INFLUENCE PREDICTION FOR CONTINUOUS-TIME INFORMATION PROPAGATION ON NETWORKS

Shui-Nee Chow

School of Mathematics, Georgia Institute of Technology
Atlanta, GA 30332, USA

Xiaojing Ye*

Department of Mathematics & Statistics, Georgia State University
Atlanta, GA 30303, USA

Hongyuan Zha

School of Computational Science and Engineering, Georgia Institute of Technology
Atlanta, GA 30332, USA

Haomin Zhou

School of Mathematics, Georgia Institute of Technology
Atlanta, GA 30332, USA

(Communicated by Axel Klar)

Abstract. We consider the problem of predicting the time evolution of influence, defined by the expected number of activated (infected) nodes, given a set of initially activated nodes on a propagation network. To address the significant computational challenges of this problem on large heterogeneous networks, we establish a system of differential equations governing the dynamics of probability mass functions on the state graph where each node lumps a number of activation states of the network, which can be considered as an analogue to the Fokker-Planck equation in continuous space. We provides several methods to estimate the system parameters which depend on the identities of the initially active nodes, the network topology, and the activation rates etc. The influence is then estimated by the solution of such a system of differential equations. Dependency of the prediction error on the parameter estimation is established. This approach gives rise to a class of novel and scalable algorithms that work effectively for large-scale and dense networks. Numerical results are provided to show the very promising performance in terms of prediction accuracy and computational efficiency of this approach.

1. **Introduction.** Viral signal propagation on large heterogeneous networks is an emerging research subject of both theoretical and practical importance. Influence prediction is one of the most fundamental problems about propagation on networks, and it has been arising from many real-world applications of significant societal impact, such as news spread on social media, viral marketing, computer malware

detection, and epidemics on heterogeneous networks. For instance, when considering a social network formed by people such as that of Facebook or Twitter, the viral signal can be a tweet or a trendy topic being retweeted by users (nodes) on the network formed by their followee-follower relationships. We call a user activated if he/she participates to tweet, and the followers of this user get activated if they retweet his/her tweet later, thus the activation process gradually progresses (propagates) and the tweet spreads out. A viral signal can also be a new electronic gadget that finds wide-spread adoption in the user population through a word-of-mouth viral marketing process [16, 17, 24], and a user is called activated when he/she adopts this new gadget. Influence prediction is to quantitatively estimate how influence, defined by the expected number of activated nodes, evolves over time during the propagation when a specific set (called source set) of nodes are initially activated.

Influence prediction is also the most critical step in solving problems arising from many important downstream applications such as influence maximization [7, 13, 14, 30] and outbreak detection [8, 17]. For instance, in influence maximization, the goal is to select the source node set of a given size from the propagation network such that its influence is maximized at a prescribed time. Obviously, influence prediction serves as the most fundamental subroutine in the computation, and the quality of influence maximization heavily depends on the accuracy of influence prediction.

1.1. **Problem description.** The influence prediction problem can be formulated as follows. Let $G = (V, E)$ be a given network (directed graph) with node (vertex) set $V = \{i : 1 \leq i \leq K\}$ and edge set $E \subset V \times V$. We denote $N_i^{\text{in}} := \{j : (j, i) \in E\}$ and $N_i^{\text{out}} := \{j : (i, j) \in E\}$. A piece of information on G can spread or propagate from an active node i to every inactive $j \in N_i^{\text{out}}$, and once succeeded, node j becomes active and starts to propagate information to inactive nodes in N_j^{out}, and so on. Once i is activated, the time elapsed for i to activate j follows certain probability. In addition to node-to-node activations, the nodes may also have the ability of self-activation. Namely, an inactive node i can be self-activated regardless of having any activated nodes in N_i^{in}. We assume the standard case that activated nodes cannot be activated again unless recovery scenario is considered. At any time, each node is in one of two states: inactive (susceptible) or active (infected). This model is called susceptible-infected (SI) model in classical mathematical epidemics theory. However, unlike most of existing works in this field, we here focus on efficient computational methods for influence prediction in the following settings due to practical concerns in real-world social networking applications.

- The network G is deterministically heterogeneous. This is significantly different from the case of classical SI model in mathematical biology/epidemics theory which does not consider contact network at individual level. Our network is also different from those heterogeneous but statistically homogeneous networks considered in statistical physics literature, where nodes can be partitioned into multiple categories according to certain properties, e.g., degrees, and the nodes within each category can be treated equivalently. In our case, the edges are explicitly given in the static network and the activation times have different but fixed distributions.
- Quantitative estimate of influence for time t before equilibrium. Note the equilibrium state of SI model on network is trivial: all nodes that can be reached from the source set will be infected as time tending to infinity. However, practical interests often lie in influence *before* equilibrium. For example, merchants

would like to know how many people will be influenced by commercial adver-
tisement within one month rather than three years later. In this case, we need
to estimate the time evolution of influence in early to middle stage where the
propagation is still in nonequilibrium state.

Our discussion also includes the case of self-activation where the unactivated nodes
can activate themselves automatically. If the infected nodes can recover, become
susceptible and prune to future infection, then the model is called susceptible-
infected-susceptible (SIS), for which we only provide a brief discussion near the end
of this paper.

Given network $G = (V, E)$, the stochastic propagation process is determined by
the distribution of the activation times between nodes. In this paper, we mainly
consider the model with activation times exponentially distributed which is widely
used in classical epidemic study and social network. In this model, the time for a
just activated node i to activate each unactivated j in N_i^{out}, denoted by $t_{i,j}$, follows
$\exp(\alpha_{ij})$ distribution (here t follows $\exp(\alpha)$ distribution if the probability density
function of t is $p_t(\tau) = \alpha e^{-\alpha\tau}$ for $\tau \geq 0$), and is independent of any other $t_{i',j'}$
where $i \neq i'$ and/or $j \neq j'$. Here $\alpha_{ij} > 0$ indicates the instantaneous activation
rate of j by i. If $(i, j) \notin E$, we set $\alpha_{ij} = 0$ by convention. Note that an exponential
random variable following $\exp(\alpha)$ has mean $1/\alpha$, therefore, the larger α_{ij} is, the
faster i can activate j on expectation. Hence α_{ij} can be interpreted as the impact
level (weight) of i on j. Similarly, the time for an inactive node i to get self activated
follows $\exp(\beta_i)$ for some $\beta_i > 0$. When recovery scenario is considered, an activated
node i can recover in time following $\exp(\gamma_i)$ for some $\gamma_i > 0$ since it is activated.

The continuous-time propagation model with heterogeneous activation rates ap-
pears suitable for a great number of real-world applications and has been advocated
by many recent works [11, 13, 23, 28, 29]. In addition, this model yields a time-
homogeneous Markov propagation process so that numerical simulations can be
implemented in a straightforward manner and some theoretical analysis of the algo-
rithm can be carried out. Therefore, we focus on the development of the algorithm
on this propagation model, and evaluate the performance numerically to obtain
references worthy of trust through a large amount of Monte Carlo simulations.
However, the general framework using Fokker-Planck equations–the main strategy
in the current work–as well as the error estimations developed in Section 2.3 apply
to any propagation models (e.g., activation time not exponentially distributed, such
as Hawkes processes) on networks.

To estimate the time evolution of the influence of a source set S, we define a
single stochastic process $N(t; S)$ as the number of activated nodes at time t when
the source set is S. Then we directly compute the probability distribution of $N(t; S)$:

$$\rho_k(t; S) := \Pr(N(t; S) = k), \quad \text{for } k = 0, 1, \dots, K. \tag{1}$$

The influence, defined by the expected number of activated nodes at time t, can
therefore be calculated easily by

$$\mu(t; S) = \mathbb{E}[N(t; S)] = \sum_{k=0}^{K} k\rho_k(t; S), \tag{2}$$

where $K := |V|$ is the size of the network.

The main focus of this paper is to establish a general framework for computing
(predicting) influence $\mu(t; S)$ based on (1) and (2) for any given source set S. More
precisely, we build the system of equations for the time evolution of $\{\rho_k(t; S) : 0 \leq$

$k \leq K$}, analyze its properties, estimate the parameters in the equations, and solve for all $\rho_k(t; S)$ to predict the influence $\mu(t; S)$ in (2) for all t. Since the source S is arbitrarily set in advance, we drop the symbol S in the derivation hereafter for notation simplicity.

