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

HARVARD
INTELLIGENT
PROBABILISTIC
SYSTEMS

VE RI TAS

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **03 AUG 2013**	2. REPORT TYPE	3. DATES COVERED **00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE **Discovering Implicit Networks from Point Process Data**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Harvard University ,School of Engineering and Applied Sciences,Cambridge,MA,02138**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

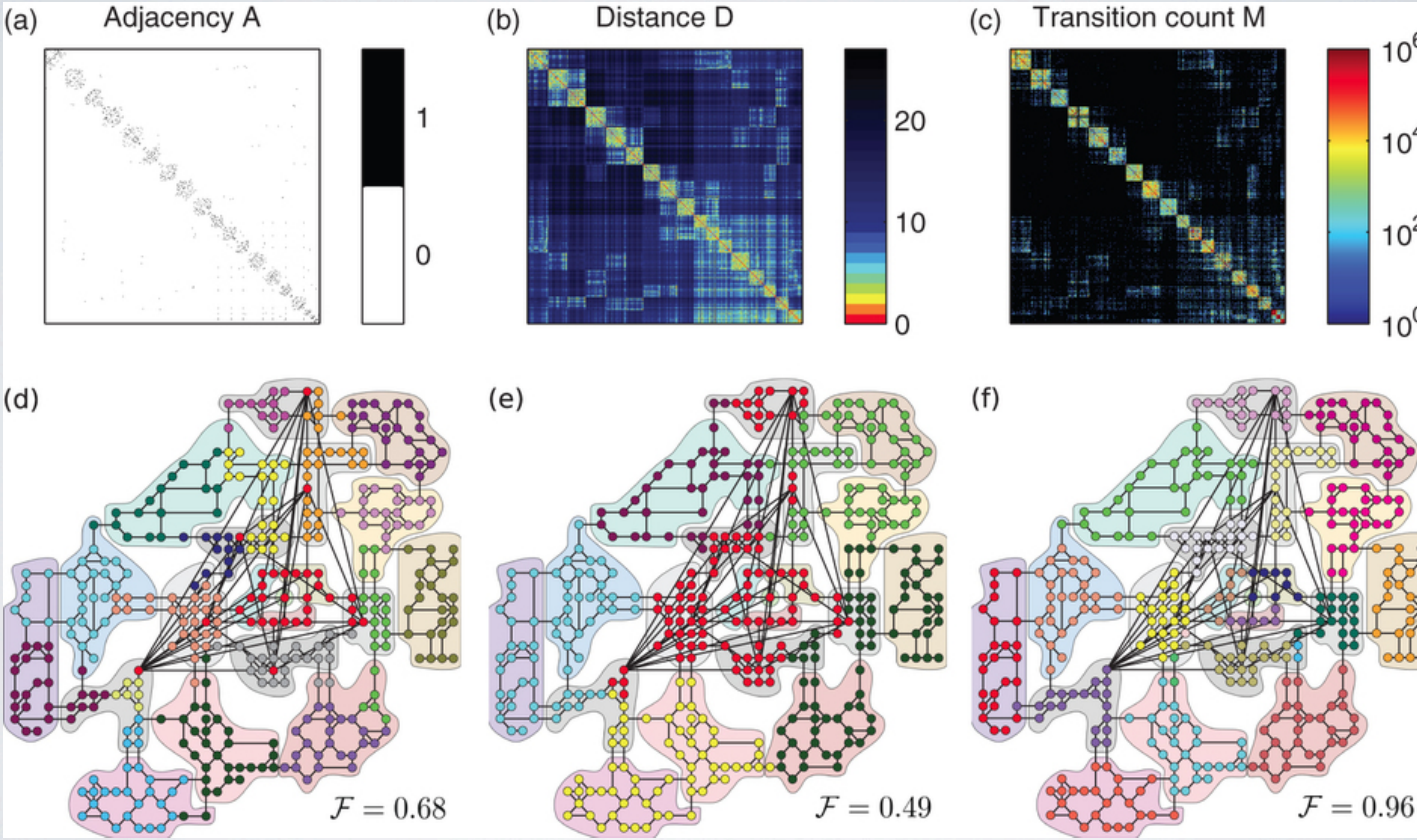
13. SUPPLEMENTARY NOTES
Third Biannual Duke University Workshop on Sensing and Analysis of High-Dimensional Data, July 23-25, 2013, Durham, NC. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

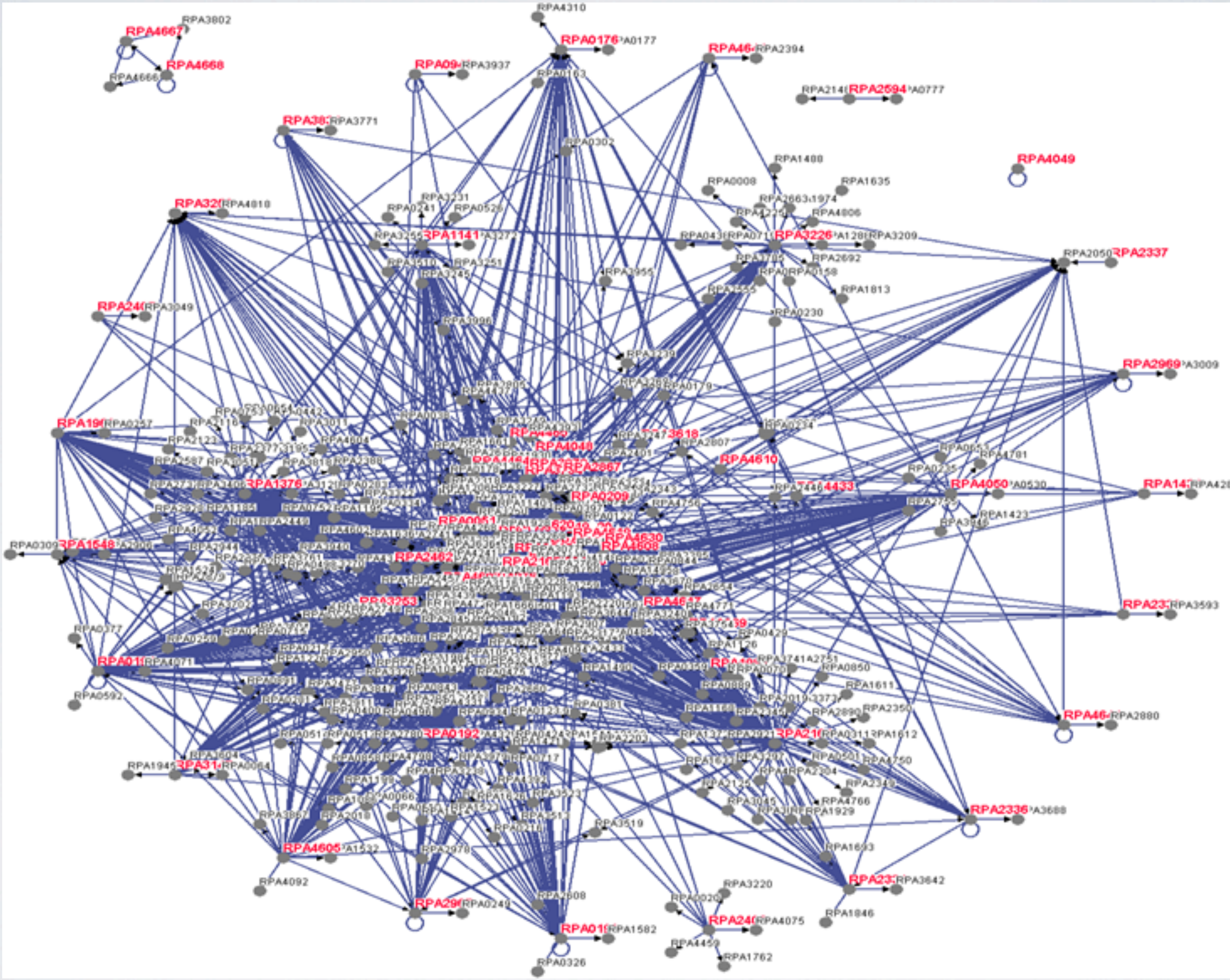
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**59**	

