NIPS

Neural Information Processing Systems

Reconstructing parameters of spreading models from partial observations

Andrey Y. Lokhov

Los Alamos National Laboratory, Theoretical Division and Center for Nonlinear Science, USA

Los Alamos National Laboratory

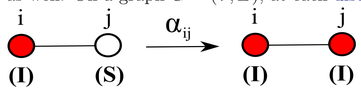
Spreading Processes are Ubiquitous

Epidemic spreading | Infrastructure failures | Financial contagion | Neural cascades | Malware propagation

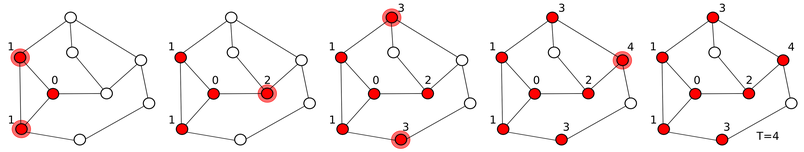
Common feature: unknown models and parameters

Example: Susceptible-Infected Model

The **SI model** is slightly more general than the popular Independent Cascade (IC) model – what follows applies to IC models as well. On a graph $G = (V, E)$, at each discrete time interval $t \to t+1$:

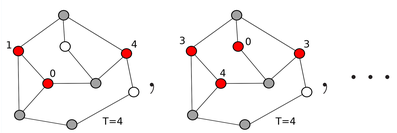


Denote $\{\alpha_{ij}\}_{(ij)\in E} \equiv G_\alpha$. Cascade Σ^c: collection of activation times $\{\tau_i^c\}_{i\in V}$ until time horizon T



Problem formulation

Each cascade is divided into observed ($\mathcal{O}$) and hidden ($\mathcal{H}$) parts, $\Sigma^c = \Sigma^c_\mathcal{O} \cup \Sigma^c_\mathcal{H}$. Examples: $|\mathcal{H}| = H$ hidden nodes not reporting their activation times, or snapshots of the network at a subset of times.



Formulation of the problem:

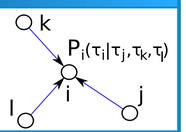
Given M partially observed cascades $\Sigma_\mathcal{O} = \cup_{c=1}^M \Sigma^c_\mathcal{O}$, reconstruct model parameters $\{\alpha^*_{ij}\}_{(ij)\in E} \equiv G_{\alpha^*}$ used to generate data

Special Case: Full Observations

Full observations $\Sigma_\mathcal{O} = \Sigma$: easy

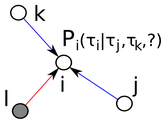
Maximize the likelihood function $P(\Sigma \mid G_\alpha) = \prod_{i\in V} \prod_{1\le c\le M} P_i(\tau_i^c \mid \Sigma^c, G_\alpha)$

$\widehat{G}_{\alpha^*} = \arg\min\left(-\ln P(\Sigma \mid G_\alpha)\right)$: local convex optimization for each $i \in V$



Maximization of the Likelihood: Intractable

Partial observations $\Sigma_\mathcal{O} \neq \Sigma$: hard



Maximization of the likelihood marginalized over unknown information: $P(\Sigma_\mathcal{O} \mid G_\alpha) = \sum_{\{\tau_h^c\},h\in\mathcal{H}} P(\Sigma \mid G_\alpha)$, computational complexity $\propto T^H$

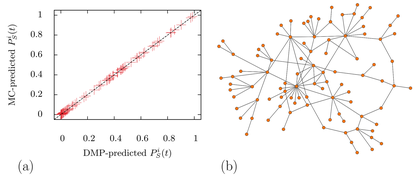
Proposed speed-up: Heuristic Two-Stage Algorithm (HTS)

1. Complete missing $\{\tau_h^c\}_{h\in\mathcal{H}}$ by most probable values $\widehat{\Sigma}_\mathcal{H} = \arg\max P(\Sigma \mid \widehat{G}_\alpha)$ using MC sampling.
2. Solve the "full observations" problem using $\Sigma = \Sigma_\mathcal{O} \cup \widehat{\Sigma}_\mathcal{H}$.
3. Iterate steps 1 and 2 until global convergence of the algorithm. □ (Still very slow.)

