coefficients $\{(\gamma_{ij}^I(t), \gamma_{ij}^N(t))\}_{i,j\in[M]}$ are learned from the observed activity sequences, while the triggering kernel functions $\{\phi_{ij}(t)\}_{i,j\in[M]}$ are estimated nonparametrically via Fourier transform without prior domain knowledge. The pseudocode of the algorithm is given in [32].

Parametric Inference. We denote the set of parameters $\{\mu_i, \gamma_{ij}^I(t), \beta_{ij}, \gamma_{ij}^N(t)\}_{i,j\in[M]}$ as Θ. We can estimate them by maximizing the likelihood over the observed data. Given the social activity collection $\mathbf{X}_t$ over the time interval $(0, t]$, the log-likelihood of conformity-aware Hawkes processes associated with the conditional intensity in Eq. 4.2 is in fact the summation of that over all dimensions, each of which can be interpreted as follows: the sum of the log-intensities of activities that happened, minus an integral of the total intensities over the observation interval $(0, t]$ [53],

$$\ln \mathcal{L}_i(\Theta|\mathbf{X}_t) = \sum\nolimits_{k=1}^{N_i(t)} \ln \lambda_i(t_{ik}) - \int_0^t \lambda_i(s)ds \tag{7.1}$$

However, $\int_0^t \lambda_i(s)ds$ is not always directly computable *w.r.t* various intensity functions $\mathcal{F}_i(\cdot)$. We propose a modified flexible-size Euler integration method to calculate $\int_0^t \lambda_i(s)ds$ in an iterative manner.

THEOREM 7.1. *Let $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$. Given an accuracy bound ξ, within the time interval $(0, t]$, taking I_m steps with the Euler method using step size $h_m = \frac{t}{I_m}$, the m^{th} iteration yields the following approximation*

$$\Lambda_i^m(t) = h_m\big(\mu_i + \lambda_i(t_1) + \ldots + \lambda_i(t_{I_m})\big) \tag{7.2}$$

and the estimation error is upper bounded by $|\Lambda_i(t) - \Lambda_i^{(m)}(t)| \leq \frac{e^{L_i t}}{L_i} O(\triangle t)$, where $\lambda_i(t)$ is Lipschitz continuous with $|(\Lambda_i^{(m)})'(t) - \Lambda_i'(t)| \leq L_i|\Lambda_i^{(m)}(t) - \Lambda_i(t)|$ and $O(\triangle t)$ denotes the first-order truncation error.

The proof of Theorem 7.1 is given in [32]. Using this numerical integration, it is straightforward to see that the log-likelihood in Eq. 7.1 can be approximately calculated regardless of the forms of $\mathcal{F}_i(\cdot)$. Next, we learn the parameters Θ by maximum likelihood estimation (MLE) using the gradient ascent method. Notably, we do not need to predefine the shape of the kernel functions. Thus the log-likelihood in Eq. 7.1 is concave, such that the global maximum and the convergence of inference can be guaranteed [45]. Additionally, Θ can be estimated in parallel over all dimensions.

Nonparametric Inference. Once the parameters Θ are estimated, we are left to estimate the kernel functions. The time shift in the kernel function $\phi_{ij}(t - t_{jl})$ in time domain corresponds to a multiplication by an exponential function in frequency domain as follows:

$$\phi_{ij}(t - t_{jl}) \Longrightarrow e^{-\mathrm{J}\omega t_{jl}} \mathbf{\Phi}_{ij}(\omega) \tag{7.3}$$

It provides a way to simplify the time-shifted kernel functions, through which the intensity function $\lambda_i(t)$ can be transformed to the frequency domain, referred to as $\mathbf{\Lambda_i}(\omega)$.

If $\lambda_i(t)$ is a linear combination of a series of time-shifted kernel functions, it is straightforward to obtain $\mathbf{\Lambda_i}(\omega)$ by applying Fourier transform directly. However, the nonlinear functions $\mathcal{F}_i(\cdot)$ prevent us from applying such strategy. To circumvent this issue, we apply Taylor approximation to relax the linearity limitation, and derive the frequency domain counterpart of $\lambda_i(t)$ as following:

$$\begin{aligned}\mathbf{\Lambda_i}(\omega) &= \int_{-\infty}^{\infty} \lambda_i(t) e^{-\mathrm{J}\omega t} dt \qquad (7.4)\\ &\approx \int_{-\infty}^{\infty} \left(\mathcal{F}_i(\mu_i) + \mathcal{F}_i'(\mu_i) \sum_{j\in[M]} \sum_{t_{jl}<t} \alpha_{ij}\phi_{ij}(t-t_{jl}) \right) e^{-\mathrm{J}\omega t} dt \\ &= 2\pi\mathcal{F}_i(\mu_i)\delta(\omega) + \mathcal{F}_i'(\mu_i) \sum_{j\in[M]} \alpha_{ij}(t) \sum_{t_{jl}<t} e^{-\mathrm{J}\omega t_{jl}} \mathbf{\Phi}_{ij}(\omega)\end{aligned}$$

where $\delta(\omega)$ is the Dirac delta function[‡].

According to Eq. 3.1, we could obtain that the expectation of an increment of the counting process $N_i(t+dt) - N_i(t)$ is essentially equivalent to $\lambda_i(t)dt$. Consequently, if we separate the period $(0, t]$ into N equal-length time slots (*i.e.,* $NT = t$) and denote the corresponding number of activities within each slot as $N_i[0], N_i[1], \ldots, N_i[N-1]$, respectively, $\mathbf{\Lambda_i}(\omega)$ can then be interpreted in terms of $N_i[k]$ as:

$$\mathbf{\Lambda_i}(\omega) = \sum_{k=0}^{N-1} N_i[k] e^{-\mathrm{J}\omega kT}$$

Since ω is a continuous variable, there are an infinite number of possible values of ω from 0 to 2π. Hence, $\mathbf{\Lambda_i}(\omega)$ could only be calculated at a finite set of frequencies. Therefore, we divide the unit circle into N equally area (*i.e.,* $\frac{1}{NT}$Hz, $\frac{2\pi}{NT}$rad/sec), and denote them as $\omega_n = \frac{2\pi}{NT} \times n$, then:

$$\mathbf{\Lambda_i}[n] = \sum_{k=0}^{N-1} N_i[k] e^{-\mathrm{J}\omega_n k} \qquad (n = 0 : N-1) \qquad (7.5)$$

wherein $\mathbf{\Lambda_i}[n]$ contains information about the amplitude and phase of the sinusoid wave of frequency ω_n. Intuitively, the triggering kernel function $\phi_{ij}(t)$ should be proportional to the decay rate of previous interactions β_{ij}. As a result, given ω_n, we could obtain :

$$\mathbf{\Phi}_{ij}[\omega_n] = \frac{\beta_{ij}\mathbf{\Lambda_i}[n]}{\mathcal{F}_i'(\mu_i) \sum_{j\in[M]} \alpha_{ij}(t)\beta_{ij} \sum_{l=1}^{N_j(t)} e^{-\mathrm{J}\omega_n t_{jl}}} \qquad (7.6)$$

In particular,

$$\mathbf{\Phi}_{ij}[\omega_0] = \frac{\mathbf{\Lambda}_i[0] - 2\pi\mathcal{F}_i(\mu_i)}{\mathcal{F}_i'(\mu_i) \sum_{j\in[M]} \alpha_{ij}(t) N_j(t)} \qquad (7.7)$$

Then we could deduce the time domain counterpart of $\mathbf{\Phi}_{ij}[\omega]$ by inversing DFT (IDFT):

$$\phi_{ij}(t) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{\Phi}_{ij}[\omega_n] e^{\mathrm{J}\omega_n t} \qquad (7.8)$$

[‡] $\delta(\omega)$ is zero everywhere except at $\omega = 0$, and its total integral is 1.

The above estimation for kernel functions is completely data-driven. We run the parametric and nonparametric inference procedures alternatively until convergence.

Applications of CHASSIS. By leveraging the estimated CHASSIS, we can predict user behaviours, such as next activity prediction and future number of activities prediction. Detailed discussion on such applications is given in [32].

8 PERFORMANCE STUDY

In the section, we present the performance of CHASSIS. We report key results here. Additional results are given in [32]. We have implemented the framework in Python. All experiments are performed on a 64-bit Windows desktop with Intel(R) Core(TM) E5-1620V2 CPU@3.70 and 16GB RAM.