SOCIAL NETWORK ANALYSIS



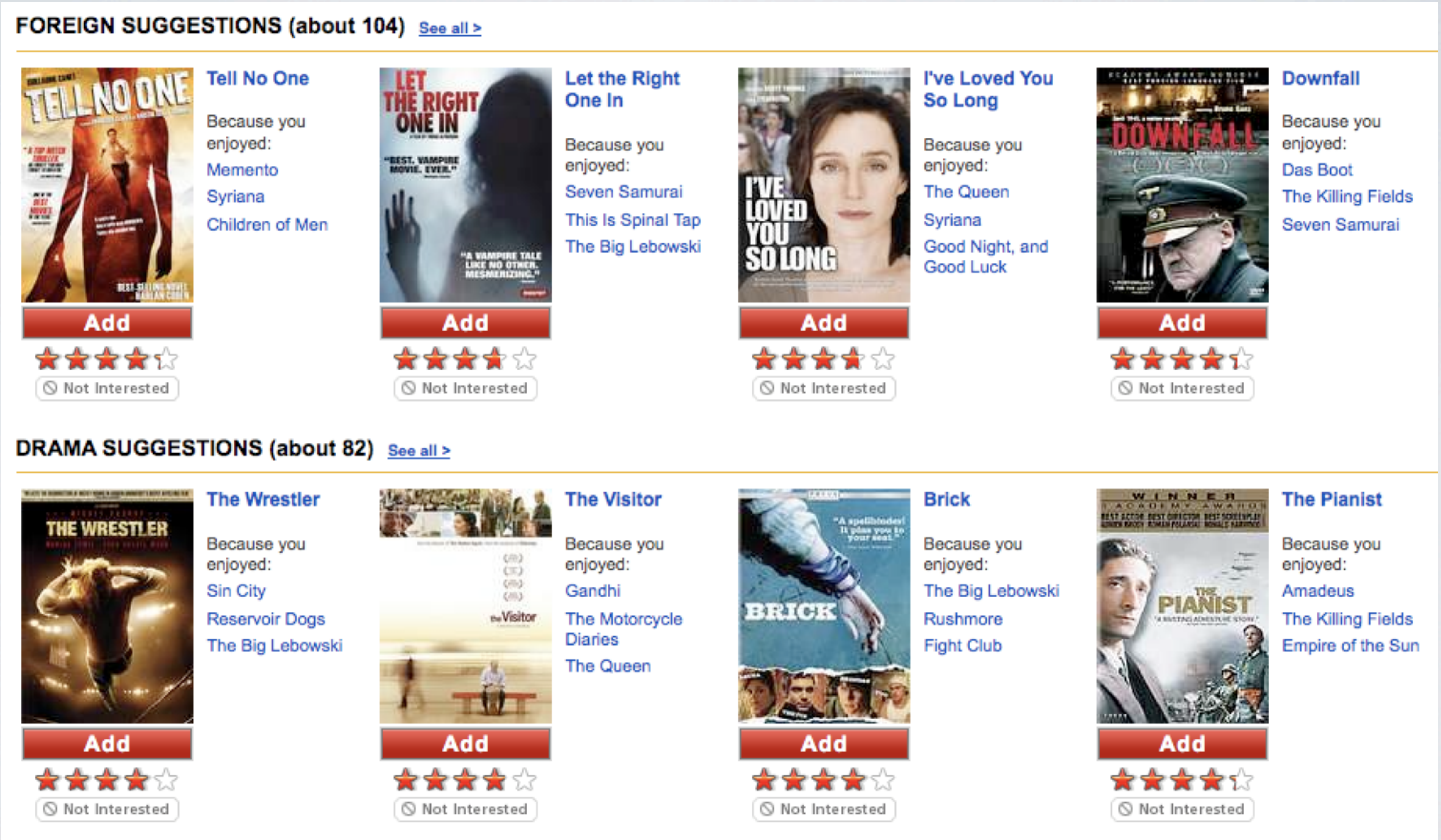
Szell et al, Nature 2012

PROTEIN-PROTEIN INTERACTIONS



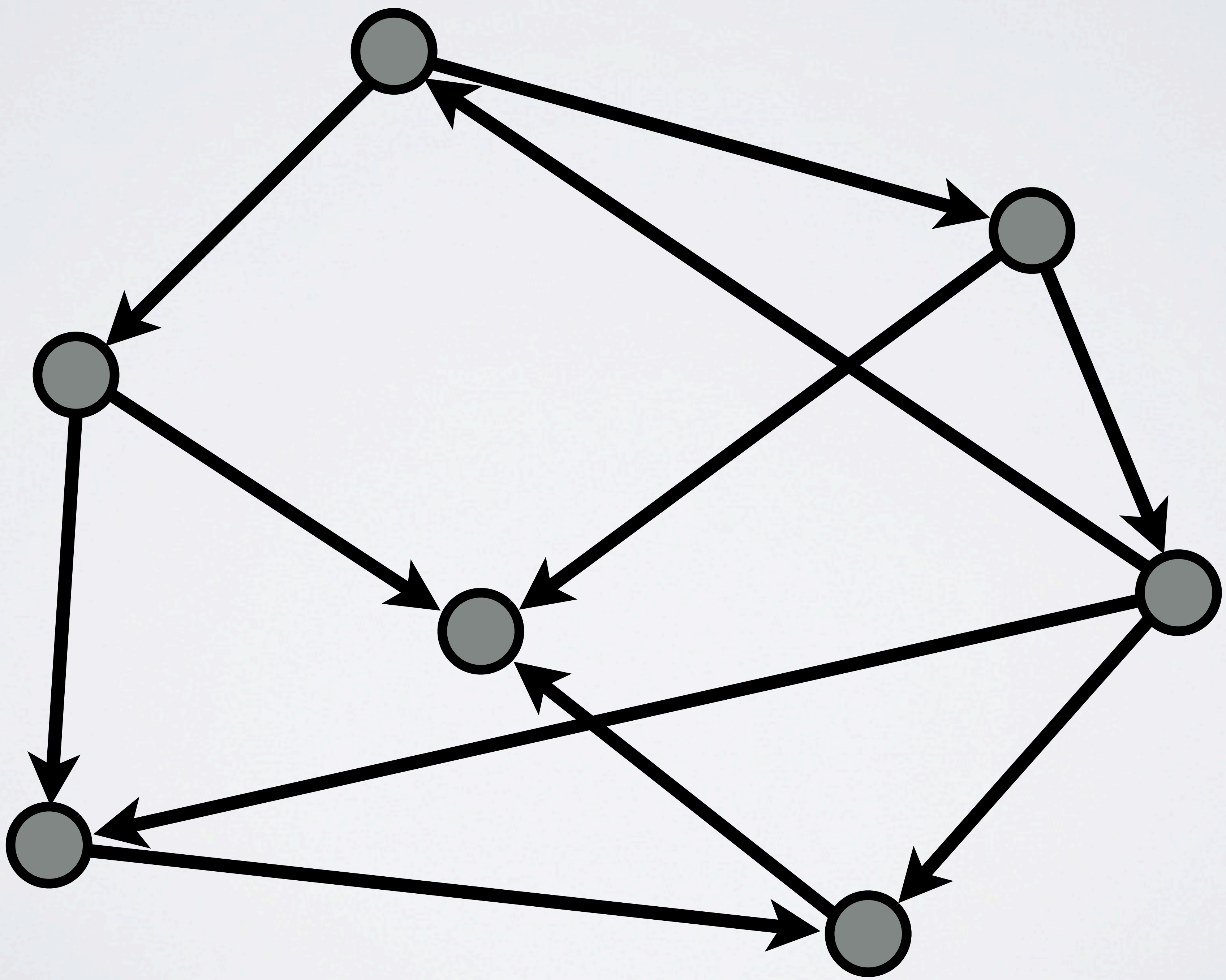
http://mippi.ornl.gov/areas/bioinfo.shtml

COLLABORATIVE FILTERING



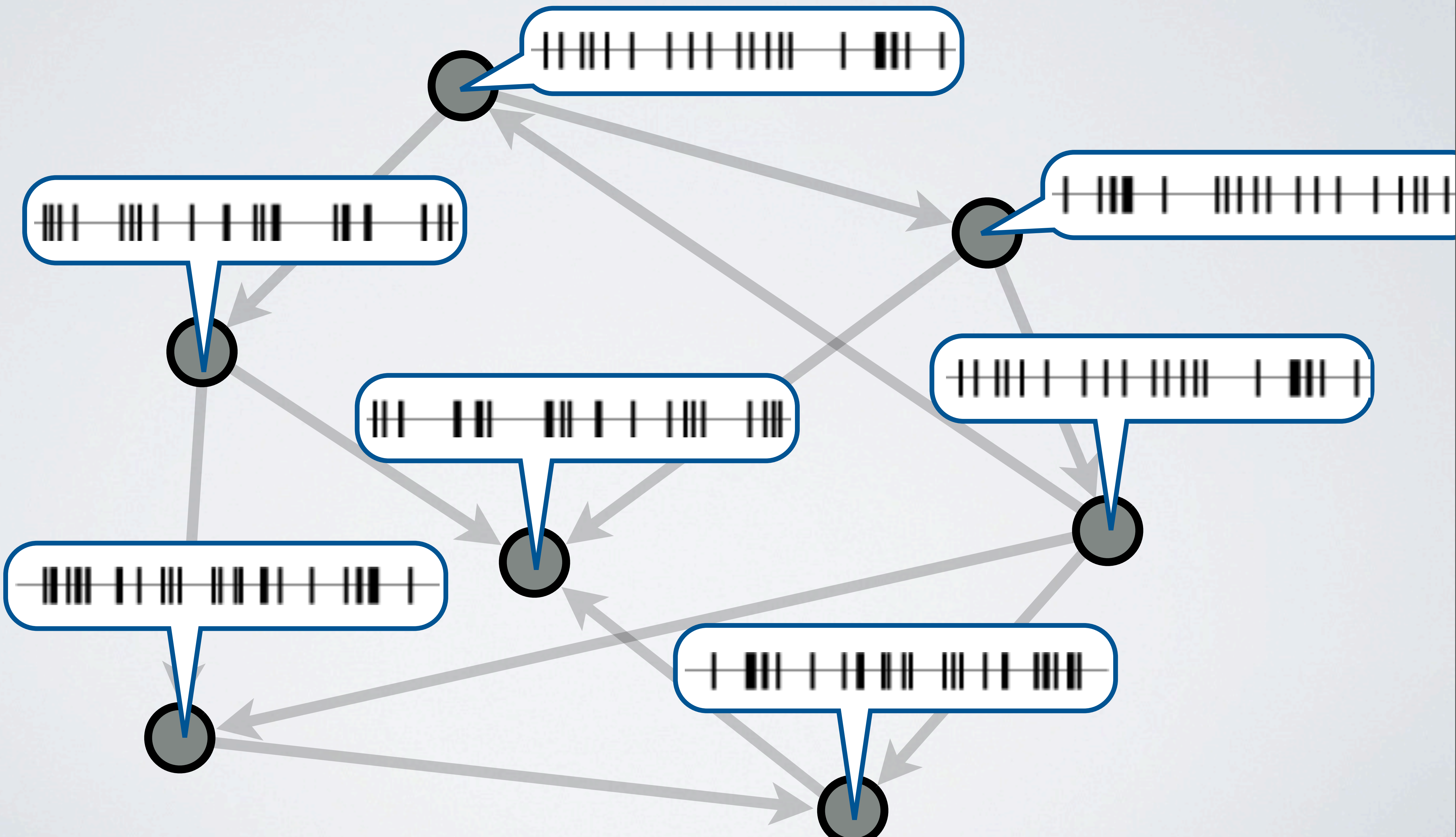
EXPLICIT VS. IMPLICIT NETWORKS

Typically, we see edges and reason about the latent properties of the vertices.



Explicit vs. Implicit Networks

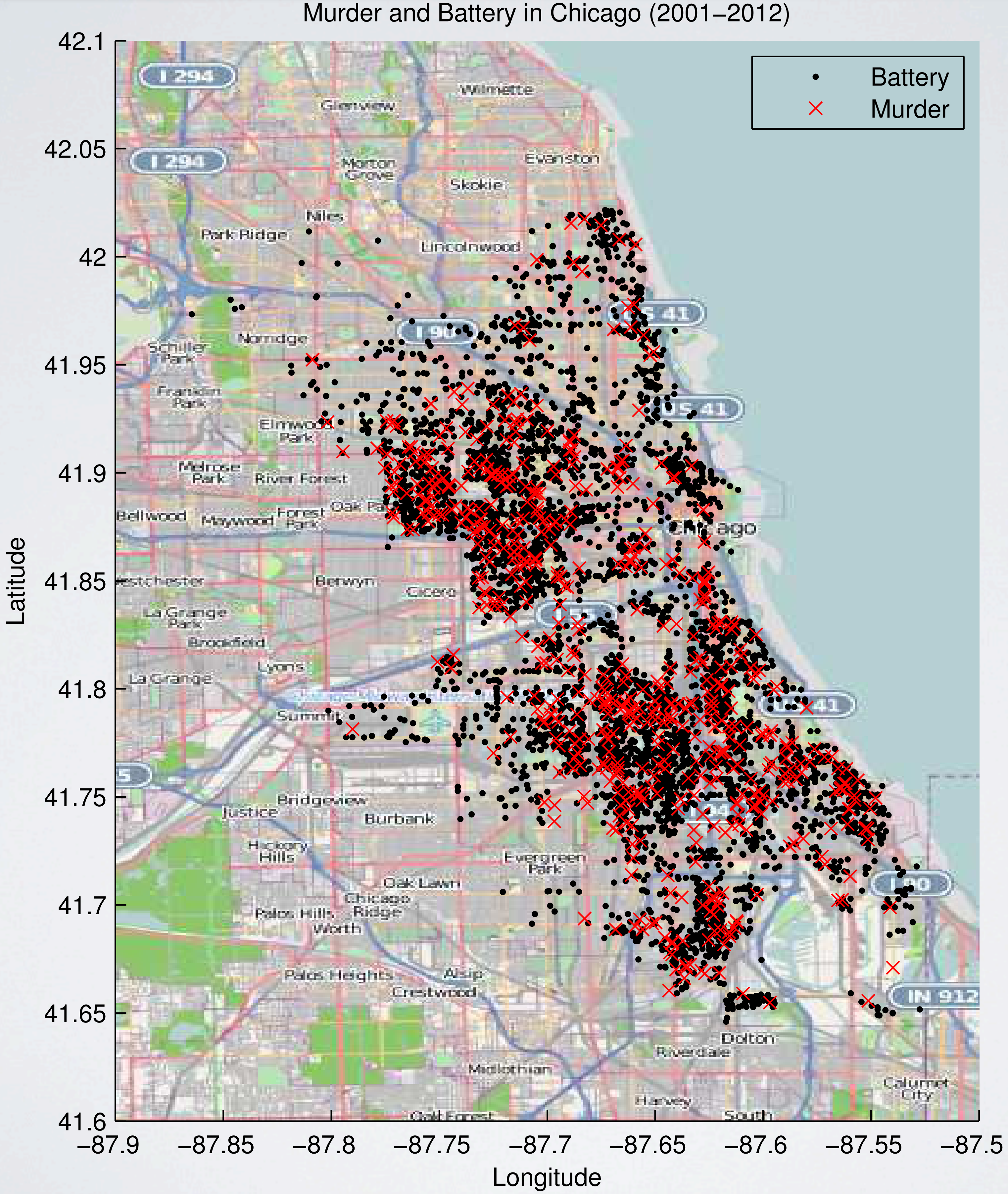
What if we don't observe edges, but only noisy emissions from each vertex?



FUNCTIONAL CONNECTIVITY OF NEURONS

Truccolo, Hochberg & Donoghue, 2010

Patterns of Gang Violence



Overview

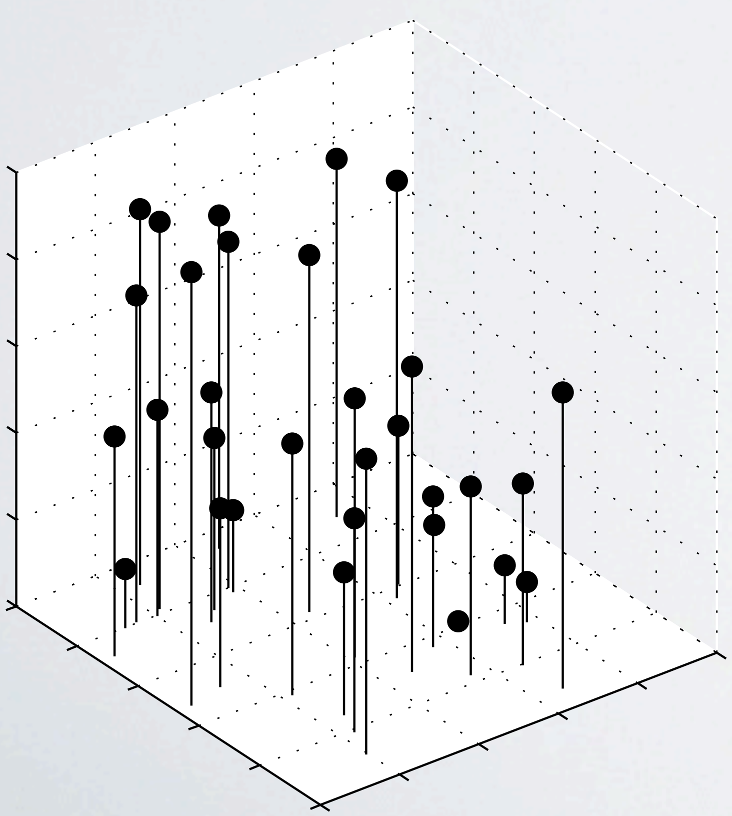
- Mutually-Exciting Point Processes
- Aldous-Hoover Graph Priors
- MCMC Inference with Data Augmentation
- Application Examples
- Extending for Neural Models with Inhibition