The idea of deriving evolution equations of $\{\rho_k(t) : 0 \leq k \leq K\}$ is closely related to the theory of Fokker-Planck equation. In continuous space $\mathbb{R}^n$, consider a classical stochastic process $X(t)$ that stands for the location of a particle at time t. Let $\rho(x, t)$ denote the probability density that $X(t)$ is located at $x \in \mathbb{R}^n$ at time t, then $\rho(x, t)$ evolves over time with a constraint $\int_{\mathbb{R}^n} \rho(x, t)\mathrm{d}x = 1$ at every t. The Fokker-Planck equation, also known as the forward Kolmogorov equation, is a deterministic partial differential equation governing the time evolution of $\rho(x, t)$. For example, if $X(t)$ moves according to a stochastic differential equation (SDE) $\mathrm{d}X(t) = -\nabla\Psi(X(t))\mathrm{d}t + \sqrt{2\beta}\mathrm{d}W(t)$ where $W(t)$ is an n-dimensional Brownian motion and $\Psi(\cdot)$ is a scalar-valued potential function, then the Fokker-Planck equation of $\rho(x, t)$ is $\partial_t\rho(x, t) = \nabla \cdot (\nabla\Psi(x)\rho(x, t)) + \beta\Delta\rho(x, t)$. Here $\Delta\rho(x, t)$ corresponds to the $W(t)$ term in the SDE and is called the diffusion term, and $\nabla \cdot (\nabla\Psi(x)\rho(x, t))$ is the drift term. Note that the statistics of $X(t)$ can be completely determined by the solution $\rho(x, t)$ of the Fokker-Planck equation.

Likewise, the probabilities $\{\rho_k(t) : 0 \leq k \leq K\}$ in our approach also evolve over time with $\sum_{k=0}^{K} \rho_k(t) = 1$ at all t. The time evolution of $\rho_k(t)$ is also governed by certain Fokker-Planck equation which is now a system of deterministic ordinary differential equations since the state space is discrete $N(t) = 0, 1, \ldots, K$ rather than continuous $\mathbb{R}^n$. In recent years, there have been growing research interests in general graph-based Fokker-Planck equations to study problems related to optimal transport on finite graphs [5, 6, 12]. In the present paper, however, our goal is to find the Fokker-Planck equation that governs $\rho_k(t)$ in (1), and solve for these $\rho_k(t)$ to obtain the influence $\mu(t)$ using (2). We also analyze how the coefficient errors in Fokker-Planck equation affect the accuracy of predicting $\mu(t)$ using this approach.

1.2. **Related Work.** Previous study of influence estimation on networks is mainly restricted to statistically homogeneous and well-mixed populations, particularly in the context of statistical properties of dynamical processes on complex networks in physics. A comprehensive survey is provided in [23]. The typical approach is based on mean-field approximation (MFA) to establish a system of differential equations for the compartment model which groups nodes with statistically identical properties into one. For example, degree-based MFA groups nodes of the same degree which are considered to have identical behavior statistically, and hence can significantly reduce the size of the system [3]. Pair approximation includes the joint distribution in the system of equations, which essentially applies moment closure after the joint distribution of paired nodes, and is shown to have improved accuracy over standard MFA [1, 9, 22]. Other generalization and improvements of MFA and pair approximation using compartment models and motif expansions can be found in [9, 18, 19, 20, 27], and references therein. Recently, a generalization to the spatial SIS model in configuration space and its associated microscopic model for the spread of an infectious disease is developed in [2].

As noted in Section 1.1, our focus in this paper is instead on influence prediction (estimation) on *deterministically heterogeneous networks* particularly in *non-equilibrium* stage, which is significantly different from existing works including those mentioned above. For influence prediction in this setting, the prototype MFA for the Markov propagation model developed in [15, 31] is generalized to

arbitrary network topology [29], and then further extended to inhomogeneous activation and recovery rate between nodes [28]. The model adopts the MFA and the first-order moment closure, i.e., substituting the joint distribution of two activated neighbor nodes by the product of marginal distributions for individual nodes, to retain a feasible size of the derived system of differential equations. A second- (or higher-) order moment closure can be considered but the limitations and instability are discussed in [4]. Assuming absence of recovery, the exact solution is available due to the Markov property of propagation, however, its computational complexity increases drastically in terms of size and density of general networks [13]. As an alternative to solving for influence based on evolution equations, methods based on sampling propagations (also called cascades) and statistical learning technique are also developed, but often posing various requirements on input data and output results. For instance, a scalable computational method based on learning the coverage function of each node based on sampling and kernel estimation is developed, which can can only predict the influence at a prescribed time [11]. The work is further extended to estimate the time-varying intensity of propagation using similar coverage function idea [10]. Learning-based methods are usually companioned with a great amount of accuracy analysis based on classical theory of sampling complexity. However, the major problem with learning-based approaches is in the use of large amount of samplings to ensemble the unknown function or probability of interests but lack of a comprehensive understanding of the underlying dynamics and unique properties associated with the stochastic propagation on networks. Moreover, learning-based methods can have special assumptions on data which may not be realistic in real-world applications. To achieve moderate accuracy level in large-scale and complex network, learning-based methods require extensive amount of sampling/simulations, which causes significant computational burden and hinders their applicability in real-world problems.

2. **Proposed method.** In this section, we first derive the Fokker-Planck equation for the probabilities $\rho_k(t)$ of $N(t)$ for the propagation model with exponentially distributed activation times. We provide two effective methods to estimate the coefficients in the Fokker-Planck equation for large heterogeneous networks. Then we establish the relation between the estimation error of the coefficients in the Fokker-Planck equation and the accuracy in the predicted influence for general propagation models using our approach.

2.1. **The Fokker-Planck equation of $\rho_k(t)$.** Let $G = (V, E)$ and $\{\alpha_{ij} : (i, j) \in E\}$ (and $\{\beta_i : i \in V\}$ for self-activation and $\{\gamma_i : i \in V\}$ for recovery) be given and the source set S be chosen arbitrarily. The number of activated nodes, $N(t)$, has $K + 1$ states corresponding to $N(t) = 0, 1, \ldots, K$. Let M_k denote the state that $N(t) = k$ nodes in G are activated. Then, for the general SIS model with self-activation and recovery, the transitions between states of $N(t)$ can be illustrated as follows,

$$\boxed{M_0} \rightleftharpoons \cdots \rightleftharpoons \boxed{M_{k-1}} \overset{q_{k-1}(t)}{\underset{r_k(t)}{\rightleftharpoons}} \boxed{M_k} \overset{q_k(t)}{\underset{r_{k+1}(t)}{\rightleftharpoons}} \boxed{M_{k+1}} \rightleftharpoons \cdots \rightleftharpoons \boxed{M_K} \qquad (3)$$

Here, $q_k(t)$ is the transition rate from M_k to M_{k+1} and $r_k(t)$ is the rate from M_k to M_{k-1} at time t, and they depend on the structure of $G = (V, E)$, the activation parameters α_{ij} (and β_i and γ_i for self-activation and recovery respectively), and the source set S.

Recall that $\rho_k(t)$ is the probability of $N(t)$ being in state M_k according to definition (1). Therefore, the time evolution of $\rho_k(t)$ is governed by the discrete Fokker-Planck equation with these $q_k(t)$ and $r_k(t)$ to be determined:

$$\rho'_0(t) = -q_0(t)\rho_0(t) + r_1(t)\rho_1(t),$$
$$\rho'_k(t) = q_{k-1}(t)\rho_{k-1}(t) - [q_k(t) + r_k(t)]\rho_k(t) + r_{k+1}(t)\rho_{k+1}(t),\ 0 < k < K, \quad (4)$$
$$\rho'_K(t) = q_{K-1}(t)\rho_{K-1}(t) - r_K(t)\rho_K(t).$$

To rewrite (4) into a concise matrix formulation, we define two $(K+1) \times (K+1)$ matrices $Q(t)$ and $R(t)$ as follows:

$$[Q(t)]_{j,j} = -q_{j-1}(t), \quad [Q(t)]_{j,j+1} = q_{j-1}(t), \quad j = 1, \ldots, K \quad (5)$$
$$[R(t)]_{j,j} = -r_{j-1}(t), \quad [R(t)]_{j,j-1} = r_{j-1}(t), \quad j = 2, \ldots, K+1. \quad (6)$$

and all other entries are zeros. Here $[P]_{j,l}$ stands for the (j,l)-th entry of matrix P. Note that only the diagonal and superdiagonal (subdiagonal) entries of $Q(t)$ $(R(t))$ are nonzeros, and $[Q(t)]_{K+1,K+1} = [R(t)]_{1,1} = 0$ for all t. With matrices $Q(t)$ and $R(t)$ given above, we define a row $(K+1)$-vector $\rho(t) := (\rho_0(t), \rho_1(t), \ldots, \rho_K(t))$ and rewrite (4) as

$$\rho'(t) = \rho(t)[Q(t) + R(t)]. \quad (7)$$

The system (7) is consistent with the nature of process $N(t)$ in (3) with a tridiagonal transition matrix $Q(t) + R(t)$. The initial value $\rho(0)$ can be easily determined given S: let $|S|$ denote the cardinality of S, then $\rho(0)$ is a binary $(K+1)$-vector such that $\rho_{|S|}(0) = 1$ and $\rho_k(0) = 0$ for all $k \neq |S|$. Therefore, we can solve (4) for $\rho(t)$ to obtain the influence $\mu(t)$ based on (2) once the transition rates $q_k(t)$ and $r_k(t)$ are determined. The following subsection is devoted to the estimation of these rates.

