The Block Point Process Model for Continuous-Time Event-Based Dynamic Networks

Ruthwik R. Junuthula, Maysam Haghdan,
Kevin S. Xu, and Vijay K. Devabhaktuni

Statistical Inference for Network Models

June 20, 2017

COLLEGE OF ENGINEERING
THE UNIVERSITY OF TOLEDO

Continuous-Time Event-Based Dynamic Networks

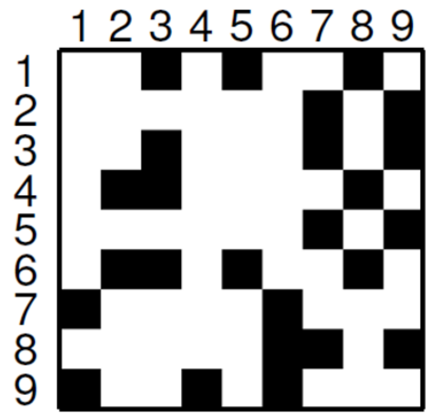
- Relational event data with fine-grained timestamps
  - Facebook wall posts (Viswanath et al., 2009)
- Represent events as triplets (i, j, t)
- Goal: build statistical model for these relations over time

Sender	Receiver	Timestamp
1595	1021	1100626783
4581	5626	1100627183
3806	991	1100640075
521	533	1100714520
521	3368	1100716404
8734	527	1100724840
1017	1015	1100828851
17377	1021	1100832283
2926	726	1100838067

THE UNIVERSITY OF TOLEDO 1872

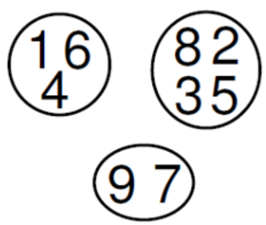
Models for Static Networks

- If we discard timestamps, events become edges (i, j) in a static network
- Represent network by $N \times N$ adjacency matrix A



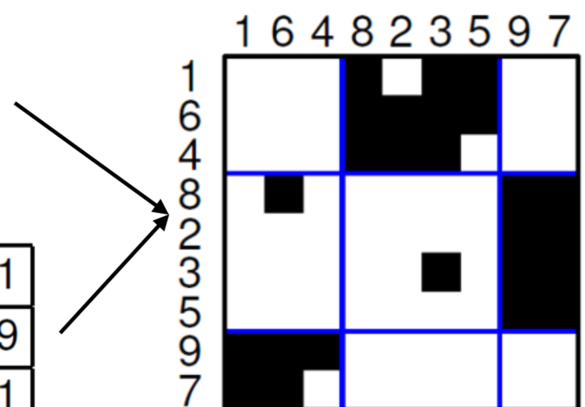
- Stochastic block model (SBM):

Latent classes



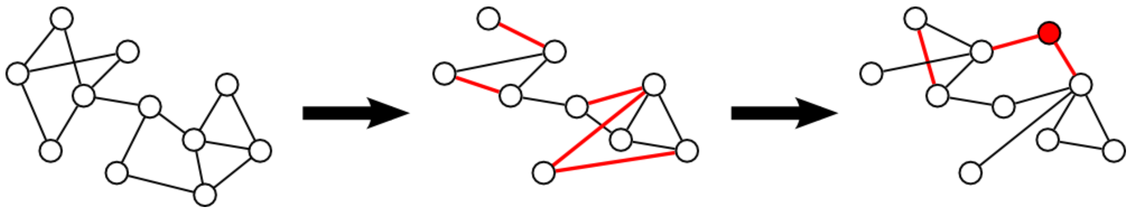
Edge probabilities between classes

0.1	0.9	0.1
0.1	0.1	0.9
0.9	0.1	0.1

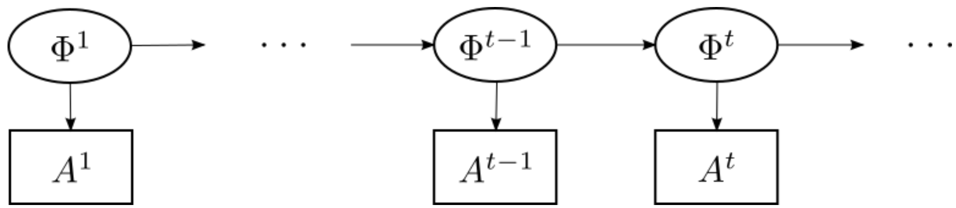


Models for Discrete-Time Dynamic Networks

- If we aggregate events over time windows, we get a discrete-time snapshot-based network representation



- Discrete-time SBMs (Yang et al., 2011; Xu and Hero, 2014; Xu, 2015; Matias and Miele, 2016)



- Trade-offs in choosing snapshot length
  - Too long: loses temporal resolution
  - Too short: increases number of snapshots and causes model to forget too quickly due to short-term memory

THE UNIVERSITY OF TOLEDO 1872

The Block Point Process Model (BPPM)