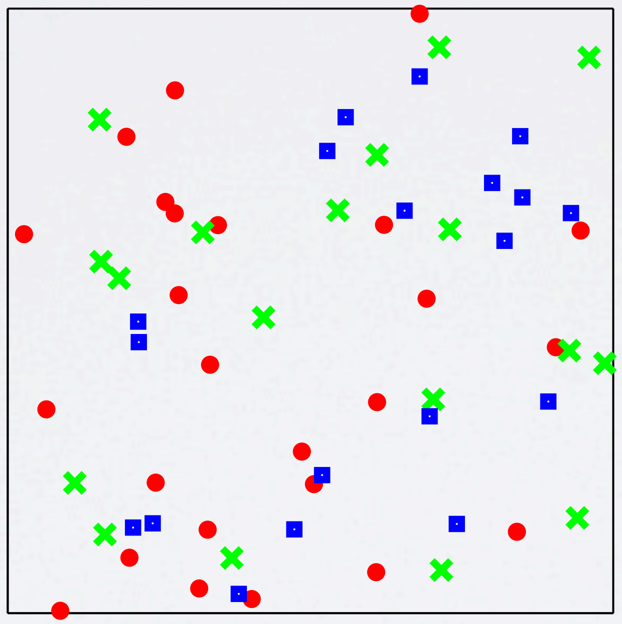
MULTIVARIATE POINT PROCESSES

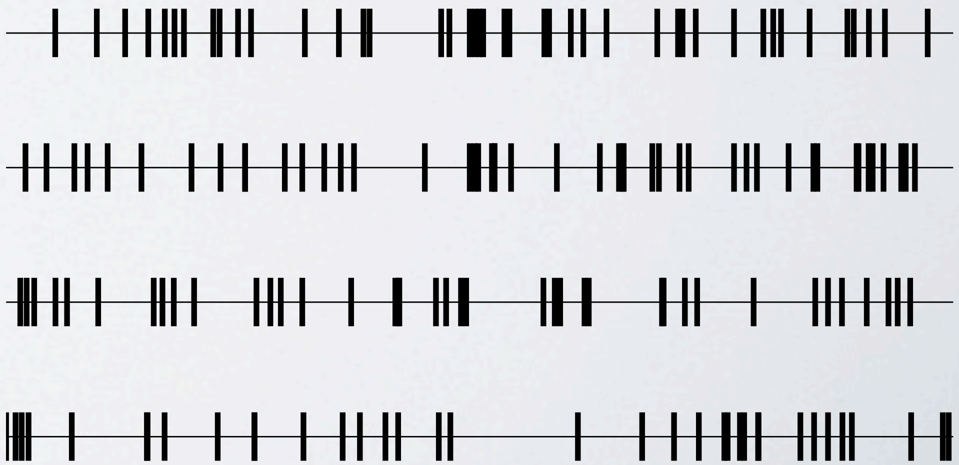
- The point process is a foundational statistical object.
- Gives us random subsets of a larger space.
- Many data are well modeled as point processes:
  - Seismology
  - Epidemiology
  - Economics
- Modeling dependence is challenging - “beyond Poisson”
  - Strauss and Gibbs Processes
  - Determinantal and Permanental Point Processes
  - TODAY: Mutually-Exciting (Hawkes) Processes

Point Process

- 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.

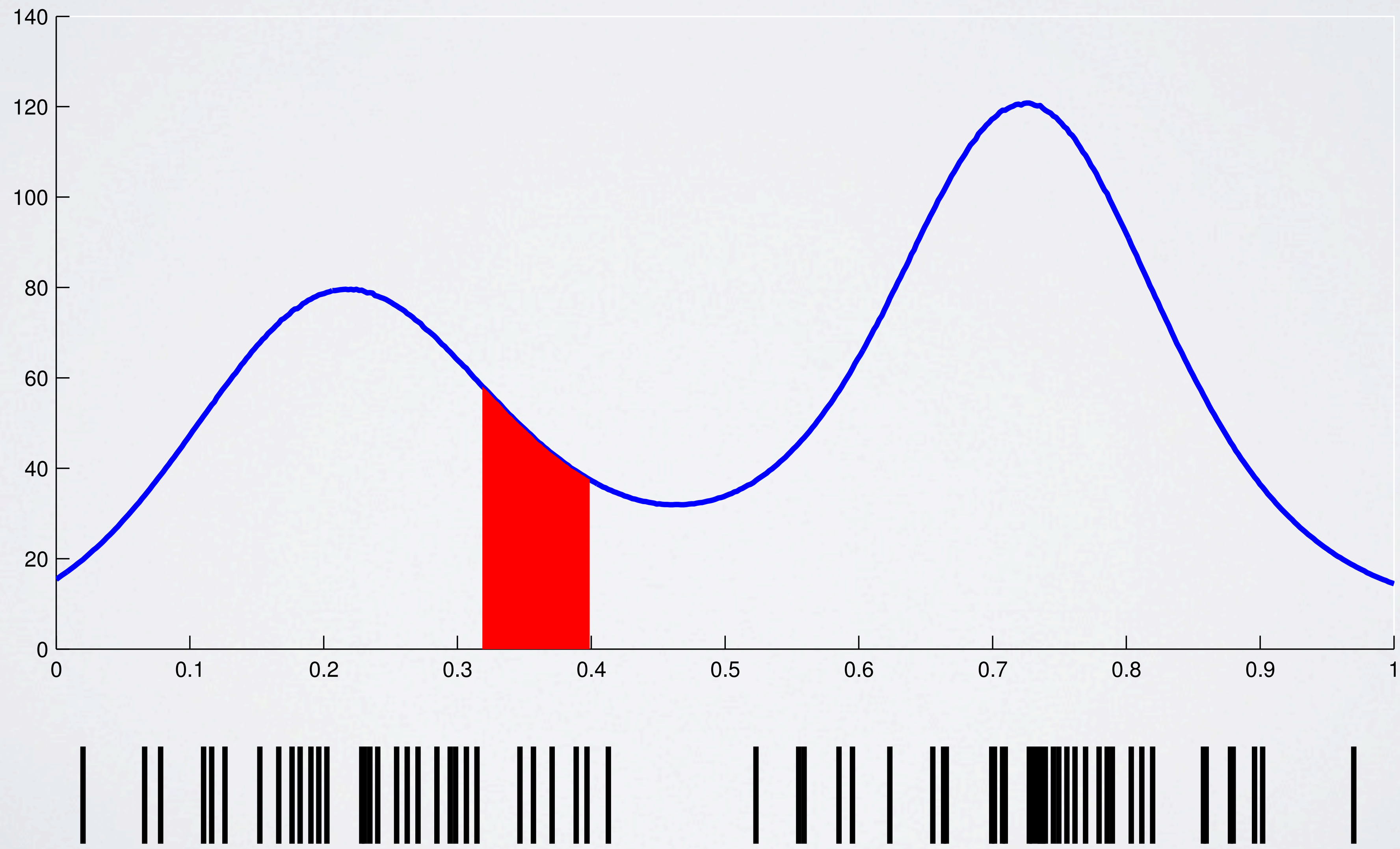






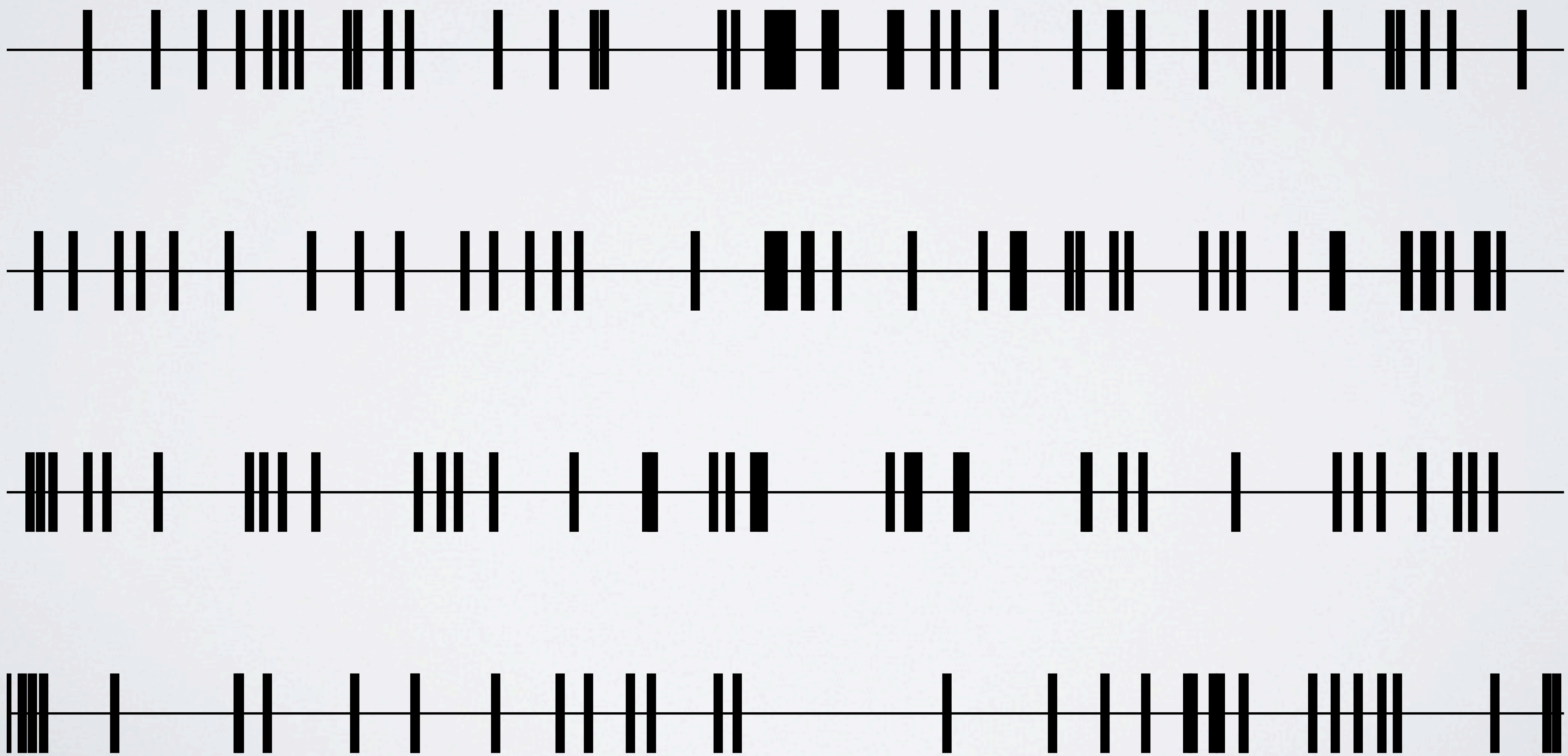
Poisson Process

- 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.