Strategies. We compare the following information diffusion models: (a) ADM4 [53]: It utilizes the mutually-exciting linear Hawkes model to capture temporal patterns of user behaviors, and infers the social influence by imposing both low-rank and sparse regularization on the influence matrix. (b) MMEL [54]: It captures the temporal dynamics of observed activities by linear Hawkes processes, and learns the triggering kernels nonparametrically. (c) CHASSIS-L: Our proposed CHASSIS model. Here, we set $\mathcal{F}_i(x) = x$ (*i.e.,* linear Hawkes processes). Initially, the base intensity μ is sampled from a uniform distribution over $[0, 0.01]$ for each dimension, and the coefficients $\{\gamma_{ij}^I(t), \beta_{ij}, \gamma_{ij}^N(t)\}_{i,j\in[M]}$ are generated from a uniform distribution on $[0, 0.1]$. (d) CHASSIS-E: Similar to CHASSIS-L, the only difference is $\mathcal{F}_i(x) = e^x$ (*i.e.,* exponential Hawkes processes).

Datasets. We use the following datasets: (a) *Facebook*: We collect the data via Facebook Graph API (https://developers.facebook.com/docs/graph-api), comprising nearly 44 million public activities posted by 109, 211 individuals, from March 2018 to May 2018; (b) *Twitter*: We gather the data via Twitter Streaming API (https://developer.twitter.com/en/docs), containing nearly 52 million public activities posted by 123, 972 individuals, from March 2018 to May 2018. Additionally, we obtain the relationships among such individuals (*i.e.,* who follows whom) in each dataset, which could be converted into an excitation matrix $\boldsymbol{A} = [\alpha_{ij}]_{i,j\in[M]}$ ($\alpha_{ij} = 1$ if U_j follows U_i, otherwise $\alpha_{ij} = 0$) as the ground truth. Utilizing such relationships, we grab the offspring activities of each immigrant activity of each individual via a depth first search algorithm. We evaluate the scalability of CHASSIS using these two datasets and extract two subsets of the datasets: 590,671 activities posted by 100,000 individuals in *Facebook* (denoted as *SF*) and the other with 671,810 activities posted by 110,000 individuals in *Twitter* (denoted as *ST*), for other experiments.

Model Fitness. We study how well CHASSIS can explain the real-world data by comparing it with other strategies. We use two evaluation metrics, namely LogLike and RankCorr. Specifically, LogLike is the log-likelihood of the estimated

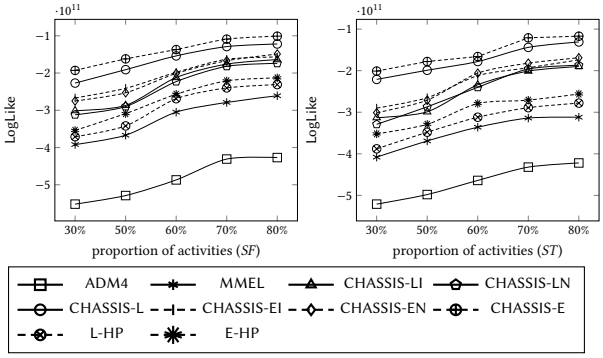


Figure 5. Model fitness (LogLike).

model on one test dataset and computed as $\ln \mathcal{L}(\mathbf{X}_{\text{test}}|\Theta_{\text{training}}) = \sum_{i\in[M]} \left(\sum_{k=1}^{N_i(t_{\text{test}})} \ln \lambda_i(t_{ik}) - \int_0^{t_{\text{test}}} \lambda_i(s)ds \right)$ [53]. RankCorr calculates the average Kendall's rank correlation coefficient between each row of *influence matrix* $\boldsymbol{A}$ and estimated $\hat{\boldsymbol{A}}$, to measure whether the relative order of the estimated social influences is correctly recovered [53]. We order all activities in a dataset chronologically, and use the first 30%, 50%, 60%, 70%, 80% samples for training, respectively.

Figure 5 shows the performance using LogLike on the testing activities. LogLike increases as the number of activities for training increases, indicating that more training data lead to better accuracy. Clearly, CHASSIS-L and CHASSIS-E perform significantly better than ADM4 and MMEL, which indicates that CHASSIS can capture the information diffusion better than the conformity-unaware strategies.

Is the superiority of CHASSIS is due to conformity-awareness or merely more flexible semi-parametric inference method? To answer this, we design two baselines, L-HP and E-HP, referring to our semi-parametric inference algorithm under linear and exponential Hawkes, respectively, with the intensities in Eq. 3.2. Both methods are conformity-unaware and are inferior to CHASSIS-L and CHASSIS-E (Figure 5). Hence, model fitness accuracy can be improved significantly when conformity is taken into account. On the other hand, both baselines exhibit better performance than ADM4 and MMEL. It implies that the proposed semi-parametric inference scheme also improves model fitness performance.

