Influence Prediction on Networks

Xiaojing Ye

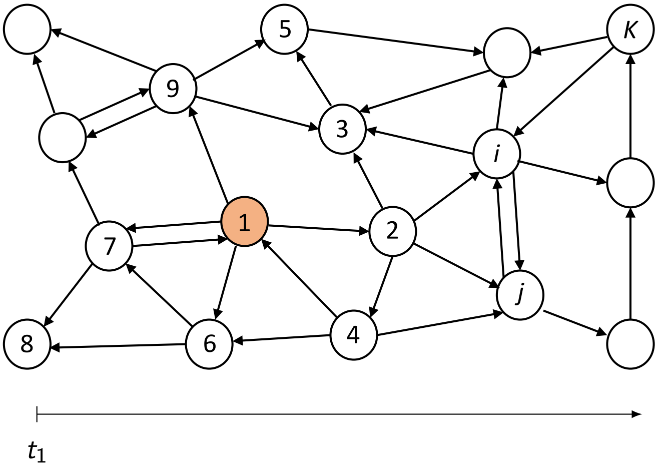
Department of Mathematics and Statistics
Georgia State University, Atlanta, GA

Joint work with
Shui-Nee Chow (GT Math), **Hongyuan Zha** (GT CSE), **Haomin Zhou** (GT Math)