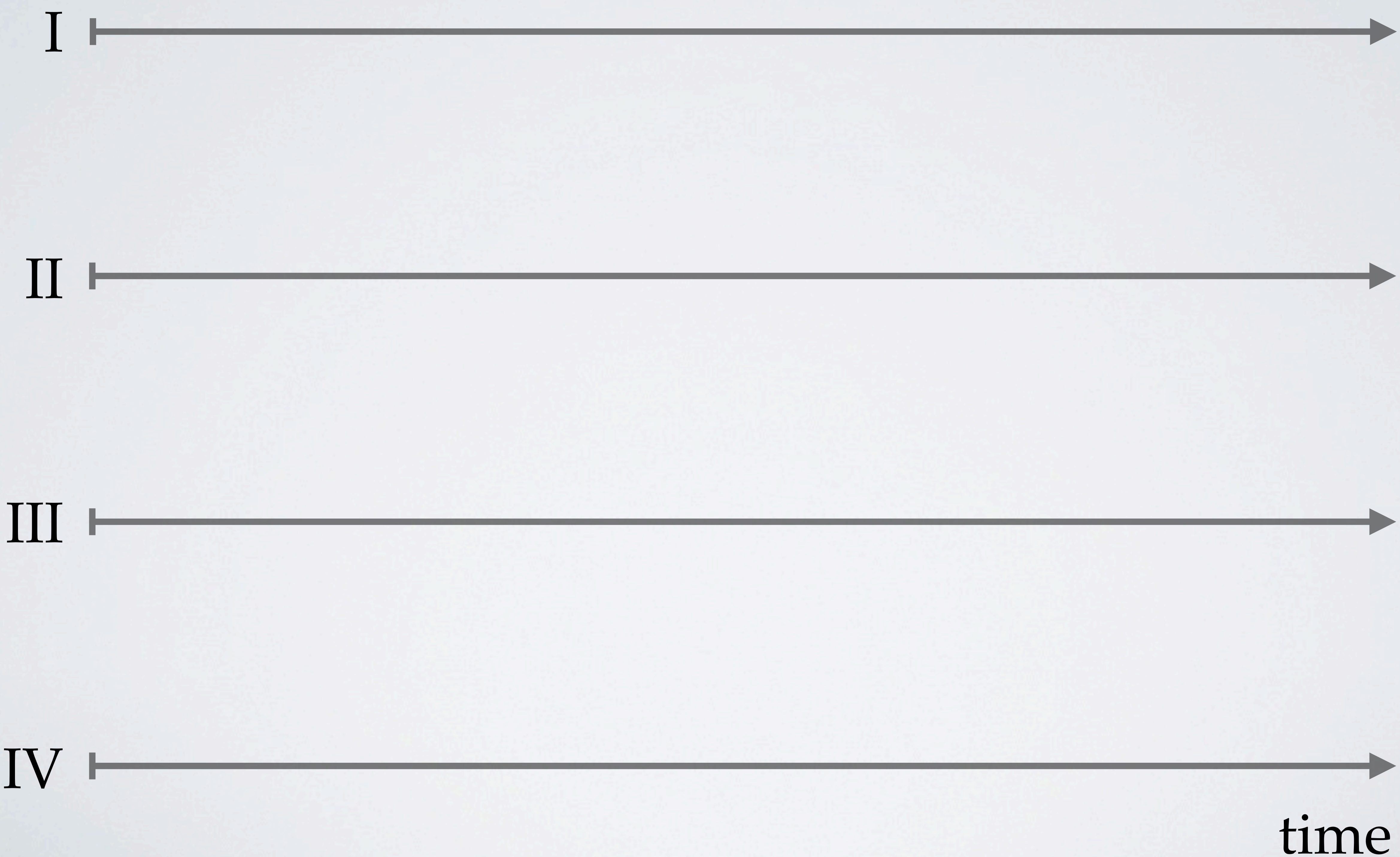


Dependence in Multivariate Point Processes

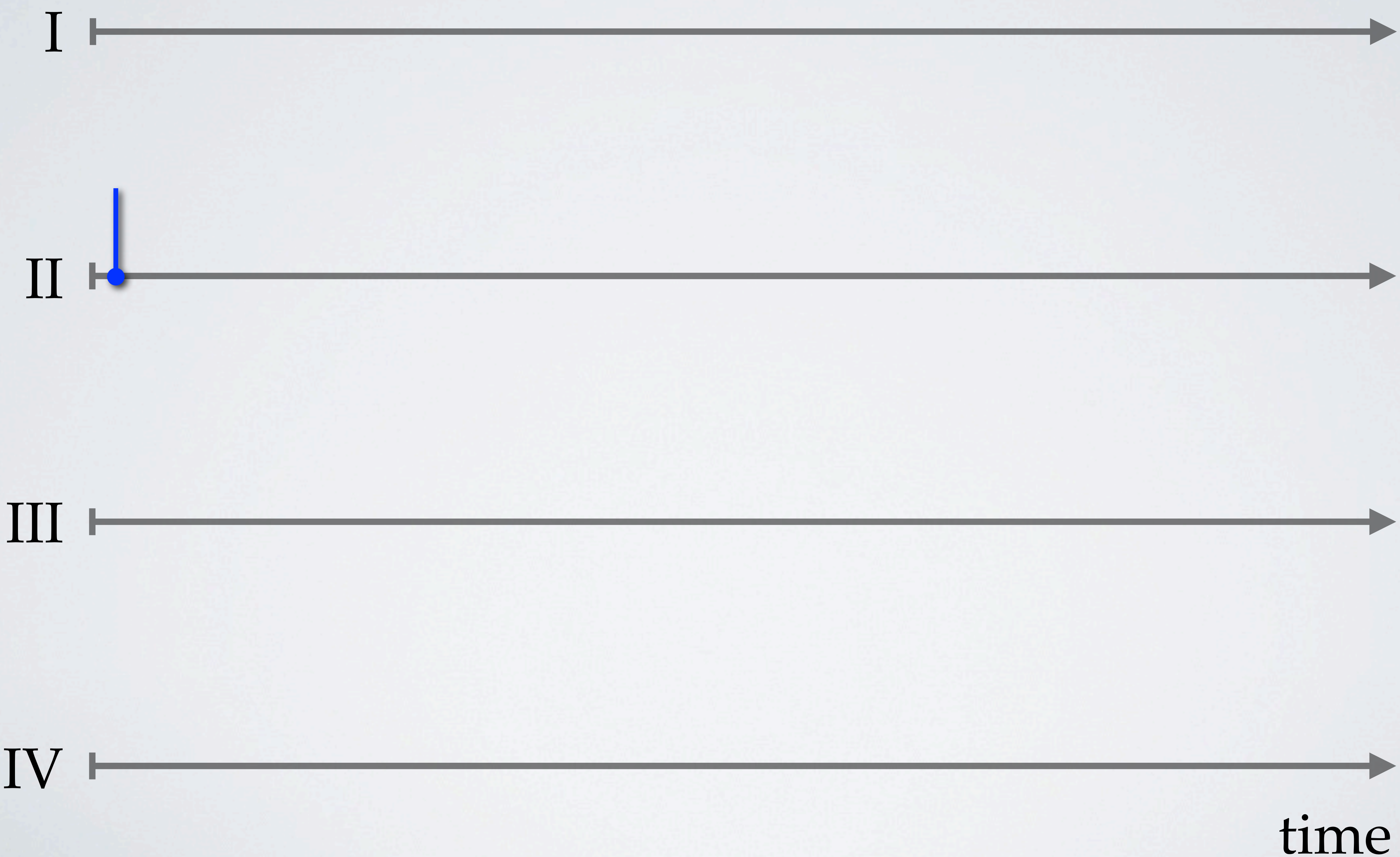
- In the Poisson process, everything is conditionally independent given the rate function.
- How can we get spike-driven dynamics?



