Discovering Implicit Networks from Point Process Data

Ryan P. Adams
School of Engineering and Applied Sciences
Harvard University

Joint work with Scott Linderman

http://hips.seas.harvard.edu
**Discovering Implicit Networks from Point Process Data**

**Harvard University, School of Engineering and Applied Sciences, Cambridge, MA, 02138**

**Approved for public release; distribution unlimited**

**Third Biannual Duke University Workshop on Sensing and Analysis of High-Dimensional Data, July 23-25, 2013, Durham, NC. U.S. Government or Federal Rights License**

**Same as Report (SAR)**

**59**
SOCIAL NETWORK ANALYSIS

(a) Adjacency A
(b) Distance D
(c) Transition count M

(d) (e) (f)

\( F = 0.68 \)
\( F = 0.49 \)
\( F = 0.96 \)

Szell et al, Nature 2012
http://mippi.ornl.gov/areas/bioinfo.shtml
Typically, we see edges and reason about the latent properties of the vertices.
What if we don’t observe edges, but only noisy emissions from each vertex?
Ensemble raster

Truccolo, Hochberg & Donoghue, 2010
Patterns of Gang Violence

Murder and Battery in Chicago (2001–2012)

- Battery
- Murder

Saturday, August 3, 13
Overview

- Mutually-Exciting Point Processes
- Aldous-Hoover Graph Priors
- MCMC Inference with Data Augmentation
- Application Examples
- Extending for Neural Models with Inhibition
The point process is a foundational statistical object.

- Gives us random subsets of a larger space.
- Many data are well modeled as point processes:
  - Seismology
  - Epidemiology
  - Economics
- Modeling dependence is challenging - “beyond Poisson”
  - Strauss and Gibbs Processes
  - Determinantal and Permanental Point Processes
  - TODAY: Mutually-Exciting (Hawkes) Processes
A point process on $\mathcal{X}$ gives us random subsets $\{x_n\}_{n=1}^N$.

Formally, a random locally-finite counting measure.

Most of the time we think of them as giving us finite subsets of compact subset of $\mathbb{R}^d$, e.g., time or space.
- The Poisson process is the most basic point process.
- Disjoint regions are independent.
- The number of points in a region is determined by integrating the rate function over that region.
In the Poisson process, everything is conditionally independent given the rate function.

How can we get spike-driven dynamics?
Hawkes Process Dynamics

I

II

III

IV

time

Saturday, August 3, 13
HAWKES PROCESS DYNAMICS

I

II

III

IV

time

Saturday, August 3, 13
Hawkes Process Dynamics

I

II

III

IV

time

Saturday, August 3, 13
HAWKES PROCESS DYNAMICS

I

II

III

IV

time
Hawkes Process Dynamics

I

II

III

IV

time

Saturday, August 3, 13
Hawkes Process Dynamics

I

II

III

IV

time

Saturday, August 3, 13
Hawkes Process Dynamics

I

II

III

IV

Time
The Hawkes process specifies conditional Poisson dynamics in causal cascades.

- Hawkes - Stochastic Point Processes (1972)

- Purely excitatory: each spike increases the intensity according to a weighted temporal kernel.
- Self-excitation: increase your own rate.
- Mutual-excitation: increase other rates.
- Spectral conditions ensure stability.
GRAPH STRUCTURE FROM HAWKES DYNAMICS
The Hawkes process provides a likelihood to connect hypotheses about an unobserved graph to observed event data.

With a Bayesian model, we can manipulate the posterior distribution over graphs, marginalizing out nuisance parameters.

We can infer the temporal kernels that modulate the interaction.

We can perform model comparison between different random graph models and their properties.
Recent work on exchangeable random graphs provides a rich set of priors for underlying networks, e.g., Diaconis & Janson (2007), Orbanz & Roy (2013).

Aldous-Hoover unifies many existing graph models:

- Erdős-Renyi
- Stochastic Block Model
- Latent Distance Model
Hawkes Process Formalism

- $K$ nodes with events in $[0, T]$.
- $N$ ordered event times $s_n \in [0, T]$.
- Node of event $n$ given by $c_n \in \{1, 2, \cdots, K\}$.
- Base rates $\lambda_k^0(t)$
- Kernel $g_\theta(t)$, such that $g_\theta(t < 0) = 0$ and $\int_0^\infty g_\theta(t) \, dt = 1$.

- Binary adjacency matrix $A \in \{0, 1\}^{K \times K}$
- Interaction weight matrix $W \in \mathbb{R}_+^{K \times K}$

- An event on $k$ induces $W_{k,k'} A_{k,k'}$ expected events on $k'$. 
Hawkes Process with Data Augmentation

- Instantaneous Poisson rate:

\[ \lambda_k(t) = \lambda_k^0(t) + \sum_{n=1}^{N} \mathbb{I}(s_n < t) A_{c_n,k} W_{c_n,k} g_\theta(t - s_n) \]

- The superposition property of Poisson processes means that each event is explained by either the background rate or exactly one previous event.