2.2. **Estimation of transition rates** $q_k(t)$ **and** $r_k(t)$. Recall that $q_k(t)$ stands for the transition rate of $N(t)$ from M_k to M_{k+1} as shown in (3). Namely, $q_k(t)$ is the instantaneous rate for the $(k+1)$-th node to be activated given that there are currently k activated node (with numerous possible choices of such k nodes in V and $q_k(t)$ aggregates all the information) at time t. Similarly, $r_k(t)$ is the instantaneous rate for any of these k activated nodes to get recovered. Therefore, we focus on the estimation of $q_k(t)$ and a similar derivation can be easily carried out for $r_k(t)$.

The estimation of rate $q_k(t)$ consists of two factors: (i) the identities of the k currently activated nodes; and (ii) the instantaneous activation rate imposed by these k nodes to all the unactivated nodes at the time t. For factor (ii), the propagation model with exponentially distributed activation times yield constant instantaneous rates if the identities of the k nodes are given. For factor (i), $q_k(t)$ need to aggregate all the $\binom{K}{k}$ possible combinations of k activated nodes. The following theorem provides the compositions of $q_k(t)$ and $r_k(t)$. Here we call U activated if all nodes in U are activated and the others in $U^c = U \setminus V$ are unactivated. The proof frequently calls two simple facts about selection probability given multiple instantaneous rates, which we provide as Propositions 1 and 2 in the Appendix for completeness.

Theorem 2.1. *For every $k = 0, 1, \ldots, K$, let $\mathcal{S}_k := \{U \subset V : |U| = k\}$ be the collection of all subsets of size k in V. Let $\Pr(t; U)$ be the probability that U is*

activated among those in $\mathcal{S}_k$, and define

$$\alpha(U) = \sum_{i \in U} \sum_{j \in N_i^{\mathrm{out}} \cap U^c} \alpha_{ij}, \quad \beta(U) = \sum_{i \in U} \beta_i, \quad \gamma(U) = \sum_{i \in U} \gamma_i. \tag{8}$$

Then the transition rates $q_k(t)$ and $r_k(t)$ in (4) are given by

$$q_k(t) = \sum_{U \in \mathcal{S}_k} [\alpha(U) + \beta(U^c)] \Pr(t; U) \quad and \quad r_k(t) = \sum_{U \in \mathcal{S}_k} \gamma(U) \Pr(t; U). \tag{9}$$

Proof. Suppose nodes in $U \subset \mathcal{S}_k$ are currently activated, then these nodes tend to activate their neighbors still in U^c independently and simultaneously. More precisely, node $i \in U$ imposes a node-to-node activation rate to each of its unactivated neighbor $j \in N_i^{\mathrm{out}} \cap U^c$ independently and simultaneously, and hence total rate is $\sum_{j \in N_i^{\mathrm{out}} \cap U^c} \alpha_{ij}$. Therefore, the combined instantaneous rate of all nodes in U for node-to-node activation is given by $\alpha(U)$ in (8) according to Proposition 1. Meanwhile, each node i in U^c tends to be self activated with rate β_i and hence their total instantaneous self-activation rate is $\beta(U)$ given in (8). As $\Pr(t; U)$ is the probability that nodes in U are activated, we obtain the total instantaneous rate $q_k(t)$ as for $N(t)$ to transit from state M_k to M_{k+1} as (9) according to Proposition 2. The derivation for $r_k(t)$ in (9) follows similarly. $\square$

Note that there are $|\mathcal{S}_k| = \binom{K}{k}$ possible combinations U and $\sum_{U \in \mathcal{S}_k} \Pr(t; U) = 1$ for all t. Moreover, $q_k(t)$ is a convex combination of the instantaneous activation rates $\alpha(U) + \beta(U^c)$ with weights given by $\Pr(t; U)$ according to Theorem 2.1. Hence $q_k(t)$ is closer to the $\alpha(U) + \beta(U^c)$ with larger $\Pr(t; U)$. The composition of $r_k(t)$ in Theorem 2.1 has similar interpretation. Although it is not practical to obtain the probability $\Pr(t; U)$ for all U, Theorem 2.1 suggests that we can approximate $q_k(t)$ and $r_k(t)$ using the activation and recovery rates of those U with large weight $\Pr(t; U)$.

We now present two estimation methods and practical implementations using this idea for the case without self-activation and recovery (which essentially yields the standard susceptible-infection (SI) propagation model) on heterogeneous networks. In this case, we have $q_k(t) = \sum_{U \in \mathcal{S}_k} \alpha(U) \Pr(t; U)$ and $r_k(t) = 0$. Therefore, the key is to approximate $q_k(t)$ using $\alpha(U)$ of few U with the largest probabilities $\Pr(t; U)$.

Estimate q_k based on the shortest distance. For every $k = 1, \ldots, K$, we can easily determine a combination U_k^* with large $\Pr(t; U)$ over all U in $\mathcal{S}_k$ as follows: recall that the expected time for node i to activate $j \in N_i^{\mathrm{out}}$ is $1/\alpha_{ij}$, it is therefore natural to define the distance from i to j as $D(i, j) := 1/\alpha_{ij}$, which can also be generated to the distance from set S to a node j as $D(S, j) := \min_{i \in S} D(i, j)$. Due to independency of all node-to-node activations and property of exponential distributions, the set U_k^* consisting of the k nodes with the shortest distance to source S has larger In practical implementation, we apply Dijkstra's method [26] on the weighted graph G with edge weights given by $1/\alpha_{ij}$ and origin S, and then sort the nodes as $i_1, i_2, \ldots, i_K$ with ascending distance from source S, i.e., $D(S, i_1) \leq D(S, i_2) \leq \cdots \leq D(S, i_K)$ (if $i \in S$ then $D(S, i) = 0$) and set $U_k^* = \{i_1, \ldots, i_k\}$ for $k = 1, \ldots, K$. Then we approximate $q_k(t)$ by $\hat{q}_k(t) = \alpha(U_k^*)$, which remains as constant for all t once the source set S is given. This method is referred to as `FPE-dist` in the numerical experiments.

Estimate q_k based on the largest overall probabilities. To refine the approximation using single U_k^* in `FPE-dist`, we can estimate $q_k(t)$ using multiple combinations U with the largest probabilities. For a fixed S, we employ the following recursive