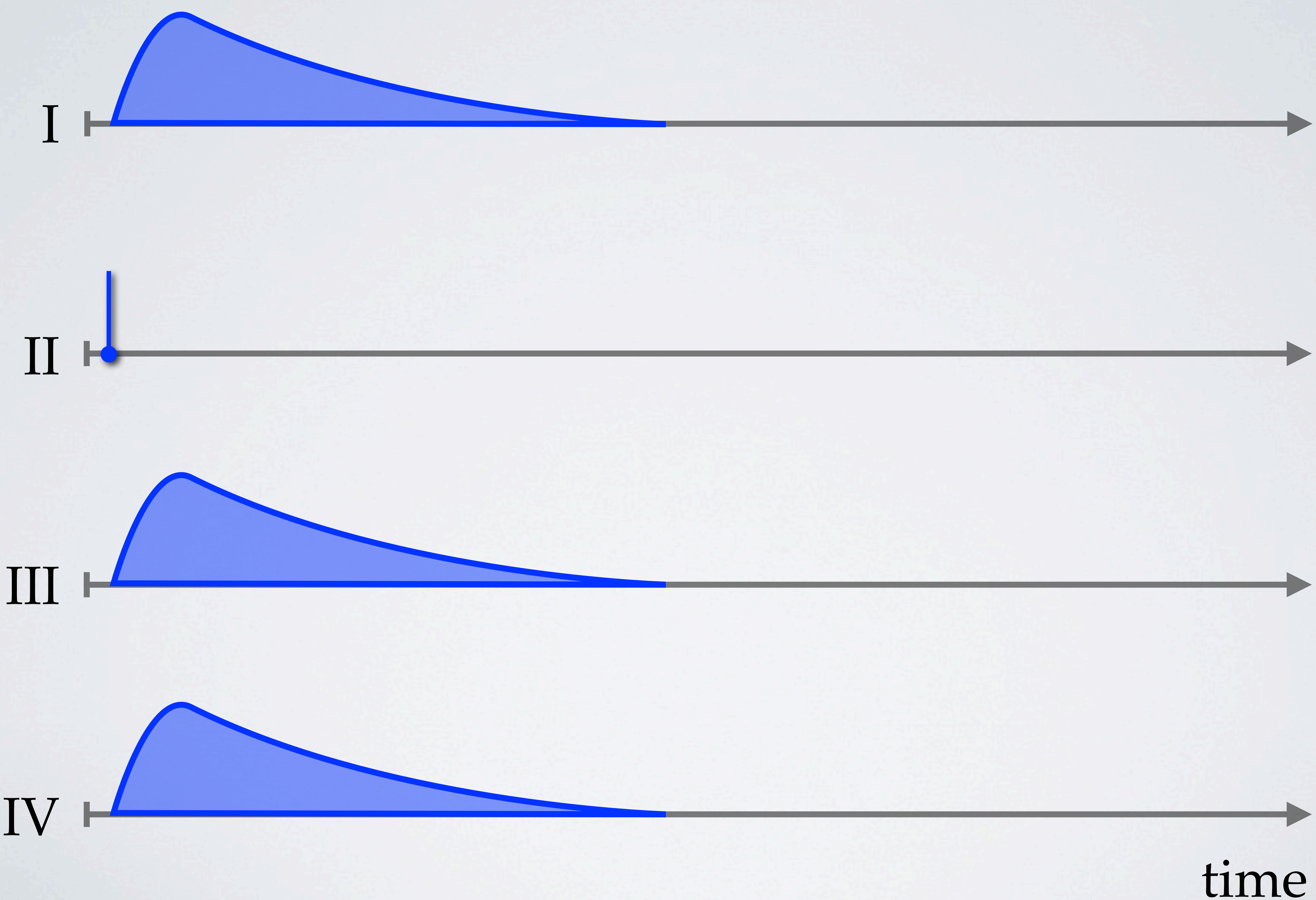
HAWKES PROCESS DYNAMICS



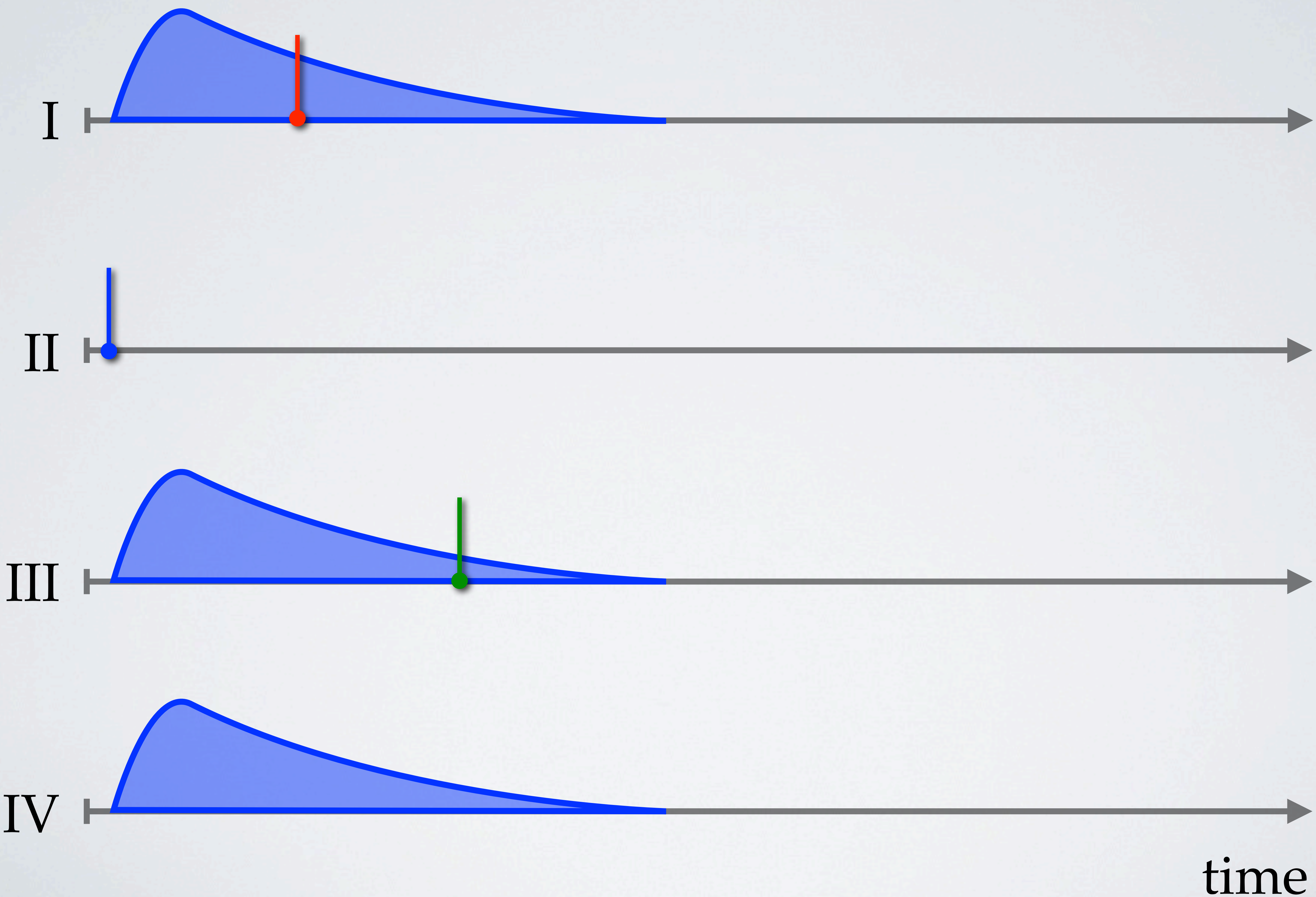
HAWKES PROCESS DYNAMICS



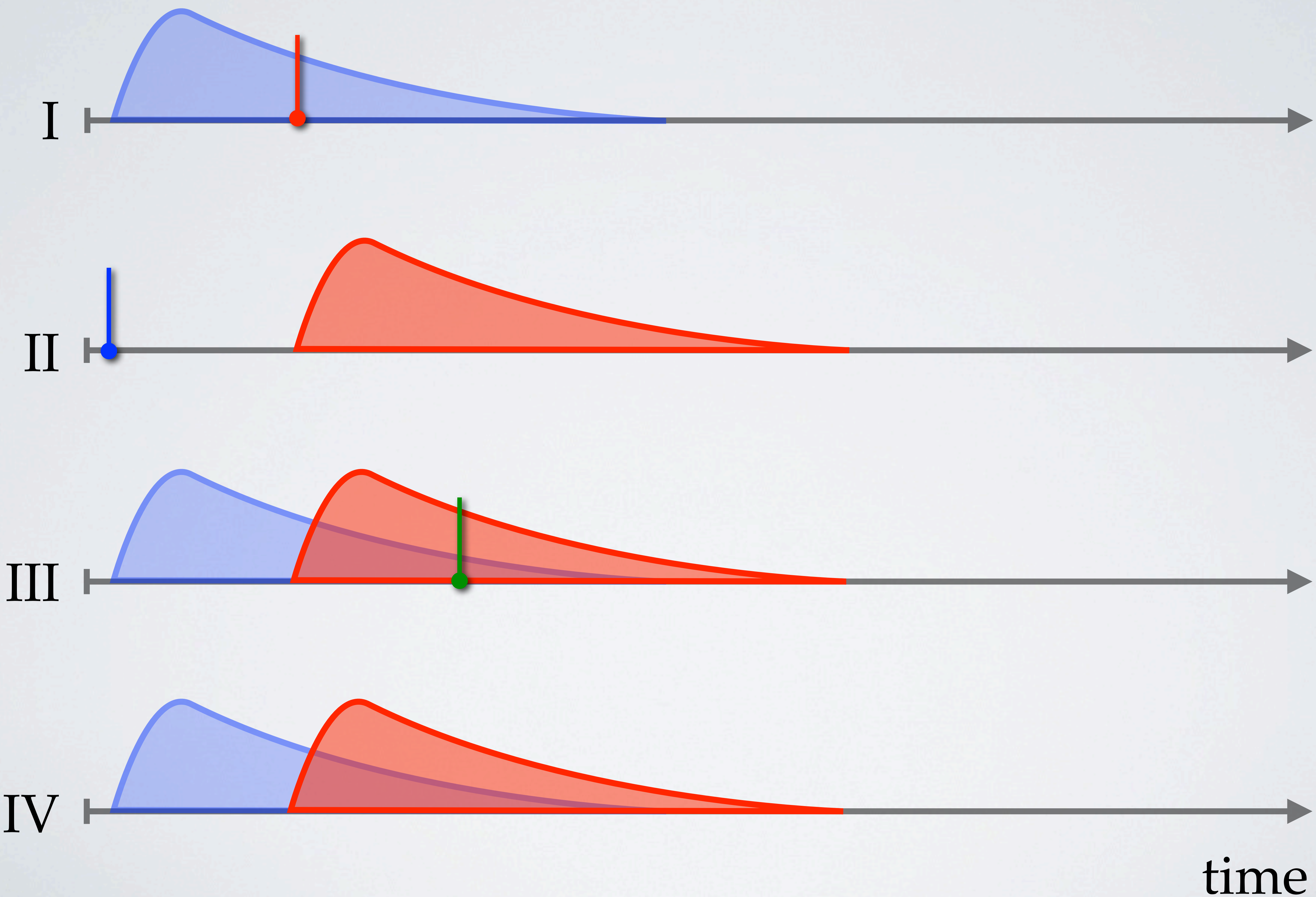
HAWKES PROCESS DYNAMICS



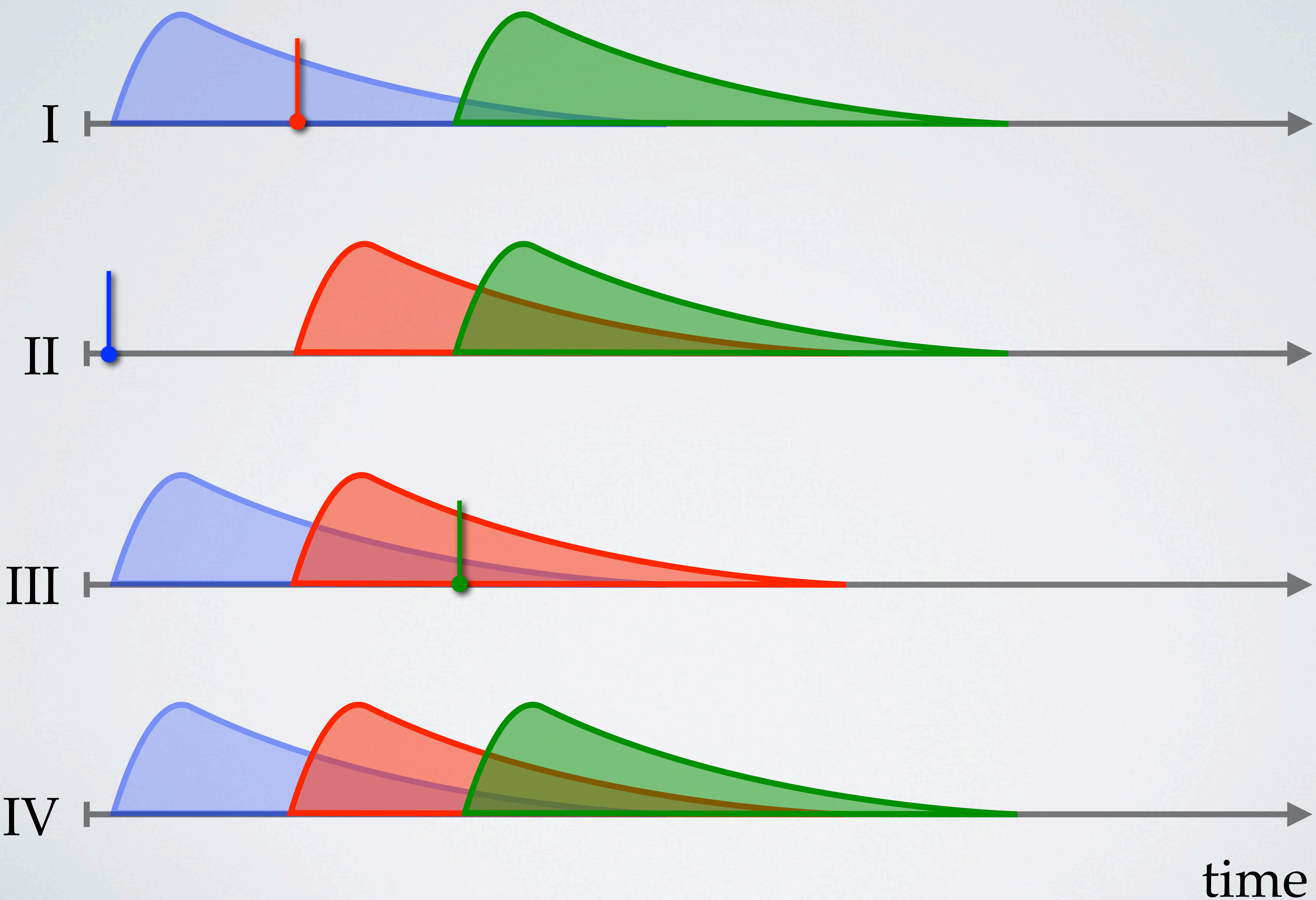
Hawkes Process Dynamics