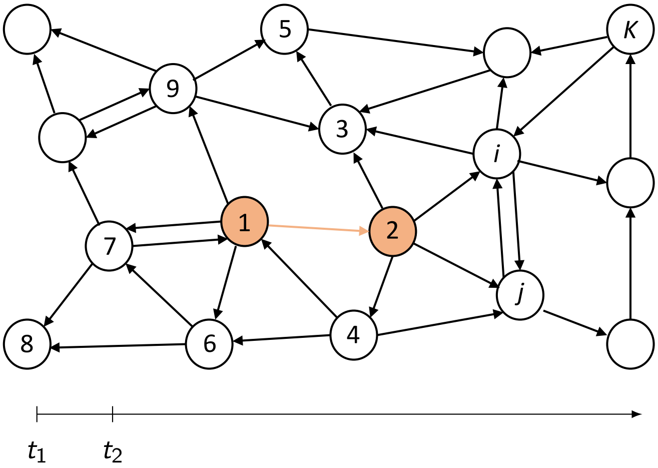
SIAM CSE'17, 02/27/17

Research supported in part by NSF grant DMS-1620342

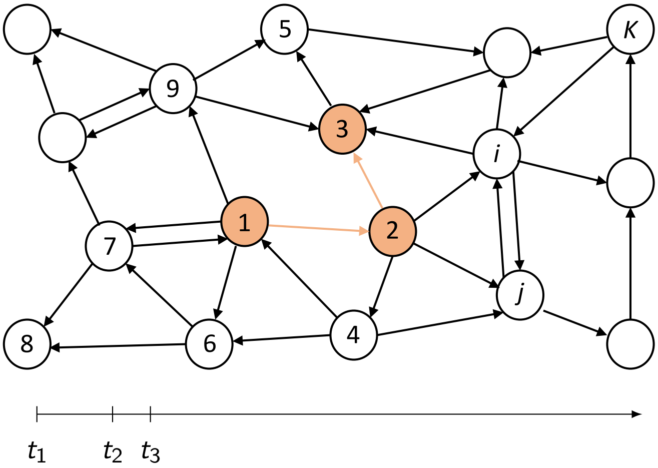
Propagation network: an example



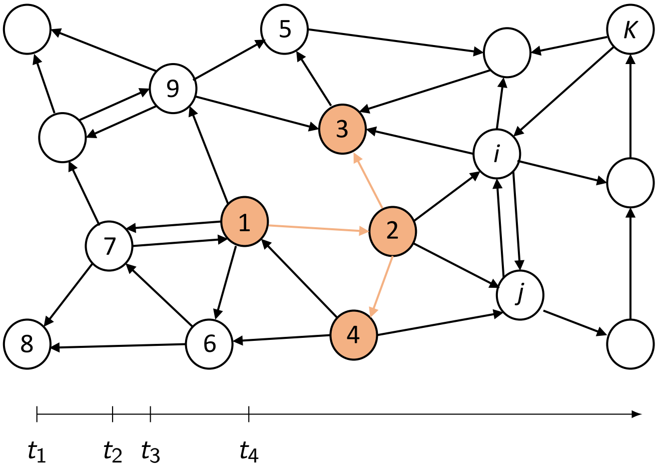
Propagation network: an example



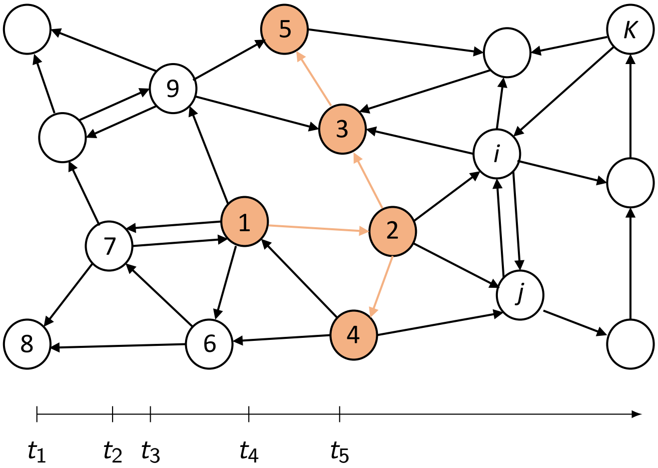
Propagation network: an example



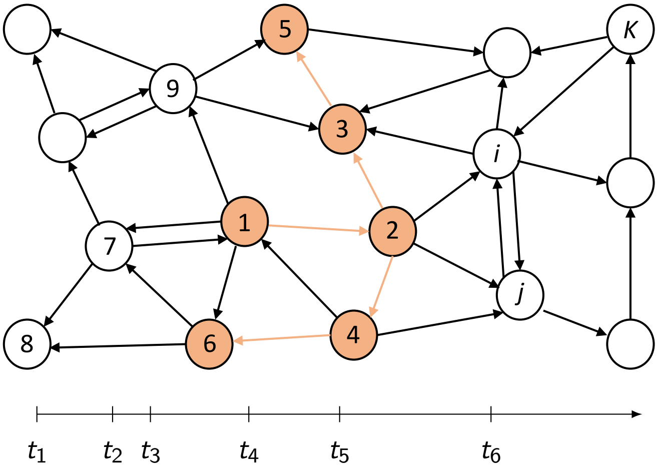
Propagation network: an example



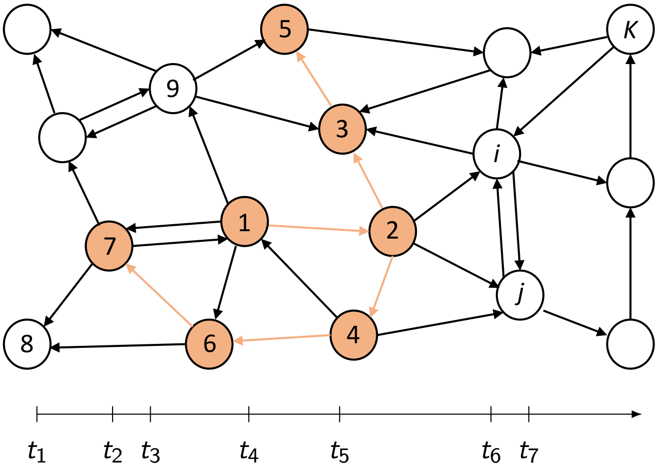
Propagation network: an example



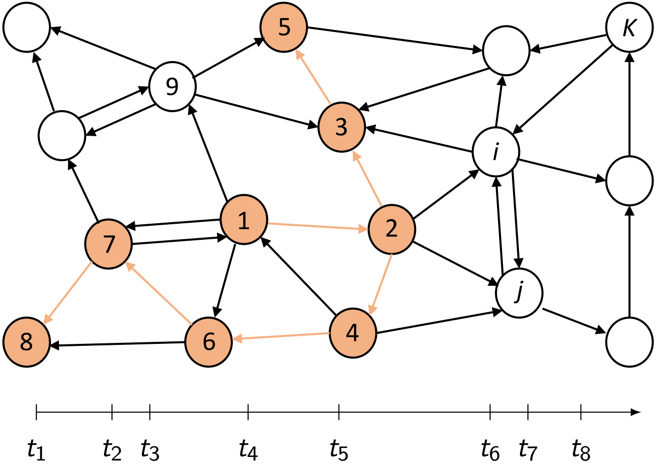
Propagation network: an example



Propagation network: an example



Propagation network: an example



Problem description

Propagation network:

- $G = (V, E)$ network (directed graph)
- $S \subset V$ source set
- $\{\alpha_{ij} : (i, j) \in E\}$: $t_{ij} = t_j - t_i \sim \mathrm{Exp}(\alpha_{ij})$

Then information propagates by gradually activating more nodes.

