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

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

Joint Statistical Meetings 2017

July 31, 2017

COLLEGE OF ENGINEERING
THE UNIVERSITY OF TOLEDO

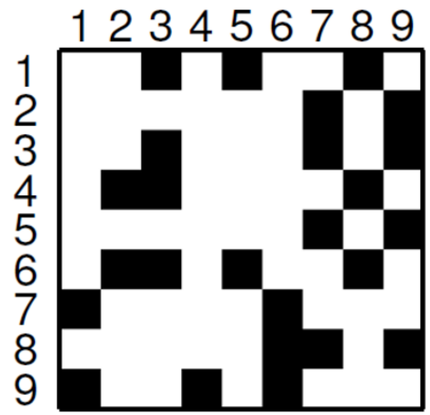
Continuous-time Event-Based Dynamic Networks

- 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

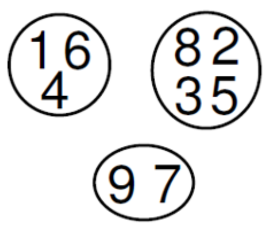
Models for Static Networks

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



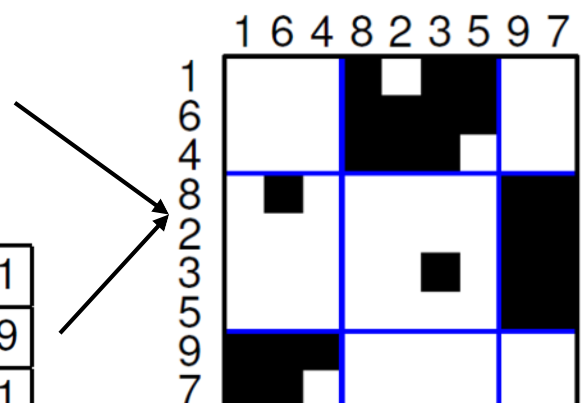
- Stochastic block model (SBM):

Latent classes



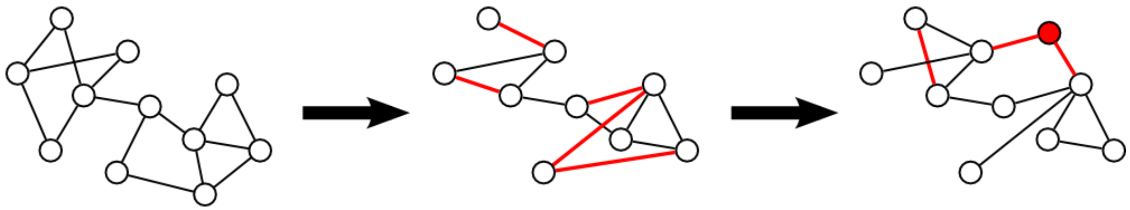
Edge probabilities between classes

0.1	0.9	0.1
0.1	0.1	0.9
0.9	0.1	0.1

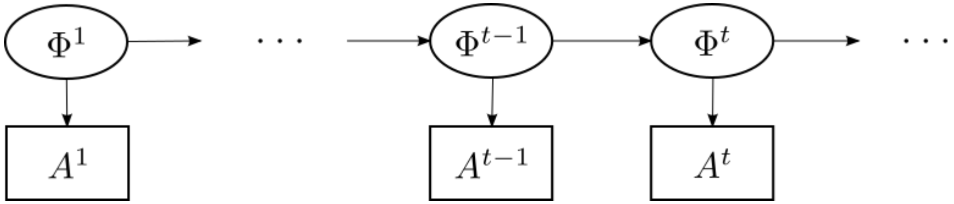


Models for Discrete-Time Dynamic Networks

- 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 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 exponential Hawkes processes in practice

THE UNIVERSITY OF TOLEDO 1872

Related Continuous-time Network Models

- Estimating structure of latent (indirectly observed) network from observations at nodes
  - Example: Estimating diffusion from information cascade (Farajtabar et al., 2015 and many others)
- Point process models for event-based networks
  - Hawkes IRM (Blundell et al., 2012)
  - Relational event model (DuBois et al., 2013)

  } MCMC
  - Individual Hawkes process for each node (Fox et al., 2016) Maximize expected complete-data log-likelihood
  - Inference for these models scales only to ~100 nodes

THE UNIVERSITY OF TOLEDO 1872

Adjacency matrix representation

- Construct adjacency matrix $A = A^{[t_1,t_2)}$ from event matrix E
  - $a_{ij} = 1$ if at least 1 event from i to j in $[t_1, t_2)$
- If E follows a BPPM, does the adjacency matrix follow an SBM?
  - If so, can we use inference techniques for SBM to fit BPPM?