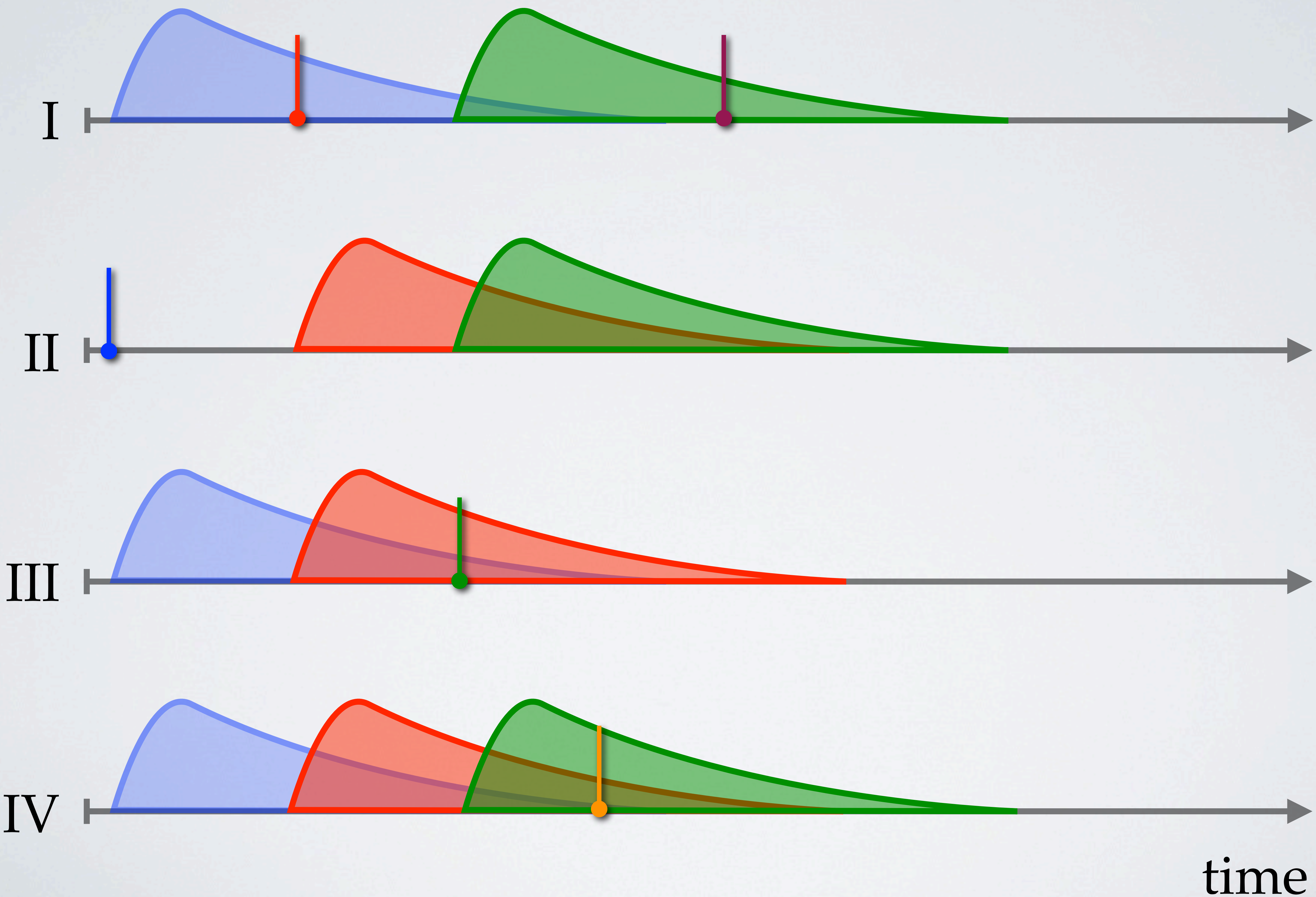
Hawkes Process Dynamics



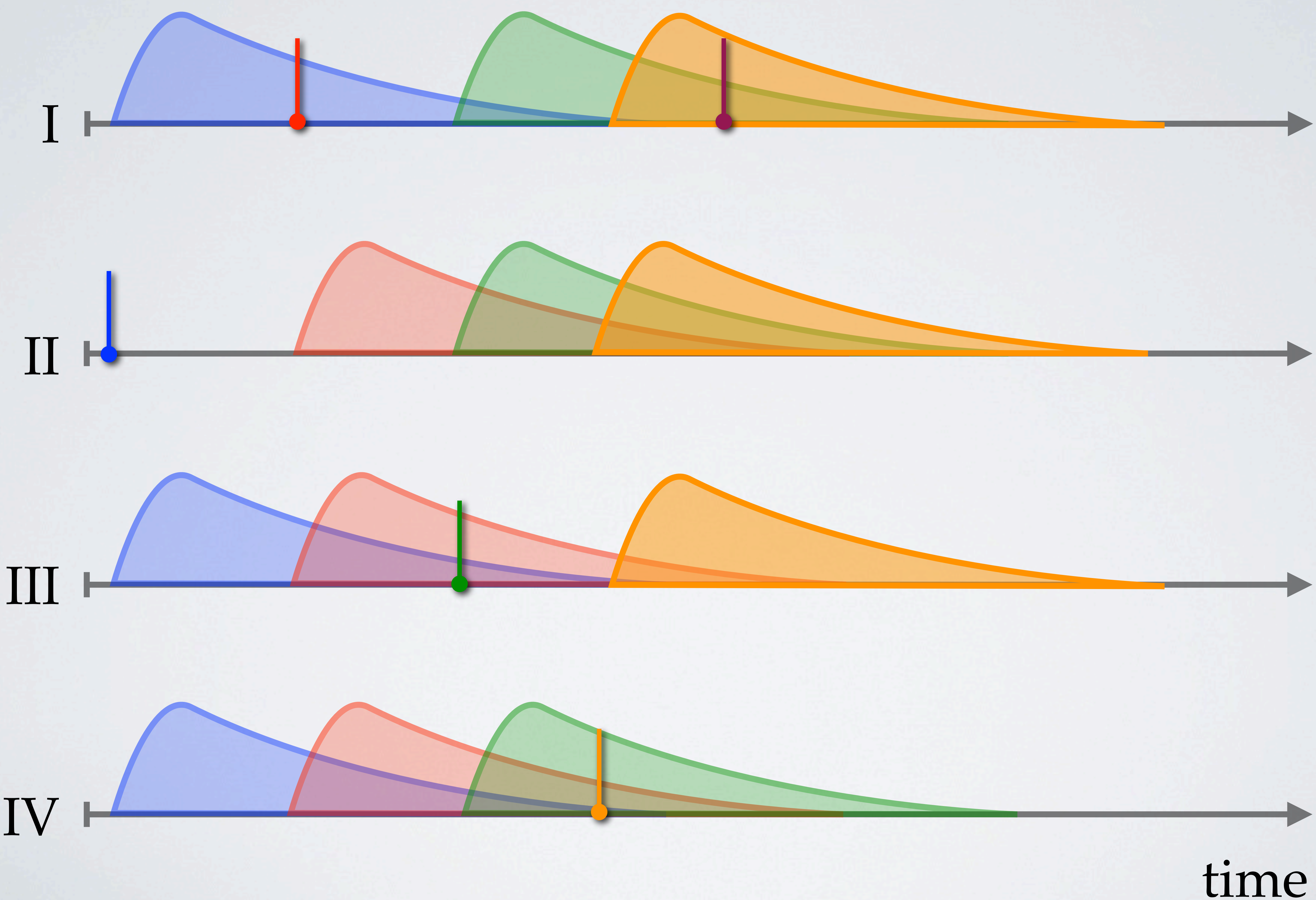
HAWKES PROCESS DYNAMICS



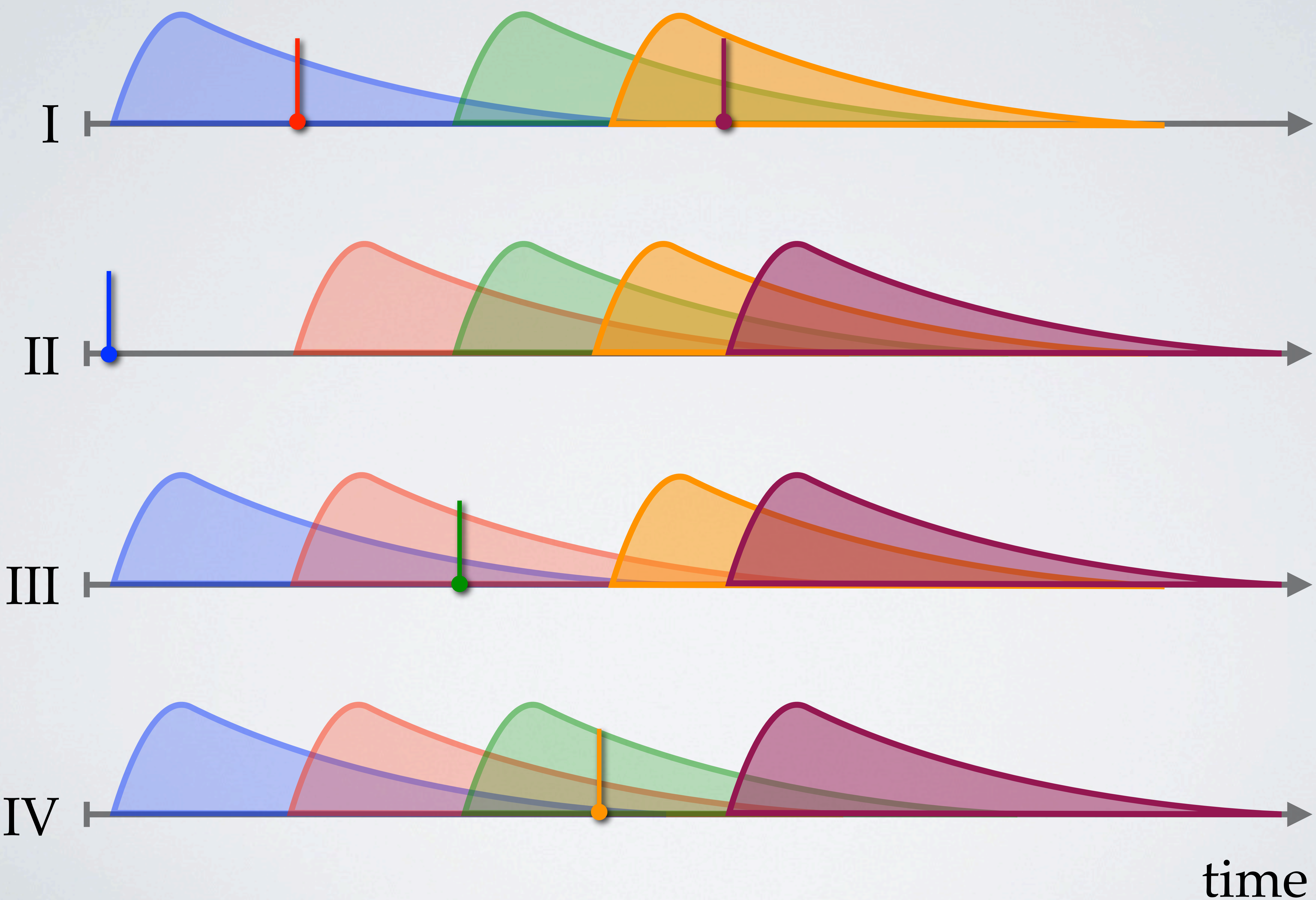
HAWKES PROCESS DYNAMICS



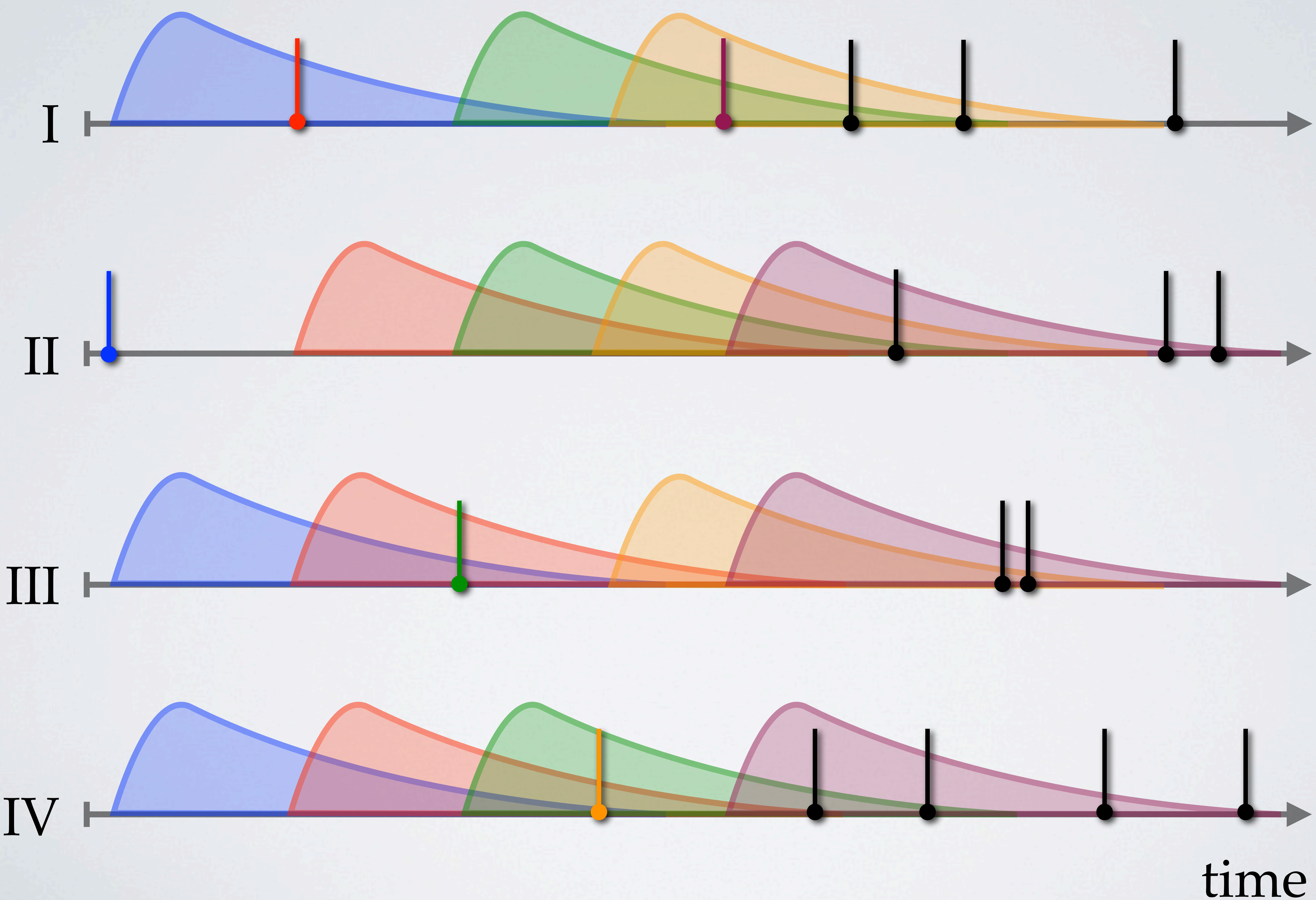
Hawkes Process Dynamics