method to determine the sets $\{U_k^1, \ldots, U_k^{m_k}\} \subset \mathcal{S}_k$ to be used in calculation of q_k in (9) for $k = 1, 2, \ldots, K-1$. Here m_k is a user-customized number of k-combinations selected from $\mathcal{S}_k$ (larger m_k yields more accurate approximation to q_k at the expense of higher computation complexity.) Suppose we have already obtained $U_k^1, \ldots, U_k^{m_k}$ for k such that $\Pr(U_k^1) \geq \cdots \geq \Pr(U_k^{m_k})$, our next step is to obtain $U_{k+1}^1, \ldots, U_{k+1}^{m_{k+1}}$. To this end, we proceed with previous U_k^l in the order of $l = 1, \ldots, m_k$ and compute $\alpha(j|U_k^l)$ for every neighbor node j of U_k^l (by neighbor j of a subset U we meant that $j \in N_i^{\text{out}}$ for some $i \in U$, and $\alpha(j|U) := \sum_{i \in U \cap N_j^{\text{in}}} \alpha_{ij}$ is the total activation rate imposed to j by nodes in U.) By Proposition 1, neighbor j of U_k^l will be activated before other neighbors j' with probability $\Pr(j|U_k^l) = \alpha(j|U_k^l)/\sum_{j'} \alpha(j'|U_k^l)$ where the summation in the denominator is over all neighbors j' of U_k^l. Therefore, $\Pr(U) = \Pr(j|U_k^l)\Pr(U_k^l)$ for $U := U_k^l \cup \{j\}$ and $T(U_k^l) \leq t_j$. Note that each neighbor j of U_k^l yields such a U of size $k+1$. All these U's are then candidates for $U_{k+1}^1, \ldots, U_{k+1}^{m_{k+1}}$ later. We proceed with each U_k^l in the aforementioned way for $l = 1, \ldots, m_k$ and obtain a number of sets U's with probabilities $\Pr(U)$. Note that if two or more of these U's are identical, then we keep only one of them and merge their probabilities $\Pr(U)$. Then we sort these U's with $\Pr(U)$ in descending order and only keep the first m_{k+1} as $U_{k+1}^1, \ldots, U_{k+1}^{m_{k+1}}$. By this way, we are likely (but not guaranteed) to maintain a list $\{U_k^1, \ldots, U_k^{m_k}\}$ with the largest probabilities among all those in $\mathcal{S}_k$ for each k. Then we approximate $q_k(t)$ by $\hat{q}_k(t) := \sum_{l=1}^{m_k} \alpha(U_k^l)\Pr(U_k^l)$ which is again constant for all t. This method essentially constructs a branching tree with $K+1$ layers, where the layer k consists of m_k nodes $U_k^1, \ldots, U_k^{m_k}$ each having a relative probability in its layer, and the others with small probabilities in $\mathcal{S}_k$ are removed so that the computation complexity is maintained within a feasible scale. We refer this method to as `FPE-tree` in the numerical experiments.

Once we obtained the estimate $\hat{q}_k(t)$, the last step is to solve the Fokker-Planck equation $\rho'(t) = \rho(t)Q(t)$ numerically. There are two straightforward methods to compute $\rho(t)$: the Runge-Kutta method which can handle time varying $Q(t)$ and very large K (with computation complexity $O(K)$) but needs to proceed the computation starting from $t = 0$; and direct computation of $\rho(t) = \rho(0)e^{\int_0^t Q(s)ds}$ with bidiagonal matrix Q. In particular, if Q is constant, then the computation $\rho(t) = \rho(0)e^{tQ}$ is very fast using matrix exponential [21, 25, 32] and can be directly done for any specific $t > 0$ rather than from $t = 0$.

The steps for influence prediction using Fokker-Planck equation (4) is summarized in Algorithm 1. For completeness we include the self-activation and recovery rates.

Algorithm 1 Influence prediction based on Fokker-Planck equation (7)

1: **input** $G = (V, E)$, $\{\alpha_{ij}, \beta_i, \gamma_i : (i,j) \in E, i \in V\}$. Give source set $S \subset V$.
2: Estimate $\{q_k(t), r_k(t) : t \geq 0\}$ defined in (9) and form matrices $Q(t), R(t)$ as in (5)-(6).
3: Solve $\rho'(t) = \rho(t)[Q(t) + R(t)]$ with initial $\rho(0)$ to obtain $\rho(t)$.
4: **return** Output influence $\mu(t) = \sum_{k=0}^K k\rho_k(t) = \rho(t)(0, 1, \ldots, K)^T$.

2.3. **Error estimate for influence prediction.** In this section, we conduct the error analysis of the proposed influence prediction method. For simplicity, we consider the case without recovery scenario, and assume that the propagation starts

with self-activation, i.e., $\rho(0) = (1, 0, \ldots, 0) \in \mathbb{R}^{K+1}$, since derivations generalize to other initials trivially. It is worth noting that the results obtained in this section apply to any propagation model.

We first observe that the solution $\rho(t) = (\rho_0(t), \ldots, \rho_K(t))$ of $\rho'(t) = \rho(t)Q(t)$ with initial value $\rho(0) = (1, 0, \ldots, 0)$ is

$$\rho_0(t) = e^{-\int_0^t q_0(s)\mathrm{d}s},$$

$$\rho_{k+1}(t) = \int_0^t \rho_k(s)\, q_k(s)\, e^{-\int_s^t q_{k+1}(u)\mathrm{d}u}\mathrm{d}s, \text{ for } k = 0, 1, \ldots, K-2, \qquad (10)$$

$$\rho_K(t) = \int_0^t \rho_{K-1}(s)q_{K-1}(s)\mathrm{d}s.$$

Now, for every $k = 0, 1, \ldots, K-1$, let Q_k denote the perturbed rate matrix as

$$Q_k(t) = \begin{pmatrix} \ddots & & \ddots & & \\ & -q_{k-1}(t) & q_{k-1}(t) & 0 & \\ & 0 & -\hat{q}_k(t) & \hat{q}_k(t) & \\ & & & \ddots & \ddots \end{pmatrix} \qquad (11)$$

That is, $Q_k(t)$ differs from the orignal $Q(t)$ by replacing $q_j(t)$ with $\hat{q}_j(t)$ for $j = k, k+1, \ldots, K-1$. Then we have the following lemma that relates the error in solution $\rho(t)$ to the error in estimating $q_k(t)$.

Lemma 2.2. *Let $\epsilon \in (0, 1)$, and ρ and $\hat{\rho}$ be the solutions of $\rho'(t) = \rho(t)Q_{k+1}(t)$ and $\hat{\rho}'(t) = \hat{\rho}(t)Q_k(t)$, respectively. Denote $\delta_k(t) := |\hat{q}_k(t) - q_k(t)|/q_k(t)$. If $\bar{\alpha} > 0$ is the upper bound of all activation rates between nodes in $G = (V, E)$ and that*

$$\delta_k(t) \le \min\left\{ \frac{\log(1 + \frac{\epsilon}{2})}{\bar{\alpha}kt\min(\bar{d}, K-k)}, \frac{\epsilon}{2+\epsilon} \right\} \qquad (12)$$

where $\bar{d} = \max\{|N_i^{\mathrm{out}}| : i \in V\}$, then $\rho_j(t) = \hat{\rho}_j(t)$ for $j = 0, 1, \ldots, k-1$ and $|\hat{\rho}_j(t) - \rho_j(t)|/\rho_j(t) \le \epsilon$ for $j = k, \ldots, K$ and all $t > 0$. Moreover, $|\hat{\mu}(t) - \mu(t)|/\mu(t) \le \epsilon$ for all t.

Proof. If $k > 0$, from the solution formulation (10), we know that $\rho_j(t) = \hat{\rho}_j(t)$ for all t and $j = 0, 1, \ldots, k-1$. Furthermore, there are

$$\rho_k(t) = \int_0^t \rho_{k-1}(s)q_{k-1}(s)e^{-\int_s^t q_k(u)\mathrm{d}u}\mathrm{d}s, \qquad (13)$$

$$\hat{\rho}_k(t) = \int_0^t \rho_{k-1}(s)q_{k-1}(s)e^{-\int_s^t \hat{q}_k(u)\mathrm{d}u}\mathrm{d}s. \qquad (14)$$

Since $q_k(t) \le \bar{\alpha}k\min(\bar{d}, K-k)$ and (12), there are $\int_0^t \delta_k(s)q_k(s)\mathrm{d}s \le \log(1 + \frac{\epsilon}{2})$ and

$$\left| e^{-\int_s^t (\hat{q}_k(u) - q_k(u))\mathrm{d}u} - 1 \right| \le e^{\int_s^t \delta_k(u)q_k(u)\mathrm{d}s} - 1 \le e^{\int_0^t \delta_k(u)q_k(u)\mathrm{d}s} - 1 \le \frac{\epsilon}{2} \qquad (15)$$

for all $s \in (0, t)$. Therefore, from (13) and (14) there is

$$\frac{|\hat{\rho}_k(t) - \rho_k(t)|}{\rho_k(t)} \le \frac{1}{\rho_k(t)} \int_0^t \rho_{k-1}(s)q_{k-1}(s)e^{-\int_s^t q_k(u)\mathrm{d}u} \left| e^{-\int_s^t (\hat{q}_k(u) - q_k(u))\mathrm{d}u} - 1 \right| \mathrm{d}s$$

$$\le \frac{\epsilon}{2\rho_k(t)} \int_0^t \rho_{k-1}(s)q_{k-1}(s)e^{-\int_s^t q_k(u)\mathrm{d}u}\mathrm{d}s = \frac{\epsilon}{2}.$$

If $k = 0$, then $\rho_0(t) = e^{-\int_0^t q_0(s)\mathrm{d}s}$ and $\hat{\rho}_0(t) = e^{-\int_0^t \hat{q}_0(s)\mathrm{d}s}$ and one can check that this inequality still holds.