Sender	Receiver	Timestamp
1	2	0.1
2	3	0.4
3	2	0.6
1	2	1.2
1	3	1.3
2	1	1.6

$$A^{[0,1)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$A^{[1,2)} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

THE UNIVERSITY OF TOLEDO 1872

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:*

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

$O(N^2)$ for fixed K

||

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.

Implications of Asymptotic Independence

- First result linking point process network models with static network models
- Main implication: for large networks that are not too dense, class estimation methods that work for SBM should also work for BPPM
  - Maximum likelihood estimation and spectral clustering are both consistent as $N \to \infty$ for polylog expected degree (Bickel et al., 2013; Lei and Rinaldo, 2015)
    - Results in $\frac{\mu_b}{n_b} = O\left(\frac{N \operatorname{poly}(\log N)}{N^2}\right) \to 0$ as $N \to \infty$ for fixed K so also satisfies Asymptotic Independence Theorem

THE UNIVERSITY OF TOLEDO 1872

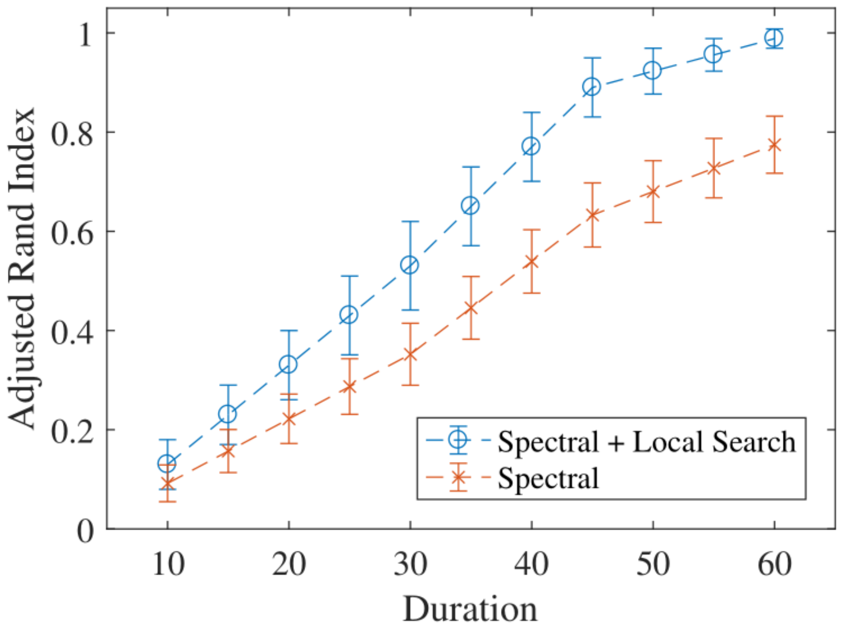
Proposed Inference Procedure

- Maximize log-likelihood of BPPM over class assignments of nodes
  - NP-hard problem: must approximate MLE!
  - We use local search (hill climbing) algorithm to reach a local maximum in a greedy manner
    - Swap each node to each different class 1 at a time and compute new log-likelihood
    - Keep best swap and keep running iterations until no better swap is available
  - Initialize local search using spectral clustering to avoid poor local maxima
- Scales to networks with thousands of nodes and hundreds of thousands of events!

THE UNIVERSITY OF TOLEDO 1872

Class Estimation Accuracy

- Simulate networks with 128 nodes and 4 classes from block Hawkes model
  - $\alpha = 0.6, \beta = 0.8$ for all blocks
  - $\lambda^{\infty} = 1.8$ for diagonal blocks
  - $\lambda^{\infty} = 0.6$ for off-diagonal blocks
- Local search greatly improves estimation accuracy compared to spectral clustering on the adjacency matrix