Additionally, we investigate the importance of modeling both informational and normative conformity in CHASSIS by disabling one of them in Eq. 4.2. Specifically, we remove $\sum_{i,j\in[M]} \gamma_{ij}^N(t)\alpha_{ij}^N(t)$ (resp. $\sum_{i,j\in[M]} \gamma_{ij}^I(t)\alpha_{ij}^I(t)$) , and only quantify $\sum_{i,j\in[M]} \gamma_{ij}^I(t)\alpha_{ij}^I(t)$ (resp. $\sum_{i,j\in[M]} \gamma_{ij}^N(t)\alpha_{ij}^N(t)$) in CHASSIS-LI (resp. CHASSIS-LN) and CHASSIS-EI (resp. CHASSIS-EN). As shown in Figure 5, CHASSIS-L (resp. CHASSIS-E) outperforms CHASSIS-LI (resp. CHASSIS-EI) and CHASSIS-LN (resp. CHASSIS-EN), endorsing the necessity of modeling both informational conformity and normative conformity. All these approaches also outperform L-HP and E-HP. Furthermore, CHASSIS-E (resp. CHASSIS-EI, CHASSIS-EN and E-HP) always has higher LogLike than CHASSIS-L (resp. CHASSIS-LI, CHASSIS-LN and L-HP), indicating that nonlinear Hawkes processes are more suitable for capturing the triggering patterns hidden in social activities.

Table 1. Branching structure inference performance

Dataset	Strategy			
	ADM4	MMEL	CHASSIS-L	CHASSIS-E
Charlie Hebdo	0.6547	0.7031	0.7966	0.8422
Sydney siege	0.6301	0.6908	0.7804	0.8380
Ferguson	0.6122	0.6743	0.7765	0.8201
Ottawa shooting	0.6003	0.6578	0.7523	0.8130
Germanwings-crash	0.5634	0.6020	0.7342	0.8002

The results using RankCorr are qualitatively similar to LogLike [32].

Convergence. The LogLike of CHASSIS-L and CHASSIS-E increase as the number of iterations grows and converge after 80 iterations on *Facebook* and *Twitter* data [32].

Inferring the Diffusion Trees. We now report the diffusion tree inference quality in CHASSIS using the datasets from PHEME (https://doi.org/10.6084/m9.figshare.6392078.v1) [27]. It contains *Twitter* conversation threads (*i.e.,* information cascades) associated with different newsworthy topics. Since the diffusion trees in each conversation are given, we use them as the ground truth for evaluating the inferred diffusion trees. We run the four Hawkes-based strategies on each dataset, and construct the diffusion trees accordingly. Comparing the inferred branching structure with the ground truth, we evaluate the effectiveness of strategies in terms of F1-Score. Table 1 reports the results. Observe that both CHASSIS-E and CHASSIS-L outperform the conformity-unaware strategies on all datasets, reemphasizing the importance of conformity in information diffusion. Again, CHASSIS-E outperforms CHASSIS-L significantly, indicating that nonlinear Hawkes processes are more appropriate than the linear ones.

9 CONCLUSIONS

A significant omission in existing online information diffusion models is the role played by conformity, a fundamental human trait according to social psychology theories. We propose a novel framework called CHASSIS to address this limitation by integrating conformity into Hawkes process-based information diffusion model. Specifically, we detect and quantify conformity by analyzing the diffusion trees. We propose an efficient semi-parametric inference algorithm, wherein the parametric evaluation procedure assists in identifying conformity of individuals, and the nonparametric procedure learns the triggering kernel functions flexibly in a data-driven way without the need of prior domain knowledge. Our experimental study not only demonstrates superiority of CHASSIS compared to conformity-unaware models but also emphasizes the pivotal role conformity plays in information diffusion.

REFERENCES

[1] Lisa R Anderson and Charles A Holt. 2006. Information cascades and rational conformity. *Encyclopedia of Cognitive Science* (2006).

[2] E. Aronson, T.D. Wilson, and R.M. Akert. 2007. *Social Psychology*. Pearson Education International.

[3] Akhil Arora, Sainyam Galhotra, and Sayan Ranu. 2017. Debunking the Myths of Influence Maximization: An In-Depth Benchmarking Study. In *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*. 651–666.