Definition (Influence)

Given S, the expected number of activated nodes at time t is called the <u>influence</u> of S, denoted by $\mu(t; S)$.

Influence prediction

Question:

Given S, how to compute influence $\mu(t; S)$ for all t?

Influence prediction has many applications

- Influence maximization: fix t and $n \in \mathbb{N}$, solve

$$\underset{S \subset V}{\text{maximize}}\, \mu(t; S) \quad \text{s.t. } |S| \leq n$$

- Outbreak detection
- Propagation control

Exact solution? Not tractable.

Exact solution requires working in a state space of size $O(2^K)$.

$N(t)$ and its transition states

From now on, since S is arbitrary and fixed, we drop it for notation simplicity.

Let $N(t)$ be the (random) number of activated nodes in G, and M_k be the state that $N(t) = k$. Then

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

where $q_k(t)$ is the transition rate from M_k to M_{k+1}, and $r_k(t)$ is the transition rate from M_k to M_{k-1} at time t.

Key quantities

Number of activated nodes:

$$N(t)$$

Probability that $N(t)$ is in state M_k:

$$\rho_k(t) = \Pr(N(t) = k)$$

Influence (i.e., expected number of activated nodes):

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

Note the key is to compute $\{\rho_k(t)\}$!

Fokker-Planck equation

Recall the state transition graph:

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

The Fokker-Planck equation is a system of deterministic differential equations that governs the time evolution of $\rho_k(t)$:

$$\begin{aligned}
\rho_0'(t) &= -q_0(t)\rho_0(t) + r_1(t)\rho_1(t), \\
\rho_k'(t) &= q_{k-1}(t)\rho_{k-1}(t) - [q_k(t) + r_k(t)]\rho_k(t) \\
&\qquad + r_{k+1}(t)\rho_{k+1}(t), \quad \text{for } 1 \le k \le K-1, \\
\rho_K'(t) &= q_{K-1}(t)\rho_{K-1}(t) - r_K(t)\rho_K(t).
\end{aligned}$$

Matrix formulation

The matrix form of the Fokker-Planck equation above is

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

where $\rho(t) = (\rho_0(t), \rho_1(t), \ldots, \rho_K(t)) \in \mathbb{R}^{K+1}$ is a row vector, and $Q(t)$ is a *bidiagonal* matrix:

$$Q(t) = \begin{pmatrix} -q_0(t) & q_0(t) & & & & \\ & -q_1(t) & q_1(t) & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & -q_{K-1}(t) & q_{K-1}(t) \\ & & & & & 0 \end{pmatrix}$$

and $R(t)$ is a (lower) *bidiagonal* matrix with $r_k(t)$'s.

Composition of Q and R

Theorem

Let $\mathcal{S}_k := \{U \subset V : |U| = k\}$ and $\Pr(t; U)$ be the probability that $U \in \mathcal{S}_k$ is activated first. Define

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

Similarly $\beta(U) = \sum_{i \in U} \beta_i$ and $\gamma(U) = \sum_{i \in U} \gamma_i$. Then there are

$$q_k(t) = \sum_{U \in \mathcal{S}_k} [\alpha(U) + \beta(U^c)] \Pr(t; U)$$

$$r_k(t) = \sum_{U \in \mathcal{S}_k} \gamma(U) \Pr(t; U)$$

for $k = 0, 1, \ldots, K$.

Estimate q_k

We assume no self-activation and recovery, and provide two ways to estimate q_k:

- Based on shortest distance (`FPE-dist`):
  Define the distance from i to j by $1/\alpha_{ij}$, let $U_k^* \in \mathcal{S}_k$ pick the k nodes with shortest distance to S, and set

$$\hat{q}_k = \alpha(U_k^*)$$

- Based on overall probability (`FPE-tree`):
  For $k = 1, 2, \ldots,$ recursively find $\{U_k^1, \ldots, U_k^{m_k}\} \subset \mathcal{S}_k$ with large probabilities in $\mathcal{S}_k$, which essentially constructs a tree of nodes $\{U_k^l\}$ with relative probabilities in each layer k. Set

$$\hat{q}_k = \sum_{l=1}^{m_k} \alpha(U_k^l) \, \mathrm{Pr}(U_k^l)$$

Experiment setup

Generating propagation networks:

- Various types of networks (directed graphs): Erdős-Rényi's random, small-world, scale-free, Kronecker, etc.
- Various sizes K and densities (average node out-degree).
- For each edge $(i,j) \in E$, draw $\alpha_{ij} \overset{i.i.d.}{\sim} \mathrm{Unif}(0,1)$.