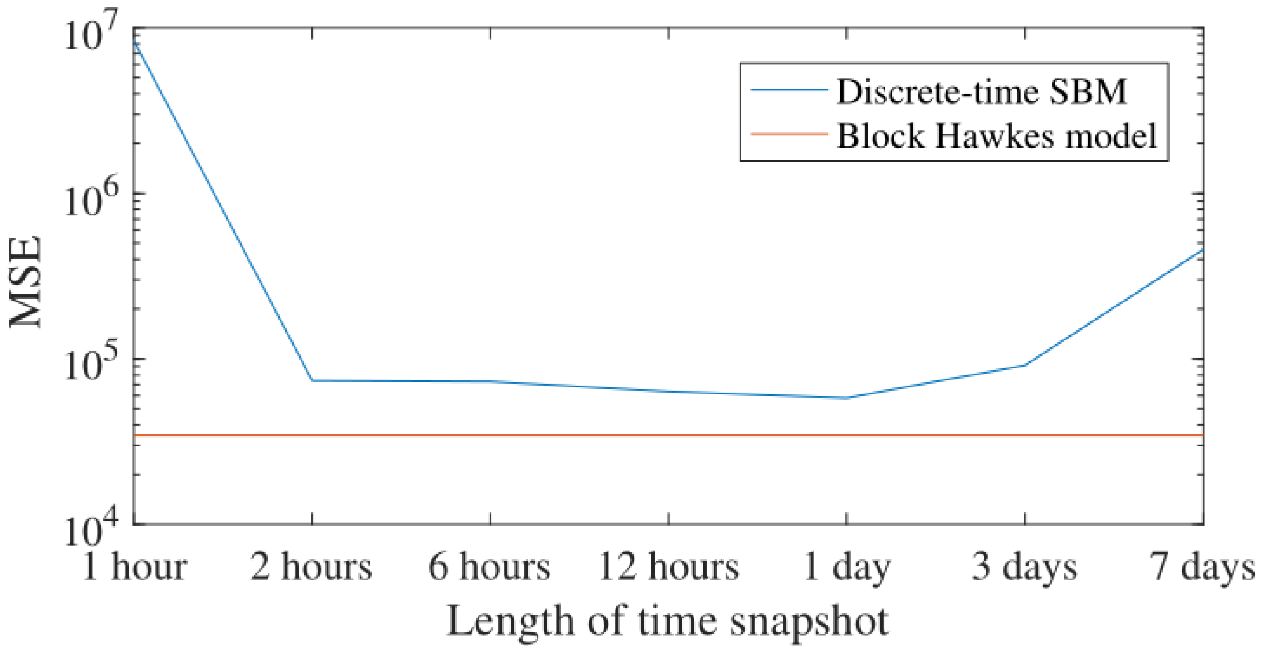


Comparison with Discrete-Time SBM

- How does block Hawkes model compare with discrete-time SBMs (Xu and Hero, 2014)?
  - How sensitive are discrete-time SBMs to snapshot length?
- Prediction task: attempt to predict time to occurrence of next event
  - Use subset of Facebook wall post data (Viswanath et al. 2009) with 3,586 nodes and 137,170 events
  - Split events into 6 folds of equal length
    - At each fold $t \geq 2$, train all models on all folds up to $t - 1$ and attempt to predict time to next event in each block

THE UNIVERSITY OF TOLEDO 1872

Comparison of Prediction Error



- Block Hawkes model has lower prediction MSE than discrete-time SBM regardless of snapshot length
- We use spectral clustering (without local search) for class estimation in both models to make a fair comparison on dynamics

Summary of Contributions

- We prove that static networks resulting from the BPPM follow an SBM as number of nodes $N \to \infty$
  - We provide an upper bound on the deviation from independence for finite N
- We develop a principled inference procedure for the BPPM
  - Local search initialized by spectral clustering on the static network adjacency matrix
  - Scales to thousands of nodes
- We demonstrate that the BPPM is superior to discrete-time network models regardless of snapshot length

THE UNIVERSITY OF TOLEDO 1872

Future Work

- Theoretical
  - Translate conditions required conditions on SBM parameters for theoretical guarantees to required conditions on point process parameters
  - Relationships between discrete-time and continuous-time dynamic network models
- Computational
  - Scaling to extremely large networks (tens or hundreds of thousands of nodes) using stochastic inference and GPU computation
  - Joint inference with discrete-time and continuous-time network data

THE UNIVERSITY OF TOLEDO 1872