As $|\hat{\rho}_k(t) - \rho_k(t)|/\rho_k(t) \leq \epsilon/2$, there is $\hat{\rho}_k(t) \leq (1 + \epsilon/2)\rho_k(t)$ and hence

$$
|\rho_k(t)q_k(t) - \hat{\rho}_k(t)\hat{q}_k(t)| \leq |\rho_k(t) - \hat{\rho}_k(t)|q_k(t) + \hat{\rho}_k(t)|q_k(t) - \hat{q}_k(t))|
$$
$$
\leq \frac{\epsilon}{2}\rho_k(t)q_k(t) + \left(1 + \frac{\epsilon}{2}\right)\rho_k(t)\delta_k(t)q_k(t) \leq \epsilon\rho_k(t)q_k(t) \quad (16)
$$

for all $t \geq 0$, where we used the fact that $(1 + \frac{\epsilon}{2})\delta_k(t) \leq \frac{\epsilon}{2}$ from (12). Due to the general formulation of solution (10), there are

$$
\rho_{k+1}(t) = \int_0^t \rho_k(s)q_k(s)e^{-\int_s^t \hat{q}_{k+1}(u)\mathrm{d}u}\mathrm{d}s \tag{17}
$$

$$
\hat{\rho}_{k+1}(t) = \int_0^t \hat{\rho}_k(s)\hat{q}_k(s)e^{-\int_s^t \hat{q}_{k+1}(u)\mathrm{d}u}\mathrm{d}s \tag{18}
$$

Then we can bound their difference as follows,

$$
\frac{|\hat{\rho}_{k+1}(t) - \rho_{k+1}(t)|}{\rho_{k+1}(t)} \leq \frac{1}{\rho_{k+1}(t)} \int_0^t |\rho_k(s)q_k(s) - \hat{\rho}_k(s)\hat{q}_k(s)|e^{-\int_s^t \hat{q}_{k+1}(u)\mathrm{d}u}\mathrm{d}s
$$
$$
\leq \frac{\epsilon}{\rho_{k+1}(t)} \int_0^t \rho_k(s)q_k(s)e^{-\int_s^t \hat{q}_{k+1}(u)\mathrm{d}u}\mathrm{d}s = \epsilon.
$$

For $j = k+1, \ldots, K$, $\hat{q}_j(t)$ is the same for both $\rho(t)$ and $\hat{\rho}(t)$ in (10), one can readily check that $|\hat{\rho}_j(t) - \rho_j(t)|/\rho_j(t) \leq \epsilon$ implies that $|\hat{\rho}_{j+1}(t) - \rho_{j+1}(t)|/\rho_{j+1}(t) \leq \epsilon$. Therefore,

$$
\frac{|\hat{\mu}(t) - \mu(t)|}{\mu(t)} \leq \frac{1}{\mu(t)} \sum_{j=k}^K j|\hat{\rho}_j(t) - \rho_j(t)| \leq \frac{\epsilon}{\mu(t)} \sum_{j=k}^K j\rho_j(t) \leq \epsilon \tag{19}
$$

for all $t \geq 0$, which completes the proof. $\square$

Theorem 2.3. *Let $\epsilon \in (0, 1)$, and $\rho(t)$ and $\hat{\rho}(t)$ be the solutions of $\rho'(t) = \rho(t)Q(t)$ and $\hat{\rho}'(t) = \hat{\rho}(t)\hat{Q}(t)$ where $\hat{Q}(t) := Q_0(t)$, respectively, and $\mu(t) = \sum_{k=0}^K k\rho_k(t)$ and $\hat{\mu}(t) = \sum_{k=0}^K k\hat{\rho}_k(t)$. If (12) holds for $k = 0, \ldots, K-1$ and there exist upper bound $\bar{\alpha}$ and lower bound $\underline{\alpha} > 0$ for all activation rates in $G = (V, E)$, then*

$$
\frac{|\hat{\mu}(t) - \mu(t)|}{\mu(t)} \leq [(1 + \epsilon)^K - 1] \min\left\{1, c_K(t)e^{-\underline{\alpha}t}\right\}, \quad \forall t \geq 0, \tag{20}
$$

where $\bar{q} := \max_k\{q_k\}$ is bounded and $c_K(t) := \frac{1}{K} \sum_{j=0}^{K-1} \frac{K-j}{j!}(\bar{q}t)^j$.

Proof. For every $k = 0, 1, \ldots, K-1$, let $\mu_k(t)$ be the influence estimated by solving differential equation $\rho'(t) = \rho(t)Q_k(t)$. Then Lemma 2.2 shows that $|\mu_k(t) - \mu_{k+1}(t)|/\mu_{k+1}(t) \leq \epsilon$ for $k = 0, \ldots, K-2$ and $|\mu_{K-1} - \mu(t)|/\mu(t) \leq \epsilon$ provided (12) holds for all k. Therefore $1 - \epsilon \leq \frac{\mu_k(t)}{\mu_{k+1}(t)} \leq 1 + \epsilon$ and $1 - \epsilon \leq \frac{\mu_{K-1}(t)}{\mu(t)} \leq 1 + \epsilon$, and hence

$$
(1 - \epsilon)^K \leq \frac{\hat{\mu}(t)}{\mu(t)} = \frac{\mu_0(t)}{\mu(t)} = \frac{\mu_{K-1}(t)}{\mu(t)} \cdots \frac{\mu_1(t)}{\mu_2(t)} \frac{\mu_0(t)}{\mu_1(t)} \leq (1 + \epsilon)^K. \tag{21}
$$

Therefore $|\hat{\mu}(t) - \mu(t)|/\mu(t) \leq \max\{1 - (1 - \epsilon)^K, (1 + \epsilon)^K - 1\} = (1 + \epsilon)^K - 1$.

On the other hand, we have $\underline{\alpha} \leq q_k(t) \leq \bar{\alpha}k\min\{\bar{d}, K-d\}$ and hence $\underline{\alpha} \leq q_k(t) \leq \bar{q}$ for all $k = 0, 1, \ldots, K-1$ and $t \geq 0$. Here $\bar{q} \leq \bar{\alpha}\bar{d}(K - \bar{d})$ if $\bar{d} \leq \frac{K}{2}$ and $\bar{q} \leq \frac{\bar{\alpha}K^2}{4}$

if $\bar{d} > \frac{K}{2}$. By induction we claim that $\rho_k(t) \leq \frac{(\bar{q}t)^k}{k!} e^{-\underline{\alpha}t}$ for $k = 0, \ldots, K-1$ as follows: the claim is obviously true for $k = 0$; suppose it is true for $k \leq K-2$, then

$$\rho_{k+1}(t) = \int_0^t \rho_k(s) q_k(s) e^{-\int_s^t q_{k+1}(u)du} ds \leq \int_0^t \frac{(\bar{q}s)^k}{k!} e^{-\underline{\alpha}s} \bar{q} e^{-\underline{\alpha}(t-s)} ds$$
$$= \frac{\bar{q}^{k+1} e^{-\underline{\alpha}t}}{k!} \int_0^t s^k ds = \frac{(\bar{q}t)^{k+1}}{(k+1)!} e^{-\underline{\alpha}t}.$$

Moreover, from Lemma (2.2) we can readily deduce that $(1-\epsilon)^{j+1} \leq \hat{\rho}_j(t)/\rho_j(t) \leq (1+\epsilon)^{j+1}$ similar as for (21). Hence $|\hat{\rho}_j(t) - \rho_j(t)|/\rho_j(t) \leq \epsilon_j := (1+\epsilon)^{j+1} - 1$ for $j = 0, 1, \ldots, K-1$. Therefore, we have

$$\frac{|\hat{\mu}(t) - \mu(t)|}{\mu(t)} = \frac{1}{\mu(t)} \left| \sum_{j=0}^K j(\hat{\rho}_j(t) - \rho_j(t)) \right| = \frac{1}{\mu(t)} \left| \sum_{j=0}^{K-1} (K-j)(\hat{\rho}_j(t) - \rho_j(t)) \right|$$

$$\leq \frac{1}{\mu(t)} \sum_{j=0}^{K-1} (K-j) |\hat{\rho}_j(t) - \rho_j(t)| \leq \frac{1}{\mu(t)} \sum_{j=0}^{K-1} (K-j) \epsilon_j \rho_j(t)$$

$$\leq \frac{\epsilon_{K-1}}{|S|} \sum_{j=0}^{K-1} (K-j) \rho_j(t) \leq \frac{\epsilon_{K-1} e^{-\underline{\alpha}t}}{|S|} \sum_{j=0}^{K-1} \frac{K-j}{j!} (\bar{q}t)^j$$

$$= \epsilon_{K-1} c_K(t) e^{-\underline{\alpha}t}$$

where we used the fact that $\rho_K(t) = 1 - \sum_{j=0}^{K-1} \rho_j(t)$ in the second equality, $\epsilon_{K-1} \geq \epsilon_j$ for all $j = 0, \ldots, K-1$ and $\mu(t) \geq |S|$ in the fourth inequality[1]. Combining the two bounds of $|\hat{\mu}(t) - \mu(t)|/\mu(t)$ above, we obtain (20). $\qquad \square$