Ground truth by MCMC:

Obtained by simulating 5000 cascades and calculating average number of activated nodes. (expensive!)

Experimental results: small networks

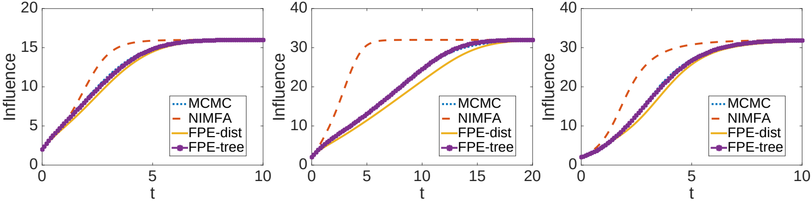


Figure: **Left**: Erdős-Rényi's network ($K = 16$, $d^{\text{avg}} = 4$). **Middle**: Erdős-Rényi's network ($K = 32$, $d^{\text{avg}} = 4$). **Right**: Small-world network ($K = 32$, $d^{\text{avg}} = 4$). Here $d^{\text{avg}} = (1/K) \sum_i |N_i^{\text{out}}|$.

Experimental results: large networks

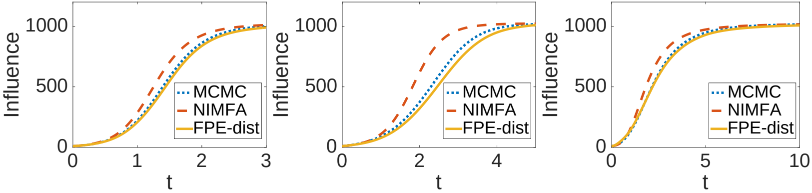


Figure: **Left**: Erdős-Rényi's network ($K = 1024$, $d^{\text{avg}} = 8$). **Middle**: Small-world network ($K = 1024$, $d^{\text{avg}} = 6$). **Right**: Scale-free network ($K = 1024$, $d^{\text{avg}} = 6$).

More experimental results

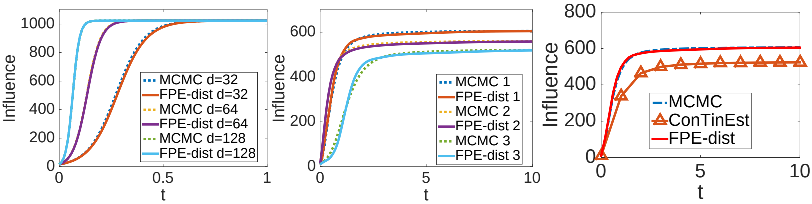


Figure: **Left**: Dense Erdős-Rényi's random network ($K = 1024$ and $d^{\text{avg}} = 32, 64, 128$ respectively). **Middle**: Influence prediction on the <u>same</u> Kronecker network of size 1024 using three different choices of source set S_1, S_2, S_3 ($|S_i| = 10$). **Right**: Comparison with ConTinEst, a state-of-the-art method that learns coverage function using sample cascades.

Why is performance so good?

The estimation of Fokker-Planck equation coefficients q_k seems crude, but why the performance is so good?

We answer this question by building relationship between error in $q_k(t)$ and error in $\mu(t)$.

Error analysis

Lemma

Let $\epsilon \in (0,1)$, and ρ and $\hat{\rho}$ solve $\rho'(t) = \rho(t)Q_{k+1}(t)$ and $\hat{\rho}'(t) = \hat{\rho}(t)Q_k(t)$ respectively, where Q_k has q_j in Q replaced by $\hat{q}_j$ for $j \geq k$. If every $\hat{q}_k$ satisfies

$$\frac{|\hat{q}_k(t) - q_k(t)|}{q_k(t)} \leq \min\left\{\frac{\log(1+\frac{\epsilon}{2})}{\bar{\alpha} kt \min(\bar{d}, K-k)}, \frac{\epsilon}{2+\epsilon}\right\}$$

where $\bar{\alpha} = \max\{\alpha_{ij} : (i,j) \in E\}$, $\bar{d} = \max\{|N_i^{out}| : i \in V\}$, then

$$\hat{\rho}_j(t) = \rho_j(t),\ \textit{for } j = 0, \ldots, k-1$$
$$|\hat{\rho}_j(t) - \rho_j(t)|/\rho_j(t) \leq \epsilon,\ \textit{for } j = k, \ldots, K-1$$
$$|\hat{\mu}(t) - \mu(t)|/\mu(t) \leq \epsilon$$