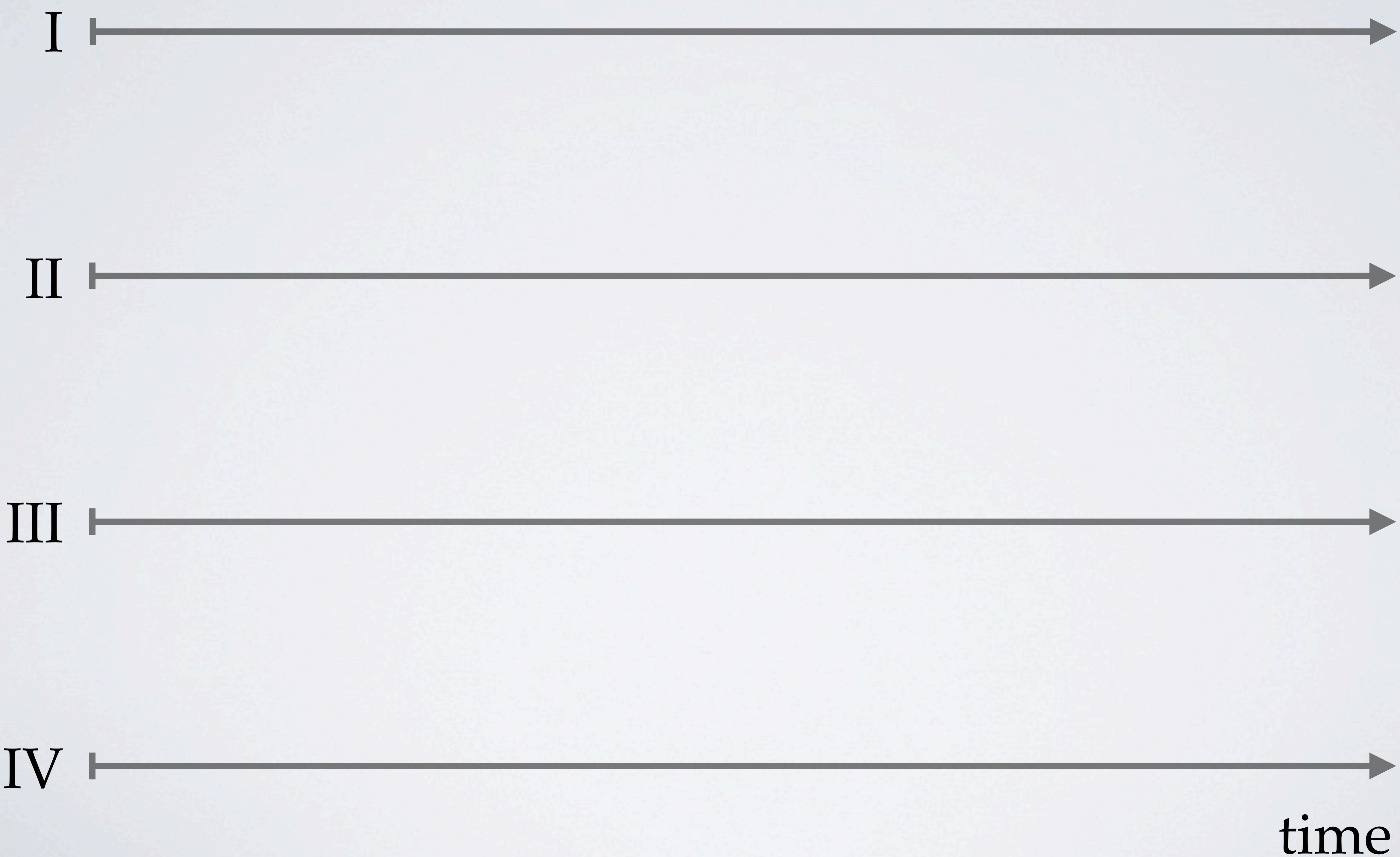
Hawkes Process Dynamics



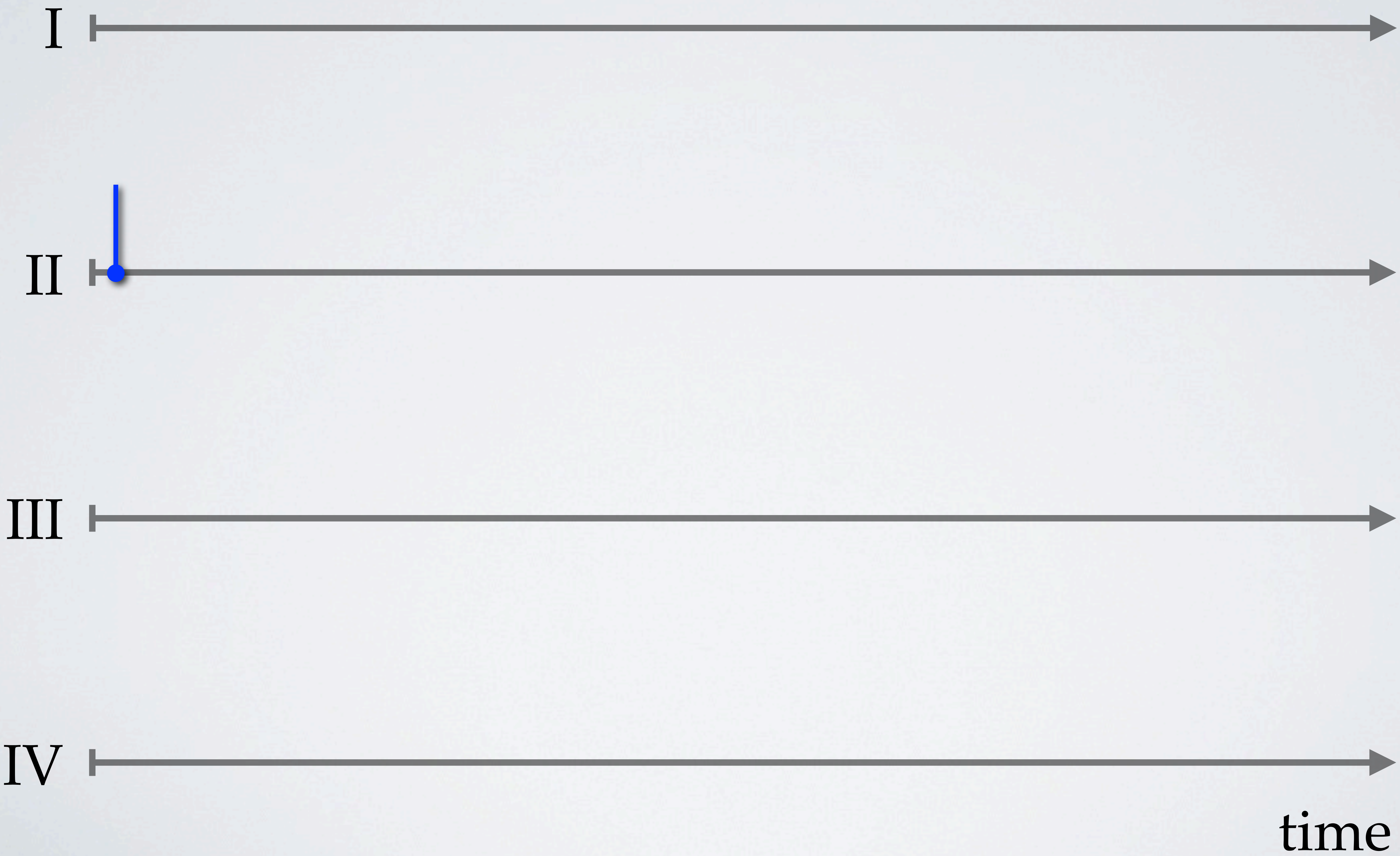
HAWKES PROCESS DYNAMICS



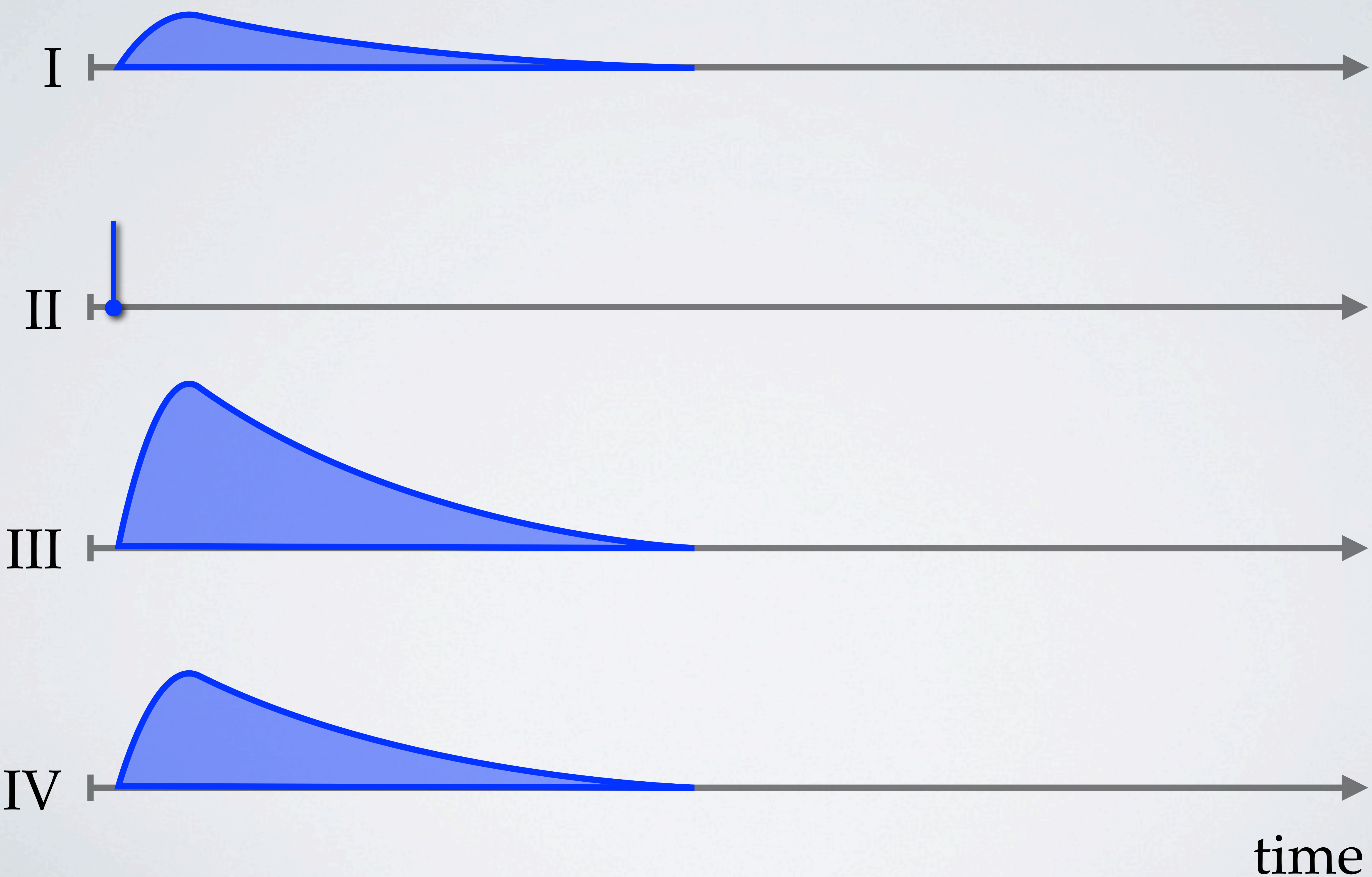
HAWKES PROCESS DYNAMICS



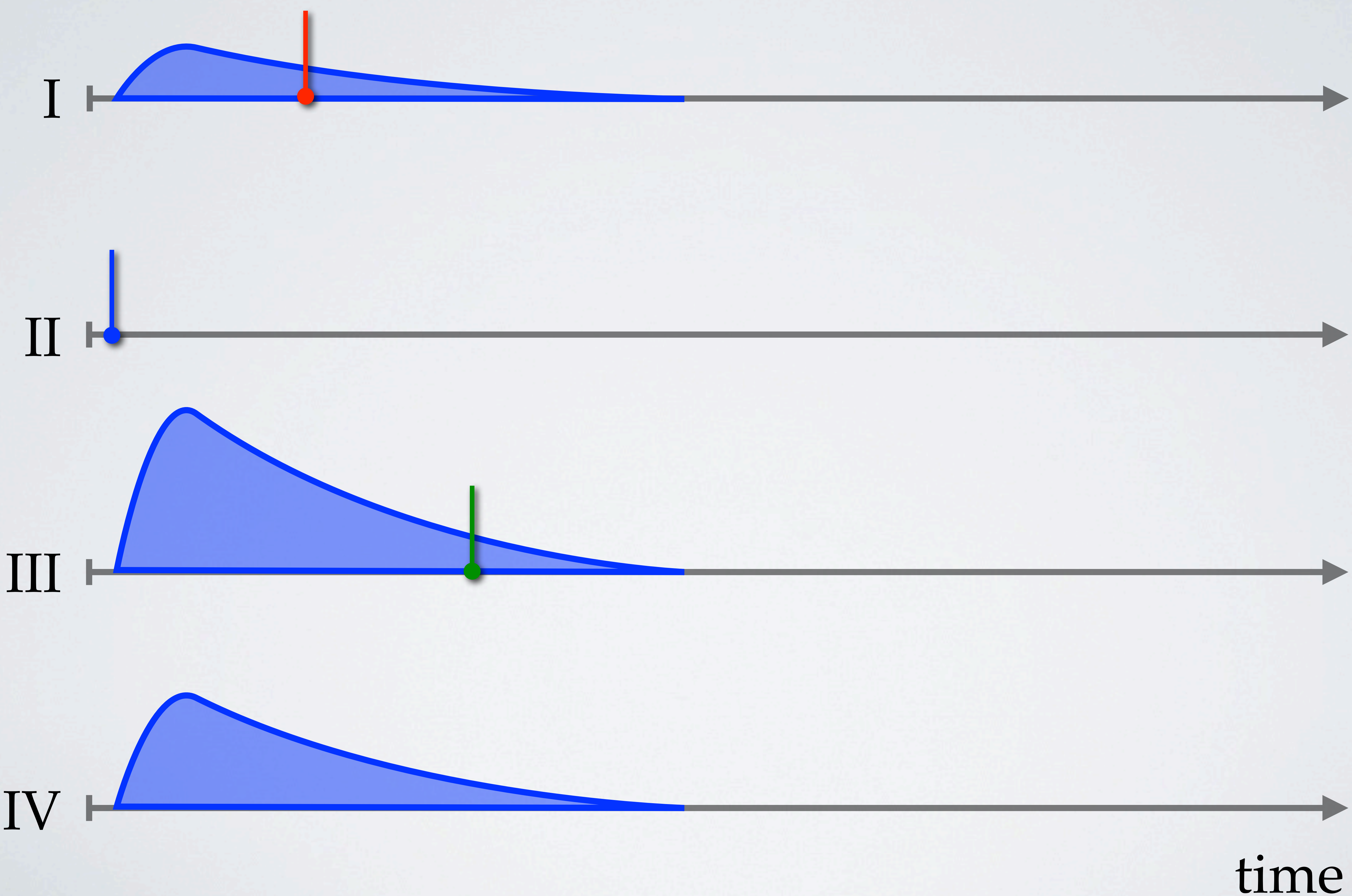
Hawkes Process Dynamics