Theorem 2.3 shows that an $O(1/t)$ decay of error in estimated $\hat{q}_k(t)$ results in an exponential $O(e^{-\underline{\alpha}t})$ decay of error in predicted influence $\hat{\mu}(t)$. This result implies that for an exponentially decaying error in $\hat{\mu}(t)$ the estimation error in $\hat{q}_k(t)$ only needs to remain about as constant for all sufficiently large t.

Corollary 1. *Suppose $\rho(t), \hat{\rho}(t), \mu(t), \hat{\mu}(t)$ are defined and conditions for $\bar{\alpha}$ and $\underline{\alpha}$ hold as in Theorem 2.3. Let $\varepsilon > 0$ and $c \in (0, \underline{\alpha})$, then $|\hat{\mu}(t) - \mu(t)|/\mu(t) \leq \varepsilon e^{-ct}$ as long as the estimated $\hat{q}_k(t)$ satisfies*

$$\frac{|\hat{q}_k(t) - q_k(t)|}{q_k(t)} \leq \frac{\underline{\alpha} - c}{K\bar{q}_k} + \frac{\log \varepsilon - K \log 2 - \log c_K(t)}{K\bar{q}_k t} = C_k - O\left(\frac{\log t}{t}\right) \quad (22)$$

for each $k = 0, 1, \ldots, K-1$, where $\bar{q}_k := \bar{\alpha}k \min\{\bar{d}, K-k\}$ and $C_k := (\underline{\alpha} - c)/K\bar{q}_k$.

Proof. By Theorem 2.3 and the bound of error $\delta_k(t)$ in (12), we can attain $|\hat{\mu}(t) - \mu(t)|/\mu(t) \leq \varepsilon e^{-ct}$ as long as $\delta_k(t)$ satisfies $\bar{q}_k t \delta_k(t) \leq \log(1 + \epsilon(t))$ for some $\epsilon(t)$ such that $[(1 + 2\epsilon(t))^K - 1]c_K(t)e^{-\underline{\alpha}t} = \varepsilon e^{-ct}$. To this end, we need $\log(2e^{\bar{q}_k t \delta_k(t)} - 1) \leq \frac{1}{K} \log(\frac{\varepsilon e^{(\underline{\alpha} - c)t}}{c_K(t)} + 1)$, to guarantee which it suffices to have $\log(2e^{\bar{q}_k t \delta_k(t)}) \leq \frac{1}{K} \log(\frac{\varepsilon e^{(\underline{\alpha} - c)t}}{c_K(t)})$, i.e., (22). $\qquad \square$

---

[1]The lower bound $\mu(t) \geq |S|$ is loose as $\mu(t)$ increases from $|S|$ to K along t. This is not an issue in the estimate above if $|S| \geq 1$. If $|S| = 0$ then one can assume existence of a pre-activated node (in addition to V) that activates each $i \in V$ at rate β_i since $t = 0$ to mimic the self-activations, and a modified estimate can be applied trivially so we omit the details here.
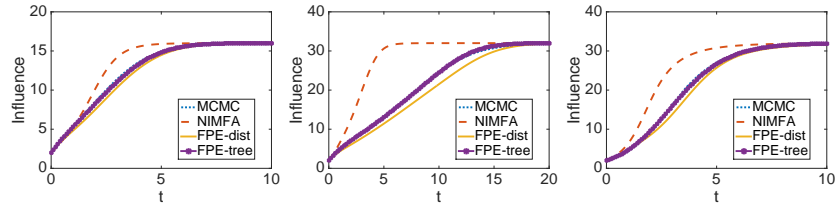
FIGURE 1. Influence prediction on small sized network (when our matlab implementation of `FPE-dist` still takes short time in computing). **Left two**: Erdős-Rényi's network of size $K = 16, 32$. **Right**: Small-world network $K = 32$. Average degree $(1/K) \sum_i |N_i^{\text{out}}| = 4$.

3. **Experimental results.** We first apply the proposed method to networks (with various sizes and parameters) generated by four models commonly used in social/biological/contact networking applications: Erdős-Rényi's random, small-world, scale-free, and Kronecker network[2]. The activation rates $\{\alpha_{ij}\}$ are drawn from interval $(0, 1)$ uniformly to simulate the inhomogeneous propagation rates across edges. Unless otherwise noted, we only consider node-to-node activations in propagations without self-activation and recovery. In all cases except those in Fig. 1, exact solutions are computationally infeasible due to the large size and heterogeneous transmission rates between nodes, we therefore use enough Monte Carlo Markov chain (MCMC) simulated cascades (5000 cascades for each network) to compute the ground truth density $\rho(t)$ and influence $\mu(t)$.

In Fig. 1, we show the performance of our method based on Fokker-Planck equation in Section 2.2 using `FPE-dist` and `FPE-tree`. The NIMFA (N-interwined mean field approximation) is a state-of-the-art method that uses mean-field theory to obtain a system of differential equations to calculate the probability $p_i(t)$ (node i gets activated at time t) [28, 29], and estimates the influence by $\sum_i p_i(t)$. Note that we take a completely different approach to calculate the probability $\rho_k(t)$ for each possible influence size k and estimate the influence by $\sum_k k \rho_k(t)$. For the influence prediction test, we find that our approach appears to be more accurate as shown in Fig. 1, especially `FPE-tree` with $m_k = 3$ for all k which matches ground truth (MCMC) very closely (but at the expense of higher computational cost to estimate transition rates $q_k(t)$). The `FPE-dist` also provides reasonably accurate solution but requires much lower computational cost, hence we only use this version in other tests with large networks. Note that NIMFA requires solving a nonlinear system of K differential equations numerically and hence has the same order of computation complexity as our approach.

In Fig. 2, we show the influence prediction result on networks of much larger size $K = 1024$. Despite of very different network structures, `FPE-dist` provides faithful influence prediction and matches ground truth (MCMC) closely.

Influence prediction problem is considered very challenging computationally, especially for dense networks. In Fig. 3 we test `FPE-dist` on very dense Erdős-Rényi's random networks of size $K = 1024$ where average degrees are $(1/K) \sum_i |N_i^{\text{out}}| = 32$,

---

[2]Code for generating Kronecker network is at https://github.com/snap-stanford/snap/tree/master/examples/krongen and other three using CONTEST package at http://www.mathstat.strath.ac.uk/outreach/contest/toolbox.html
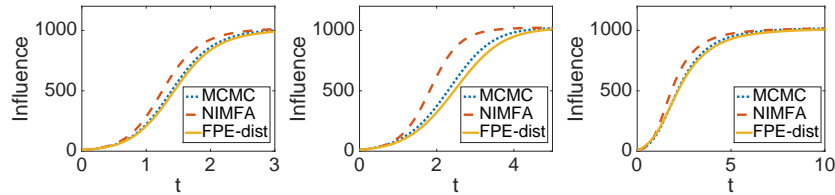
FIGURE 2. Influence prediction on **Left**: Erdős-Rényi's network, **Middle**: small-world network, and **Right**: scale-free network. All have size $K = 1024$ and average degree are $(1/K) \sum_i |N_i^{\text{out}}| = 8, 6, 6$ respectively.

64, and 128 respectively. On all of these networks, `FPE-dist` returns highly accurate prediction of influence which justifies its robustness.