[4] Solomon E Asch. 1951. Effects of group pressure upon the modification and distortion of judgments. *Organizational influence processes* (1951), 295–303.

[5] Solomon E Asch. 1955. Opinions and social pressure. *Scientific American* 193, 5 (1955), 31–35.

[6] Prithu Banerjee, Wei Chen, and Laks V. S. Lakshmanan. 2019. Maximizing Welfare in Social Networks under A Utility Driven Influence Diffusion model. In *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*. 1078–1095.

[7] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* 100, 5 (1992), 992–1026.

[8] Steven J Breckler, James Olson, and Elizabeth Wiggins. 2005. *Social psychology alive*. Cengage Learning.

[9] Pierre Brémaud and Laurent Massoulié. 1996. Stability of nonlinear Hawkes processes. *The Annals of Probability* (1996), 1563–1588.

[10] Jennifer D Campbell and Patricia J Fairey. 1989. Informational and normative routes to conformity: The effect of faction size as a function of norm extremity and attention to the stimulus. *Journal of personality and social psychology* 57, 3 (1989), 457.

[11] Shuo Chen, Ju Fan, Guoliang Li, Jianhua Feng, Kian-Lee Tan, and Jinhui Tang. 2015. Online Topic-Aware Influence Maximization. *PVLDB* 8, 6 (2015), 666–677.

[12] Justin Cheng, Lada A. Adamic, P. Alex Dow, Jon M. Kleinberg, and Jure Leskovec. 2014. Can cascades be predicted?. In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*. 925–936.

[13] Nicolas Claidière and Andrew Whiten. 2012. Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological bulletin* 138, 1 (2012), 126.

[14] William D Crano. 1970. Effects of sex, response order, and expertise in conformity: A dispositional approach. *Sociometry* (1970), 239–252.

[15] Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology* 51, 3 (1955), 629.

[16] Kuntal Dey, Ritvik Shrivastava, and Saroj Kaushik. 2017. Twitter Stance Detection - A Subjectivity and Sentiment Polarity Inspired Two-Phase Approach. In *2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017*. 365–372.

[17] Sainyam Galhotra, Akhil Arora, and Shourya Roy. 2016. Holistic Influence Maximization: Combining Scalability and Efficiency with Opinion-Aware Models. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. 743–758.

[18] Jacob K Goeree, Thomas R Palfrey, Brian W Rogers, and Richard D McKelvey. 2007. Self-correcting information cascades. *The Review of Economic Studies* 74, 3 (2007), 733–762.

[19] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Schölkopf. 2011. Uncovering the Temporal Dynamics of Diffusion Networks. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*. 561–568.

[20] Caitlin Gray, Lewis Mitchell, and Matthew Roughan. 2018. Superblockers and the Effect of Network Structure on Information Cascades. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*. 1435–1441.

[21] Adrien Guille, Hakim Hacid, Cécile Favre, and Djamel A. Zighed. 2013. Information diffusion in online social networks: a survey. *SIGMOD Record* 42, 2 (2013), 17–28.

[22] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[23] Alan G Hawkes and David Oakes. 1974. A cluster process representation of a self-exciting process. *Journal of Applied Probability* 11, 3 (1974), 493–503.

[24] Herbert W. Hethcote. 2000. The Mathematics of Infectious Diseases. *SIAM Rev.* 42, 4 (2000), 599–653.

[25] Harold H Kelley and Martin M Shapiro. 1954. An experiment on conformity to group norms where conformity is detrimental to group achievement. *American Sociological Review* 19, 6 (1954), 667–677.

[26] David Kempe, Jon M. Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003*. 137–146.

[27] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. PHEME dataset for Rumour Detection and Veracity Classification.

[28] Dong Li, Shengping Zhang, Xin Sun, Huiyu Zhou, Sheng Li, and Xuelong Li. 2017. Modeling Information Diffusion over Social Networks for Temporal Dynamic Prediction. *IEEE Trans. Knowl. Data Eng.* 29, 9 (2017), 1985–1997.

[29] Guoliang Li, Shuo Chen, Jianhua Feng, Kian-Lee Tan, and Wen-Syan Li. 2014. Efficient location-aware influence maximization. In *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014*. 87–98.

[30] Hui Li, Sourav S. Bhowmick, and Aixin Sun. 2011. CASINO: towards conformity-aware social influence analysis in online social networks. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011*. 1007–1012.

[31] Hui Li, Sourav S. Bhowmick, Aixin Sun, and Jiangtao Cui. 2015. Conformity-aware influence maximization in online social networks. *VLDB J.* 24, 1 (2015), 117–141.