Error analysis

Theorem

Let $\epsilon \in (0,1)$, and $\rho(t)$ and $\hat{\rho}(t)$ solve $\rho'(t) = \rho(t)Q(t)$ and $\hat{\rho}'(t) = \hat{\rho}(t)\hat{Q}(t)$ respectively, where $\hat{Q}$ has q_k in Q replaced by $\hat{q}_k$ for all k. If every q_k satisifies

$$\frac{|\hat{q}_k(t) - q_k(t)|}{q_k(t)} \leq \min\left\{\frac{\log(1+\frac{\epsilon}{2})}{\bar{\alpha}kt\min(\bar{d}, K-k)}, \frac{\epsilon}{2+\epsilon}\right\}$$

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

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

where $\underline{\alpha} := \min\{\alpha_{ij} : (i,j) \in E\}$.

Error analysis

Corollary

Suppose $\rho(t), \hat{\rho}(t), \mu(t), \hat{\mu}(t)$ are defined and conditions for $\bar{\alpha}$ and $\underline{\alpha}$ as above. 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

$$\begin{aligned}\frac{|\hat{q}_k(t) - q_k(t)|}{q_k(t)} \leq & \frac{\underline{\alpha} - c}{K\bar{q}_k} + \frac{\log \varepsilon - K \log 2 - \log c_K(t)}{K\bar{q}_k t} \\ = & C_k - O(\log t/t)\end{aligned}$$

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

Experimental results

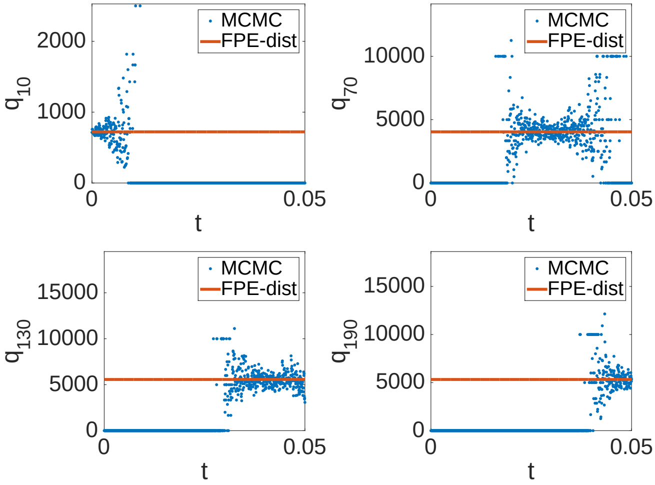


Figure: $\hat{q}_k$ (red) and q_k (blue) for for $k = 10, 70, 130, 190$ in Erdős-Rényi's network ($K = 300$, $d^{\mathrm{avg}} = 150$, $\alpha_{ij} \overset{i.i.d.}{\sim}$ Unif$(0, 1)$).

Experimental results

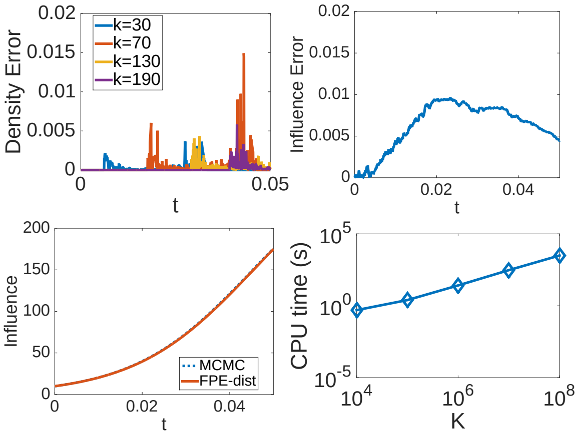


Figure: **Upper left**: $\frac{|\hat{\rho}_k(t)-\rho_k(t)|}{\rho_k(t)}$ for $k = 30, 70, 130, 190$. **Upper right**: $\frac{|\hat{\mu}(t)-\mu(t)|}{\mu(t)}$. **Lower left**: $\hat{\mu}(t)$ and $\mu(t)$. **Lower right**: CPU time (in seconds) to solve Fokker-Planck equation for networks with various sizes.

Summary

In this work, we have

- Built a general framework for influence prediction based on time evolutions of $\rho_k(t)$.
- Provided methods to estimate coefficients of the related Fokker-Planck equations.
- Established relationship between coefficient error and prediction error.

Future work

- Non-Markov propagations.
- Prediction directly based on historical cascade data.
- and more ...

Thank you