The influence prediction problem considered in this paper, as noted in Section 1.1, is significantly different from those for dynamical processes on networks in statistical physics. Our network is deterministically heterogenous, meaning that $G = (V, E)$ and α_{ij} on all edges are given, and they play critical roles in propagations. Therefore, the identities of nodes in source set S matter greatly (in contrary the nodes in a network are not distinguishable in most statistical physics problems) which leads to many important follow-up questions such as influence maximization (e.g., finding the source set S that solves $\max_{|S| \le k_0} \mu(t; S)$ for some prescribed size $k_0 \in \mathbb{N}$ and time t) [7, 13, 14, 30] and outbreak detection [8, 17]. To see the critical role of the source set S, we apply `FPE-dist` to three different choices of source set S_1, S_2, S_3 all with $|S_i| = 10$ and show the prediction results in the middle panel of Fig. 3. Here S_1 is the choice obtained by the influence maximization function from ConTinEst code [11], S_2 consists of the ten nodes with largest degrees in G, and S_3 contains ten nodes randomly chosen from the network. The plots clearly show different influences of these sources sets S_i's due to the deterministically heterogeneous structure of the network. Nevertheless, `FPE-dist` has very robust performance and matches the ground truths (MCMC) closely in every case.

We also compare `FPE-dist` to the state-of-the-arts learning-based ConTinEst algorithm [11]. The network data and its implementation are obtained from the ConTinEst package published by its authors[3]. ConTinEst is a state-of-the-arts learning-based algorithm that uses parametrized kernel functions to approximate the coverage of each node based on Monte Carlo samplings. The result is shown in the right panel of Fig. 3. From this test, we see that `FPE-dist` is very accurate as it matches the ground truth (MCMC) much better. Moreover, ConTinEst takes excessively long time to estimate influence for denser networks as those in the left panel of Fig. 3, while `FPE-dist` still works robustly without suffering the issue at all. Note that comprehensive comparison of ConTinEst with several other existing methods is reported in [11], from which significant improvement in accuracy of the proposed method `FPE-dist` can be projected.

We established the relation between estimation error in $\{q_k(t)\}$ and the prediction error in $\mu(t)$ in Section 2.3. To check this numerically, we apply `FPE-dist` to a dense Erdős-Rényi's network of size $K = 300$ and average degree $(1/K) \sum_i |N_i^{\text{out}}| = 150$ (α_{ij} again drawn from $(0, 1)$ uniformly) with source set $S = \{1, \ldots, 10\}$, and

---

[3]Data and code available at http://www.cc.gatech.edu/~ndu8/DuSonZhaMan-NIPS-2013.html.

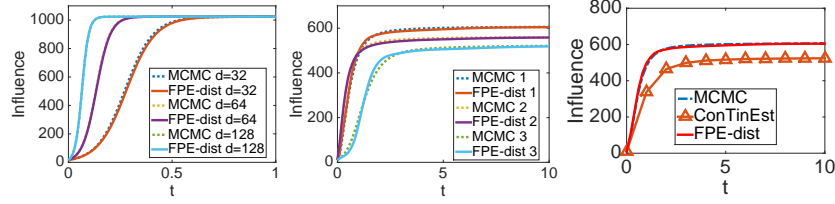FIGURE 3. **Left**: Influence prediction on dense Erdős-Rényi's random network with $K = 1024$ and $(1/K)\sum_i |N_i^{\text{out}}| = 32, 64, 128$ respectively. **Middle**: Influence prediction on the <u>same</u> Kronecker network of size 1024 using three different choices of source set S_1, S_2, S_3 ($|S_i| = 10$ in all three cases). **Right**: Comparison with the state-of-the-arts learning-based ConTinEst method.

check the estimated $q_k(t)$, $\rho_k(t)$ and $\mu(t)$ with those obtained by MCMC simulated cascades. Recall that `FPE-dist` uses very crude estimate of $q_k(t)$ by setting a constant $\hat{q}_k = \alpha(U_k^*)$ where U_k^* contains the k nodes of shortest distance from S in Section 2.2. We first plot the $\hat{q}_k$ for $k = 10, 70, 130, 190$ and compare with $q_k(t)$ given by ground truth (MCMC simulations) in the top row of Fig. 4. To this end, we observe from (4) that $q_k(t) = -(\sum_{j=0}^k \rho_j(t))'/\rho_k(t)$, so we obtain $\rho_k(t)$ from MCMC simulations and apply finite difference to get $\rho_k'(t)$ and hence $q_k(t)$. Note that $q_k(t) = 0$ for most t because $\sum_{j=0}^k \rho'(t)$ or $\rho_k(t)$ vanish there and obtaining these $q_k(t)$ is unstable numerically, so the comparison is only meaningful for t where $\rho_k(t)$ is away from zero. From the top row of Fig. 4, we can see that the estimated $\hat{q}_k$ appear to accurately capture the mean of $q_k(t)$, but can be quite deviated (i.e., with large $|\hat{q}_k(t) - q_k(t)|/q_k(t)$). However, the densities $\hat{\rho}_k(t)$ computed using these $\hat{q}_k$ are still close to the ground truth $\rho_k(t)$, as shown by the small relative error $|\hat{\rho}_k(t) - \rho_k(t)|/\rho_k(t)$ in the bottom leftmost panel of Fig. 4. This also yields a small relative error in influence prediction $|\hat{\mu}(t) - \mu(t)|/\mu(t)$ (second on bottom row), and close match of prediction result $\hat{\mu}(t)$ and ground truth $\mu(t)$ (MCMC) (third on bottom row) in Fig. 4. The small errors in $\hat{\rho}_k(t)$ and $\hat{\mu}(t)$ in our numerical tests suggest that the theoretical bound on the estimation error in $q_k(t)$ in (12) may be further relaxed without degrading solution quality.

To show the great potential of the proposed method for influence prediction on large sized networks, we plot the CPU time (in seconds) for solving the Fokker-Planck equation (4) numerically using MATLAB with single core computation on a regular desktop computer (Intel Core 3.4GHz CPU) in the bottom rightmost panel of Fig. 4. In contrast, most state-of-the-art learning-based approaches suffer drastic increase of computational cost for larger or denser networks due to the significantly amplified number of simulations required to achieve acceptable level of accuracy [11]. On the other hand, the proposed method possesses low computation complexity and is scalable for large and dense networks.

4. **Concluding remarks.** We consider the important influence (expected number of activated nodes) prediction problem on general heterogeneous networks. The problem is significantly different from those in classical mathematical epidemics theory where individual contact network is not considered nor those in statistical physics where networks are statistically homogeneous and nodes are not exactly distinguishable. In our problem, the influence depends on the following factors
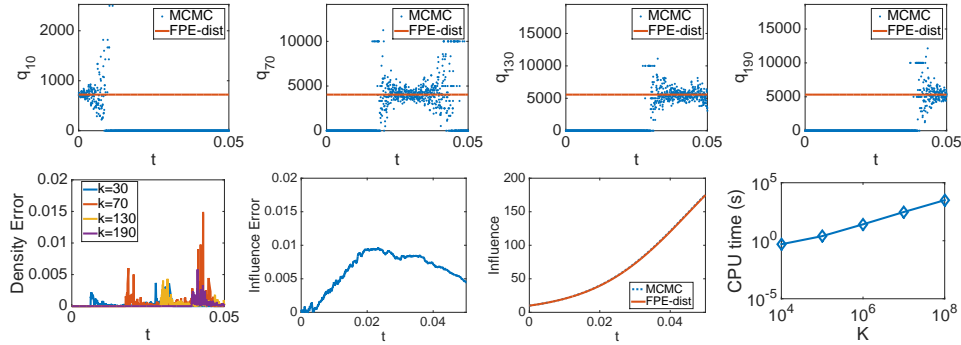
FIGURE 4. **Top row**: $\hat{q}_k$ estimated in `FPE-dist` and $q_k(t)$ shown by ground truth (MCMC simulation) for $k = 10, 70, 130, 190$ using a dense Erdős-Rényi's network of size $K = 300$ and average degree 150. **Bottom row** from left to right: $|\hat{\rho}_k(t) - \rho_k(t)|/\rho_k(t)$; $|\hat{\mu}(t) - \mu(t)|/\mu(t)$ and plot of $\mu(t)$ and $\hat{\mu}(t)$ for this $K = 300$ network; and CPU time (in seconds) of `FPE-dist` using Runge-Kutta 4th order ODE solver on networks with K range from 10^4 to 10^8.

which all play critical roles in computations: the structure of network (directed graph) $G = (V, E)$, the activation rates $\{\alpha_{ij}\}$ between every pair of nodes i and j (and self-activation rates $\{\beta_i\}$ and recovery rates $\{\gamma_i\}$ if applicable), and the source set S. In this paper, we proposed a novel approach by calculating the probability $\rho_k(t)$ (k nodes are activated at time t) for all influence sizes k to obtain influence $\mu(t) = \sum_k k\rho_k(t)$. To this end, we establish the Fokker-Planck equation as a system of deterministic differential equations that governs the dynamical evolution of $\{\rho_k(t)\}$. We provide a few instances for estimating the coefficients in the Fokker-Planck equations, and establish the relation between the coefficient estimation error and the final influence prediction error, which apply to all types of propagation models on general networks. We conducted a number of numerical experiments which justify the very promising performance of the proposed approach in terms of accuracy, efficiency and robustness.