- Our approach: Model event triplets (i, j, t) directly using SBM-like generative structure
  - Divide nodes into K classes forming $p = K^2$ blocks (assuming directed events)
  - Generate times of events in each block using a point process model
  - Randomly associate event with a pair of nodes (i, j) in the block (thinning)
  - We use an exponential Hawkes process model in practice

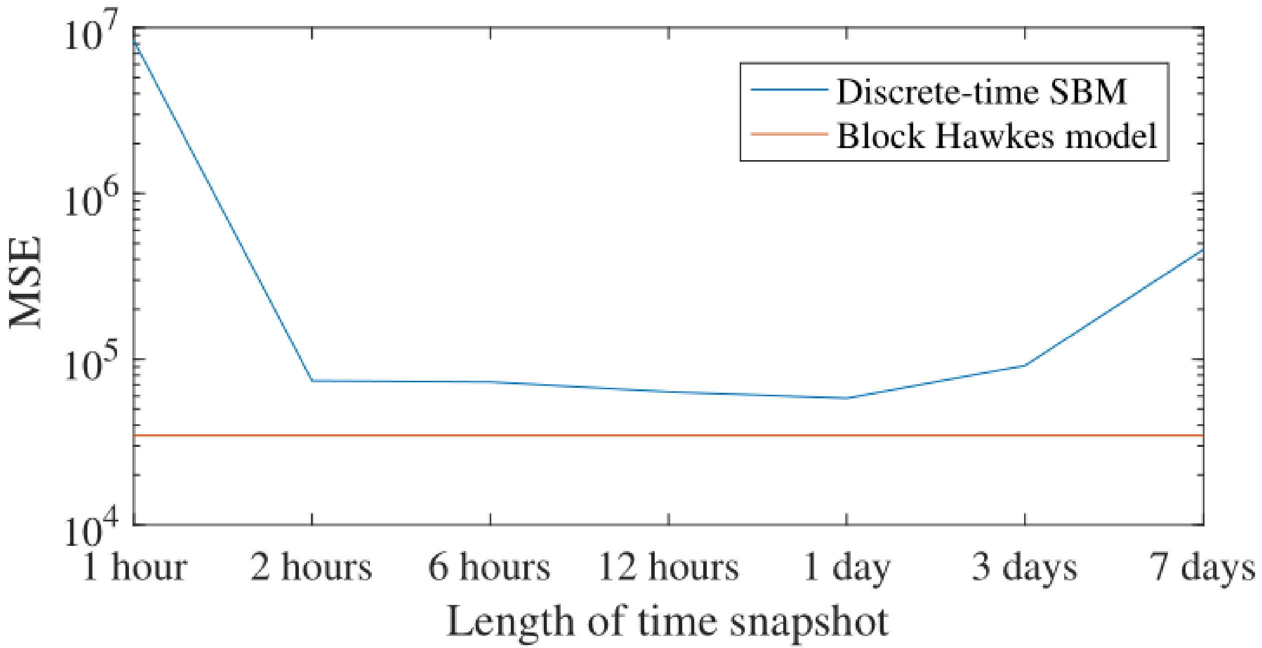
THE UNIVERSITY OF TOLEDO 1872

Our Contributions

- We prove that static networks resulting from the BPPM follow an SBM as $N \rightarrow \infty$
  - We provide an upper bound on the deviation from independence for finite N
- We develop a principled inference procedure for the BPPM using local search initialized by spectral clustering
  - Scales to 5,000+ nodes and 100,000+ events
- We demonstrate that the BPPM is superior to discrete-time network models regardless of snapshot length

Comparison with Discrete-Time SBM

- Prediction task: Attempt to predict time to next event (Facebook wall post) in each block
  - 3,586 nodes and 137,170 events in data set



Relationship to SBM

- Identical distribution of adjacency matrix entries within block satisfied by BPPM generative procedure
- But independence of entries is not satisfied!
  - Denote deviation from independence by

$$\delta_0 = \Pr(a_{ij} = 0 | a_{i'j'} = 0) - \Pr(a_{ij} = 0)$$
$$\delta_1 = \Pr(a_{ij} = 0 | a_{i'j'} = 1) - \Pr(a_{ij} = 0)$$

Theorem (Asymptotic Independence Theorem)**.** *Consider an adjacency matrix A constructed from the BPPM over some time interval $[t_1, t_2)$. Then, for any two entries a_{ij} and $a_{i'j'}$ both in block b, the deviation from independence given by δ_0, δ_1 defined in (1) is bounded in the following manner:*

$$|\delta_0|, |\delta_1| \leq \min\{1, \mu_b/n_b\}$$

where μ_b denotes the expected number of events in block b in $[t_1, t_2)$, and n_b denotes the size of block b. In the limit as the block size $n_b \to \infty$, $\delta_0, \delta_1 \to 0$ provided μ_b is fixed or growing at a slower rate than n_b. Thus a_{ij} and $a_{i'j'}$ are asymptotically independent in the block size n_b.