[32] Hui Li, Hui Li, and Sourav S Bhowmick. 2020. *CHASSIS: Conformity Meets Online Information Diffusion*. Technical Report. https://www.ntu.edu.sg/home/assourav/TechReports/CHASSIS-TR.pdf

[33] Yuchen Li, Ju Fan, Dongxiang Zhang, and Kian-Lee Tan. 2017. Discovering Your Selling Points: Personalized Social Influential Tags Exploration. In *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*. 619–634.

[34] Yiqing Li, Xiaoying Gan, Luoyi Fu, Xiaohua Tian, Zhida Qin, and Yanhong Zhou. 2018. Conformity-Aware Influence Maximization with User Profiles. In *10th International Conference on Wireless Communications and Signal Processing, WCSP 2018, Hangzhou, China, October 18-20, 2018*. 1–6.

[35] Linyuan Lu, Duanbing Chen, and Tao Zhou. 2011. Small world yields the most effective information spreading. *CoRR* abs/1107.0429 (2011).

[36] Michal Lukasik, Trevor Cohn, and Kalina Bontcheva. 2015. Point Process Modelling of Rumour Dynamics in Social Media. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2: Short Papers*. 518–523.

[37] Hongyuan Mei and Jason Eisner. 2017. The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*. 6754–6764.

[38] Mark E. J. Newman. 2003. The Structure and Function of Complex Networks. *SIAM Rev.* 45, 2 (2003), 167–256.

[39] Hung T. Nguyen, My T. Thai, and Thang N. Dinh. 2016. Stop-and-Stare: Optimal Sampling Algorithms for Viral Marketing in Billion-scale Networks. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. 695–710.

[40] Sen Pei, Lev Muchnik, Shaoting Tang, Zhiming Zheng, and Hernán A. Makse. 2015. Exploring the complex pattern of information spreading in online blog communities. *CoRR* abs/1504.00495 (2015).

[41] Chanthika Pornpitakpan. 2004. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology* 34, 2 (2004), 243–281.

[42] Liudmila Ostroumova Prokhorenkova, Alexey Tikhonov, and Nelly Litvak. 2019. Learning Clusters through Information Diffusion. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. 3151–3157.

[43] Jie Tang, Sen Wu, and Jimeng Sun. 2013. Confluence: conformity influence in large social networks. In *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*. 347–355.

[44] Youze Tang, Yanchen Shi, and Xiaokui Xiao. 2015. Influence Maximization in Near-Linear Time: A Martingale Approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015*. 1539–1554.

[45] William Trouleau, Jalal Etesami, Matthias Grossglauser, Negar Kiyavash, and Patrick Thiran. 2019. Learning Hawkes Processes Under Synchronization Noise. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. 6325–6334.

[46] Maria Nicolette Margaretha vanLieshout. 2006. Campbell and moment measures for finite sequential spatial processes. *CWI. Probability, Networks and Algorithms [PNA]* R 0601 (2006).

[47] David Vere-Jones. 2003. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer.

[48] Ke Wu, Song Yang, and Kenny Q. Zhu. 2015. False rumors detection on Sina Weibo by propagation structures. In *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*. 651–662.

[49] Shuang-Hong Yang and Hongyuan Zha. 2013. Mixture of Mutually Exciting Processes for Viral Diffusion. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*. 1–9.

[50] Linyun Yu, Peng Cui, Fei Wang, Chaoming Song, and Shiqiang Yang. 2015. From Micro to Macro: Uncovering and Predicting Information Cascading Process with Behavioral Dynamics. In *2015 IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17, 2015*. 559–568.

[51] Fue Zeng, Ran Tao, Yanwu Yang, and Tingting Xie. 2017. How social communications influence advertising perception and response in online communities? *Frontiers in psychology* 8 (2017), 1349.

[52] Jing Zhang, Jie Tang, Honglei Zhuang, Cane Wing-Ki Leung, and Juan-Zi Li. 2014. Role-Aware Conformity Modeling and Analysis in Social Networks. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*. 958–965.

[53] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning Social Infectivity in Sparse Low-rank Networks Using Multi-dimensional Hawkes Processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2013, Scottsdale, AZ, USA, April 29 - May 1, 2013*. 641–649.

[54] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning Triggering Kernels for Multi-dimensional Hawkes Processes. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*. 1301–1309.