Our novel approach also gives rise to a number of new research problems. For example: How to approximate the transition rates q_k and r_k accurately for general propagation models (e.g., activation time is not exponentially distributed and hence the propagation is not Markov)? How to apply the Fokker-Planck equation approach to influence prediction when only propagation cascade data is available (i.e., only the activation times and identities are observed during a number of propagations but not the actual network $G = (V, E)$ and/or activation parameters in practice)? These problems are important from both of theoretical and practical points of view, and we plan to investigate them in our future research.

Appendix.

Proposition 1. *Let $T_1, T_2, \ldots, T_n$ be independent random variables and $T_j \sim \exp(\alpha_i)$ for all j, then the probability that $T_i = \min_{1 \leq j \leq n} T_j$ is $\alpha_i / (\sum_{j=1}^{n} \alpha_j)$, and the minimum $\min_{1 \leq j \leq n} T_j \sim \exp(\sum_{i=1}^{n} \alpha_i)$.*

Proof. The proof is by direct computation and hence details are omitted here. □

Proposition 2. *Let $T_i \sim \exp(\alpha_i)$ and Y be a multinomial random variable such that $\Pr(Y = i) = p_i$ for $i = 1, \ldots, n$, then the probability density function of T_Y is $f_{T_Y}(t) = \sum_{i=1}^{n} p_i \alpha_i e^{-\alpha_i t}$, and the instantaneous hazard rate of point process associated to time T_Y is $\alpha_{T_Y}(t) = (\sum_{i=1}^{n} p_i \alpha_i e^{-\alpha_i t})/(\sum_{i=1}^{n} p_i e^{-\alpha_i t})$. In particular, $\alpha_{T_Y}(0) = \sum_{i=1}^{n} p_i \alpha_i$.*

Proof. We use the rule of total probability to obtain

$$\Pr(T_Y \geq t) = \sum_{i=1}^{n} \Pr(T_Y \geq t | Y = i) \Pr(Y = i) = \sum_{i=1}^{n} p_i e^{-\alpha_i t}. \qquad (23)$$

Hence the cumulative distribution function of T_Y is $F_{T_Y}(t) = 1 - \Pr(T_Y \geq t)$ and probability density function is $f_{T_Y}(t) = F'_{T_Y}(t) = \sum_{i=1}^{n} p_i \alpha_i e^{-\alpha_i t}$. The instantaneous hazard rate is then given by $\alpha_{T_Y}(t) = f_{T_Y}(t)/\Pr(T_Y \geq t)$. $\qquad \square$

REFERENCES

[1] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang and M. Zanin, The structure and dynamics of multilayer networks, *Physics Reports*, **544** (2014), 1–122.

[2] W. Bock, T. Fattler, I. Rodiah and O. Tse, An analytic method for agent-based modeling of spatially inhomogeneous disease dynamics, in *AIP Conference Proceedings*, vol. 1871, AIP Publishing, 2017, 1–11.

[3] M. Boguná and R. Pastor-Satorras, Epidemic spreading in correlated complex networks, *Physical Review E*, **66** (2002), 047104.

[4] E. Cator and P. Van Mieghem, Second-order mean-field susceptible-infected-susceptible epidemic threshold, *Physical review E*, **85** (2012), 056111.

[5] R. Che, W. Huang, Y. Li and P. Tetali, Convergence to global equilibrium for fokker–planck equations on a graph and talagrand-type inequalities, *Journal of Differential Equations*, **261** (2016), 2552–2583.

[6] S.-N. Chow, W. Huang, Y. Li and H. Zhou, Fokker–planck equations for a free energy functional or markov process on a graph, *Archive for Rational Mechanics and Analysis*, **203** (2012), 969–1008.

[7] E. Cohen, D. Delling, T. Pajor and R. F. Werneck, Timed influence: Computation and maximization, arXiv preprint, arXiv:1410.6976.

[8] P. Cui, S. Jin, L. Yu, F. Wang, W. Zhu and S. Yang, Cascading outbreak prediction in networks: A data-driven approach, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2013, 901–909.

[9] G. Demirel, F. Vazquez, G. Böhme and T. Gross, Moment-closure approximations for discrete adaptive networks, *Physica D: Nonlinear Phenomena*, **267** (2014), 68–80.

[10] N. Du, Y. Liang, M.-F. Balcan and L. Song, Influence function learning in information diffusion networks, in *International Conference on Machine Learning*, 2014, 2016–2024.

[11] N. Du, L. Song, M. Gomez-Rodriguez and H. Zha, Scalable influence estimation in continuous-time diffusion networks, in *Advances in Neural Information Processing Systems*, 2013, 3147–3155.

[12] M. Erbar and J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, *Arch. Ration. Mech. Anal.*, **206** (2012), 997–1038.

[13] M. Gomez Rodriguez, B. Schölkopf, L. J. Pineau et al., Influence maximization in continuous time diffusion networks, in *29th International Conference on Machine Learning (ICML 2012)*, International Machine Learning Society, 2012, 1–8.

[14] D. Kempe, J. Kleinberg and É. Tardos, Maximizing the spread of influence through a social network, *Theory Comput.*, **11** (2015), 105–147.

[15] J. O. Kephart and S. R. White, Directed-graph epidemiological models of computer viruses, in *Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on*, IEEE, 1991, 343–359.

[16] J. Leskovec, L. Backstrom and J. Kleinberg, Meme-tracking and the dynamics of the news cycle, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2009, 497–506.

[17] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen and N. Glance, Cost-effective outbreak detection in networks, in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2007, 420–429.

[18] J. Lindquist, J. Ma, P. Van den Driessche and F. H. Willeboordse, Effective degree network disease models, *Journal of mathematical biology*, **62** (2011), 143–164.

[19] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard and L. J. Dubé, Adaptive networks: Coevolution of disease and topology, *Physical Review E*, **82** (2010), 036116, 10pp.

[20] J. C. Miller and I. Z. Kiss, Epidemic spread in networks: Existing methods and current challenges, *Mathematical Modelling of Natural Phenomena*, **9** (2014), 4–42.

[21] C. Moler and C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, *SIAM review*, **45** (2003), 3–49.

[22] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

[23] R. Pastor-Satorras, C. Castellano, P. Van Mieghem and A. Vespignani, Epidemic processes in complex networks, *Reviews of modern physics*, **87** (2015), 925–979.

[24] H. Ryu, M. Lease and N. Woodward, Finding and exploring memes in social media, in *Proceedings of the 23rd ACM conference on Hypertext and social media*, ACM, 2012, 295–304.

[25] R. B. Sidje, Expokit: A software package for computing matrix exponentials, *ACM Transactions on Mathematical Software (TOMS)*, **24** (1998), 130–156.

[26] M. Sniedovich, Dijkstra's algorithm revisited: the dynamic programming connexion, *Control and cybernetics*, **35** (2006), 599–620.

[27] T. J. Taylor and I. Z. Kiss, Interdependency and hierarchy of exact and approximate epidemic models on networks, *Journal of mathematical biology*, **69** (2014), 183–211.

[28] P. Van Mieghem and J. Omic, In-homogeneous virus spread in networks, arXiv preprint, arXiv:1306.2588.

[29] P. Van Mieghem, J. Omic and R. Kooij, Virus spread in networks, *Networking, IEEE/ACM Transactions on*, **17** (2009), 1–14.

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

[31] Y. Wang, D. Chakrabarti, C. Wang and C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, IEEE, 2003, 25–34.

[32] J. Xue and Q. Ye, Computing exponentials of essentially non-negative matrices entrywise to high relative accuracy, *Mathematics of Computation*, **82** (2013), 1577–1596.

E-mail address: chow@math.gatech.edu

E-mail address: xye@gsu.edu

E-mail address: zha@cc.gatech.edu

E-mail address: hmzhou@math.gatech.edu