Solution: Dynamic Message-Passing Algorithm

Dynamic Message-Passing (DMP) equations allow to approximately solve the dynamics, i.e. compute the marginal probability $m^i(t)$ of activation of node i at time t for each $i \in V$ in time $O(|E|T)$

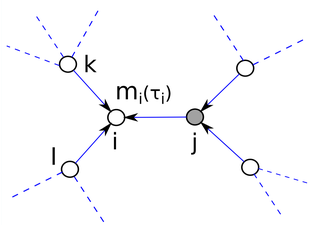
$$m^i(t) = P^i_S(t-1) - P^i_S(t),$$
$$P^i_S(t) = P^i_S(0) \prod_{k\in\partial i} \theta^{k\to i}(t),$$
$$\theta^{k\to i}(t) = \theta^{k\to i}(t-1) - \alpha_{ki}\phi^{k\to i}(t-1),$$
$$\phi^{k\to i}(t) = (1-\alpha_{ki})\phi^{k\to i}(t-1)$$
$$+ P^k_S(0) \prod_{l\in\partial k\setminus i} \theta^{l\to k}(t-1) - P^k_S(0) \prod_{l\in\partial k\setminus i} \theta^{l\to k}(t)$$



Theorem: The quantities $P^i_S(t)$ are exact on tree graphs, and give lower bounds on values of marginal probabilities for general loopy graphs. □ (Accurate in practice.)

Key idea: approximation of the likelihood with marginal probabilities

$$P(\Sigma_\mathcal{O} \mid G_\alpha) \approx \prod_{c=1}^M \prod_{i\in\mathcal{O}} \left[m^i(\tau_i^c \mid G_\alpha)\mathbb{1}_{\tau_i^c \le T} + P^i_S(\tau_i^c \mid G_\alpha)\mathbb{1}_{\tau_i^c = T}\right]$$



Each $m^i(\tau_i^c)$ summarizes the effect of all possible propagation paths. Minimization of the "free energy" $f_{\rm DMP} = -\ln P(\Sigma_\mathcal{O} \mid G_\alpha) = \sum_{i\in\mathcal{O}} f^i_{\rm DMP}$ will yield the most likely consensus among the ensemble of parameters.

Gradient $\partial f^i_{\rm DMP}/\partial\alpha_{rs}$ through the DMP eqs. for derivatives $p^{k\to i}_{rs}(t) \equiv \partial\theta^{k\to i}(t)/\partial\alpha_{rs}$, e.g.

$$\frac{\partial P^i_S(t)}{\partial\alpha_{rs}} = P^i_S(0) \sum_{k\in\partial i} p^{k\to i}_{rs}(t) \prod_{l\in\partial i\setminus k} \theta^{l\to i}(t)$$

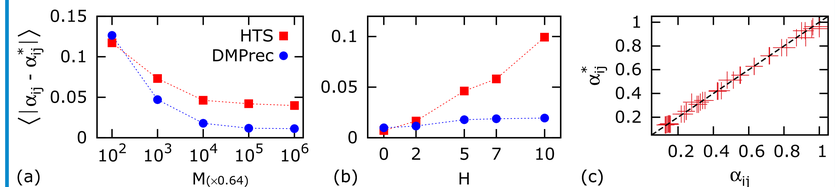
The DMPrec has a much lower complexity per step $O(|E|^2MT)$ vs. $O(|E|^2MNL_{H,T})$ for the HTS.

Consistency of DMPrec on Tree Networks

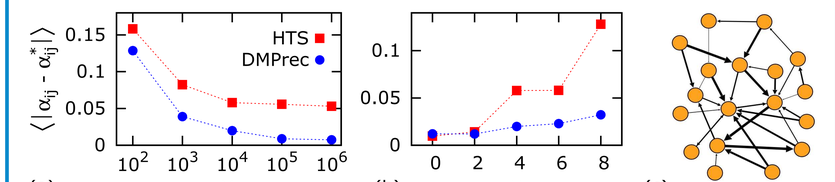
Claim: On tree graphs (regime in which DMP eqs. are derived), $\lim_{M\to\infty} \frac{\partial f_{\rm DMP}}{\partial\alpha_{rs}}|_{G_{\alpha^*}} = 0$. □

Numerical Results

Comparison between HTS and DMPrec:
days → minutes, for even more accurate results

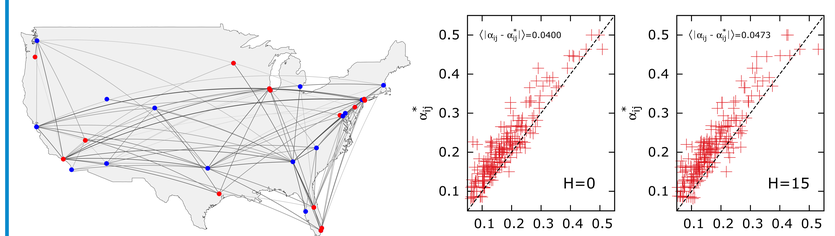


Small power-law network $N = 20$ with random α^*_{ij} in $[0, 1]$ and $T = 10$: (a) $H = 5$, (b) $M = 6400$. Illustration for incomplete observations in time: (c) $M = 6400$, every other time stamp is missing.



Results for the network of relationships in a New England monastery: (b) $H = 4$, (b) $M = 6400$.

Application to real data: air-traffic mediated epidemic spreading



Results for a sub-network of flights ($|E| = 210$) between $N = 30$ major U.S. hubs. $M = 10,000$.