- Use \( Z \in \{0, 1\}^{N \times N} \) to represent these latent variables, where \( Z_{n,n'} = 1 \) means that event \( n \) induced event \( n' \).
\[ p\left( \{ s_n, c_n \}_{n=1}^N, Z \mid A, W, \{ \lambda_k^0(t) \}_{k=1}^K, \theta \right) = \]
\[ \prod_{k=1}^K \exp \left\{ - \int_0^T \lambda_k^0(\tau) \, d\tau \right\} \prod_{n=1}^N \lambda_k^0(s_n) I(c_n = k) I(1 - \sum_{n'} Z_{n', n}) \]
\[ \times \prod_{k'=1}^K \left\{ - \int_{s_n}^T A_{c_n, k'} W_{c_n, k'} g_\theta(\tau - s_n) \, d\tau \right\} \]
\[ \times \prod_{n'=1}^N \left( A_{c_n, k'} W_{c_n, k'} g_\theta(s_{n'} - s_n) \right)^{Z_{n, n'}} \]
\[
p(\{s_n, c_n\}_{n=1}^N, Z \mid A, W, \{\lambda_k^0(t)\}_{k=1}^K, \theta) =
\prod_{k=1}^K \exp \left\{ - \int_0^T \lambda_k^0(\tau) \, d\tau \right\} \prod_{n=1}^N \lambda_k^0(s_n) \mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n})
\times \prod_{k'=1}^K \left\{ - \int_{s_n}^T A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n) \, d\tau \right\}
\times \prod_{n'=1}^N (A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n))^{Z_{n,n'}}
\]

Poisson process likelihood for events from the background process.
DATA-AUGMENTED LIKELIHOOD

\begin{equation}
\begin{aligned}
p(\{s_n, c_n\}_{n=1}^{N}, Z \mid A, W, \{\lambda_k^0(t)\}_{k=1}^{K}, \theta) &= \\
\prod_{k=1}^{K} \exp \left\{ - \int_0^T \lambda_k^0(\tau) \, d\tau \right\} \prod_{n=1}^{N} \lambda_k^0(s_n) \mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n}) \\
\times \prod_{k'=1}^{K} \left\{ - \int_{s_n}^{T} A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n) \, d\tau \right\} \\
\times \prod_{n'=1}^{N} (A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n))^{Z_{n,n'}}
\end{aligned}
\end{equation}

Poisson process likelihood for events induced by previous events.
\[
p(\{s_n, c_n\}_{n=1}^{N}, Z \mid A, W, \{\lambda_k^0(t)\}_{k=1}^{K}, \theta) =
\prod_{k=1}^{K} \exp \left\{ - \int_0^T \lambda_k^0(\tau) \, d\tau \right\} \prod_{n=1}^{N} \lambda_k^0(s_n) \mathbb{I}(c_n=k) \mathbb{I}(1 - \sum_{n'} Z_{n',n})
\times \prod_{k'=1}^{K} \left\{ - \int_{s_n}^{T} A_{c_n,k'} W_{c_n,k'} g_{\theta}(\tau - s_n) \, d\tau \right\}
\times \prod_{n'=1}^{N} (A_{c_n,k'} W_{c_n,k'} g_{\theta}(s_{n'} - s_n))^{Z_{n,n'}}
\]

Ugly looking, but just normalization constants.
- Logistic-normal impulse response
  - Reasonably flexible, with compact support.
- Gaussian process for log background rate
  - Smoothly-varying or periodic external effects.
- Conjugate gamma priors for weights
  - Can also be coupled via, e.g., latent block identity.
\textbf{Inference with MCMC}

- Graph structure: collapsed block Gibbs
- Edge weights: Gibbs (conjugate gamma posterior)
- Latent parent explanations: parallel Gibbs
- Background rates: elliptical slice sampling
- Temporal kernels: Gibbs or slice sampling

- Many transitions can be efficiently simulated on a GPU, enabling models with hundreds of nodes and millions of events.
FINANCIAL UPTICKS / DOWNTICKS

10 stocks, finance and tech sectors, intraday price moves over one week in Sept 2009.

Weighted Edges

Block Membership

Inferred Block Structure (SBM+GP)
HOMICIDES AND ARMED BATTERIES IN CHICAGO

Murder and Battery in Chicago (2001–2012)

• Battery
× Murder

Saturday, August 3, 13
Gang territories identified by CPD
Gang Affiliations as Identified by Chicago PD

Folk

People
Homicide rates in Chicago gang territories (2001-9/2012)
We also infer the event sources, using a spatial model and a Dirichlet process prior.
Trying to Predict Hotspots

Predicted Homicide Rate, Memorial Day 2012

Homicide
Battery

Latitude
Longitude

42.1
42.05
42
41.95
41.9
41.85
41.8
41.75
41.7
41.65
41.6
41.5

-87.9
-87.85
-87.8
-87.75
-87.7
-87.65
-87.6
-87.55
-87.5

Saturday, August 3, 13
Neural functional connectivity involves both excitation and inhibition, so the pure Hawkes is inappropriate.

We extend the model to allow for negative weights and include a saturating nonlinearity.

This effective becomes a Bayesian variant of the popular generalized linear model (GLM) from the computational neuroscience literature.

We can leverage graph priors within this framework to discover latent neural properties and connectivity.
Bayesian GLM for Neural Data

"Background" Rate

\[ x^0(t) \]

Saturating Nonlinearities

\[ x(t) \rightarrow \lambda(t) \]

Poisson Processes

\[ \{s_n, c_n\}_{n=1}^N \]

Temporal Impulse Response \( g_\theta(\Delta t) \)

Weighted Network \( W \cdot A \)

(Scalar)
27 neurons, 50K spikes, macaque retina
(Data from Pillow and Chichilnisky)
INFERRED NETWORK PROPERTIES

RGC Connectivity

Presynaptic Neuron
- OFF
- ON

Postsynaptic Neuron
- OFF
- ON

Outgoing Weights (OFF)
- Pr(W)
- W

Outgoing Weights (ON)
- Pr(W)
- W

Pr(A) vs Distance
- P(A_{k,k'})
- \|x_k - x_{k'}\|^2
SUMMARY

‣ We cannot directly observed edges for many networks of interest.

‣ However, these latent graphs can be inferred from vertex emissions.

‣ The purely-excitatory case (Hawkes process) enables an elegant data-augmentation approach to inference.

‣ MCMC is fast and tractable, and lets us reason about different graphs and graph priors.

‣ The inhibitory case is less tractable, but important for neuroscience applications.