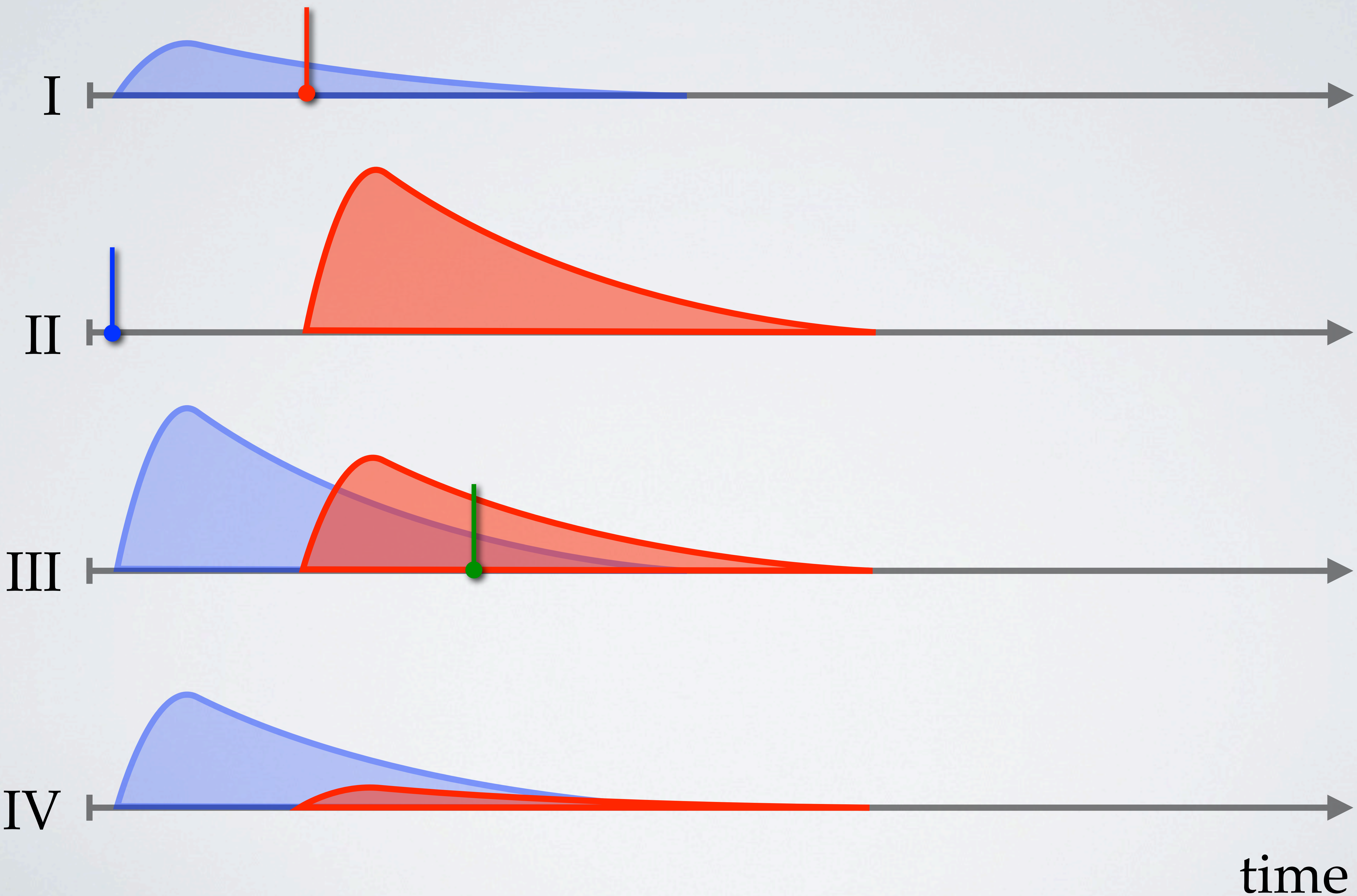
HAWKES PROCESS DYNAMICS



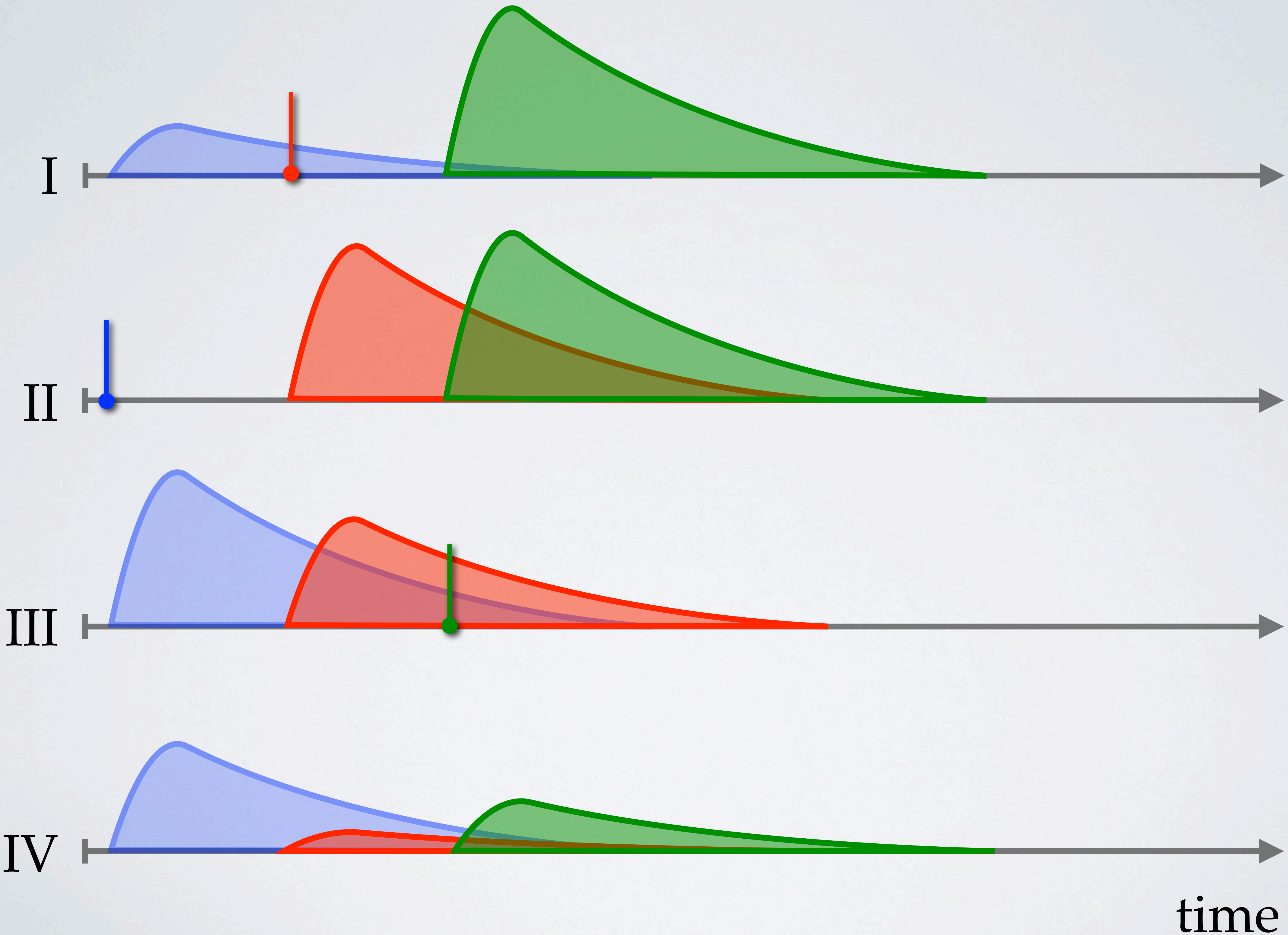
HAWKES PROCESS DYNAMICS



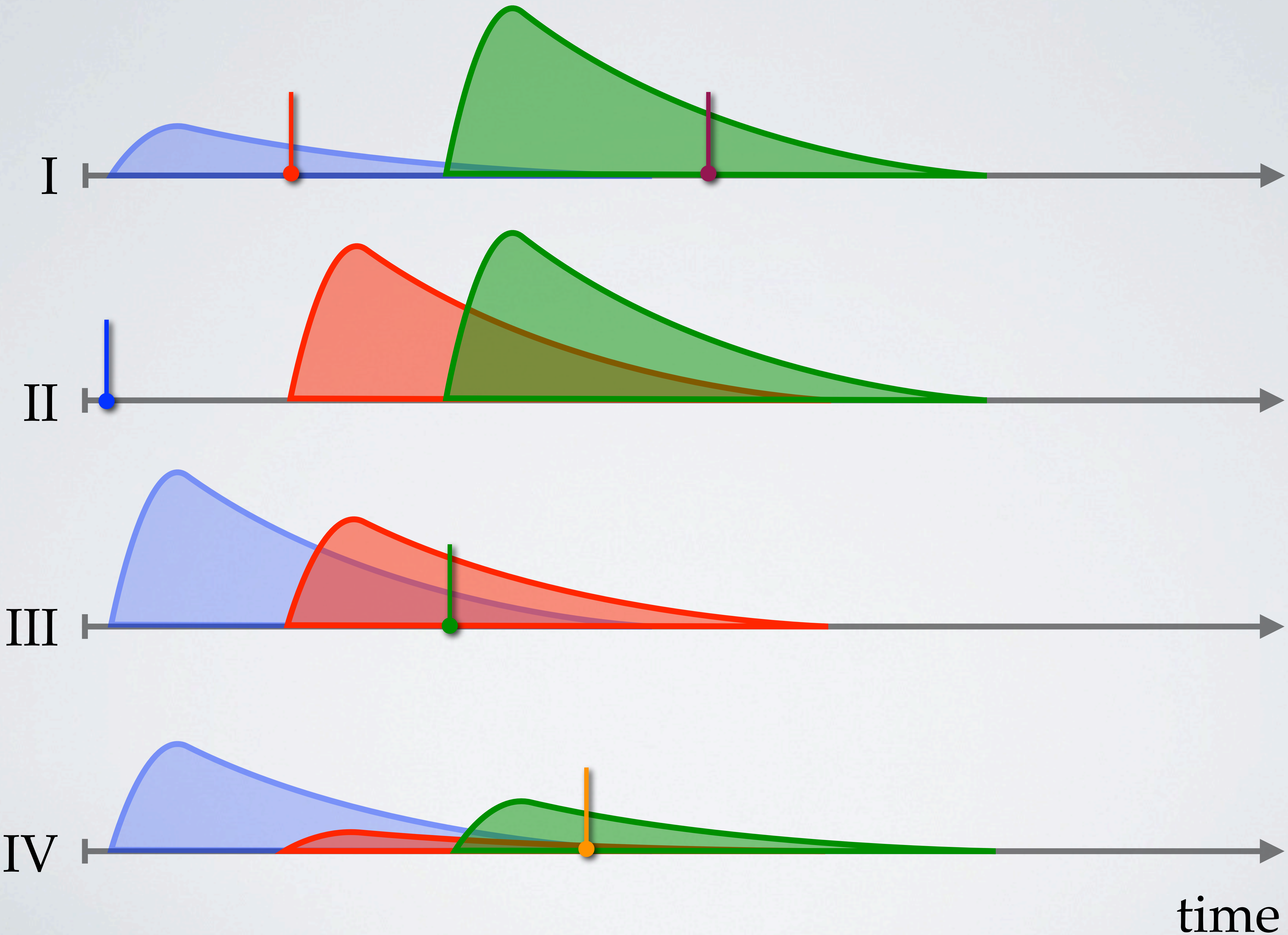
Hawkes Process Dynamics



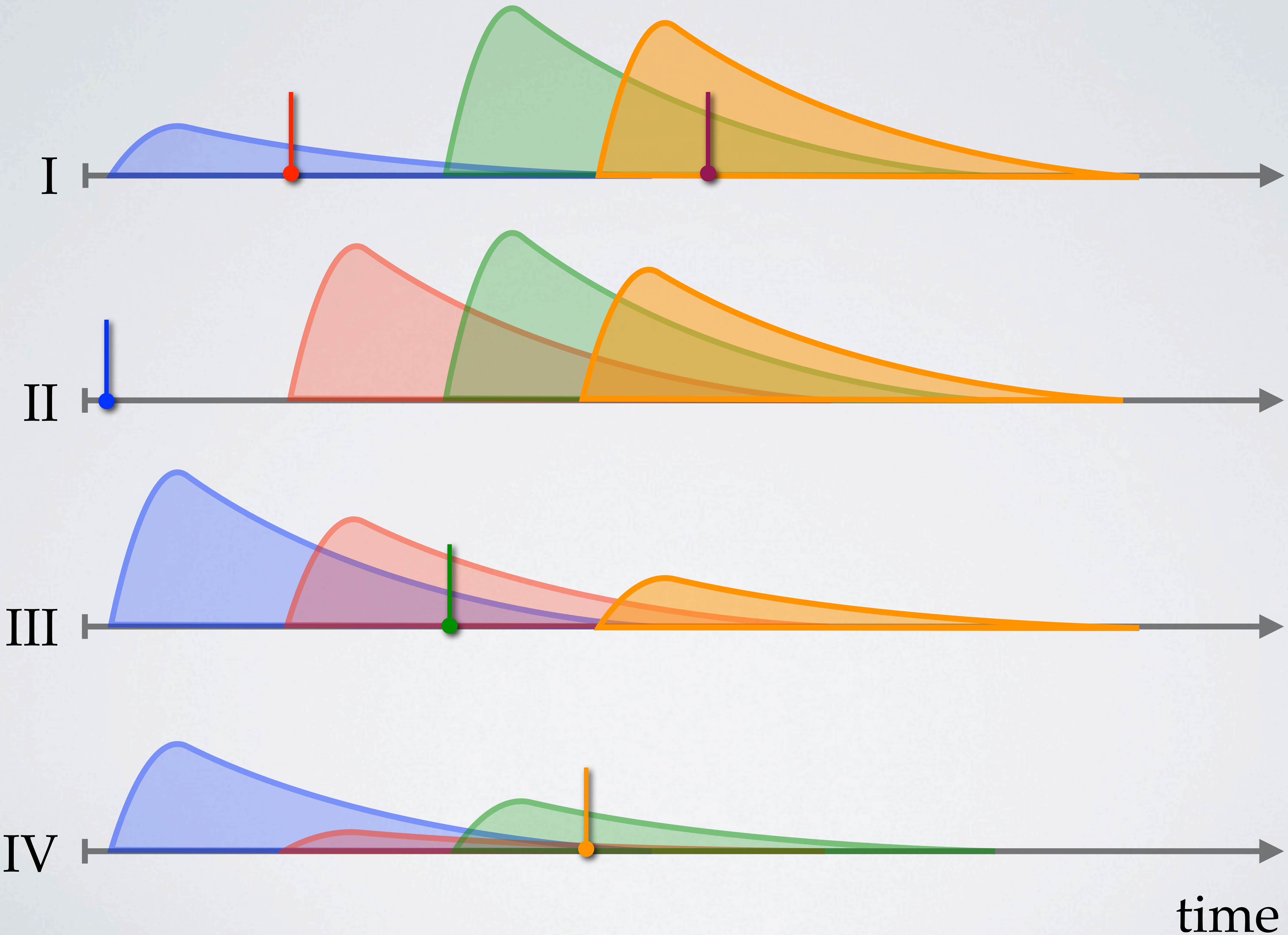
HAWKES PROCESS DYNAMICS



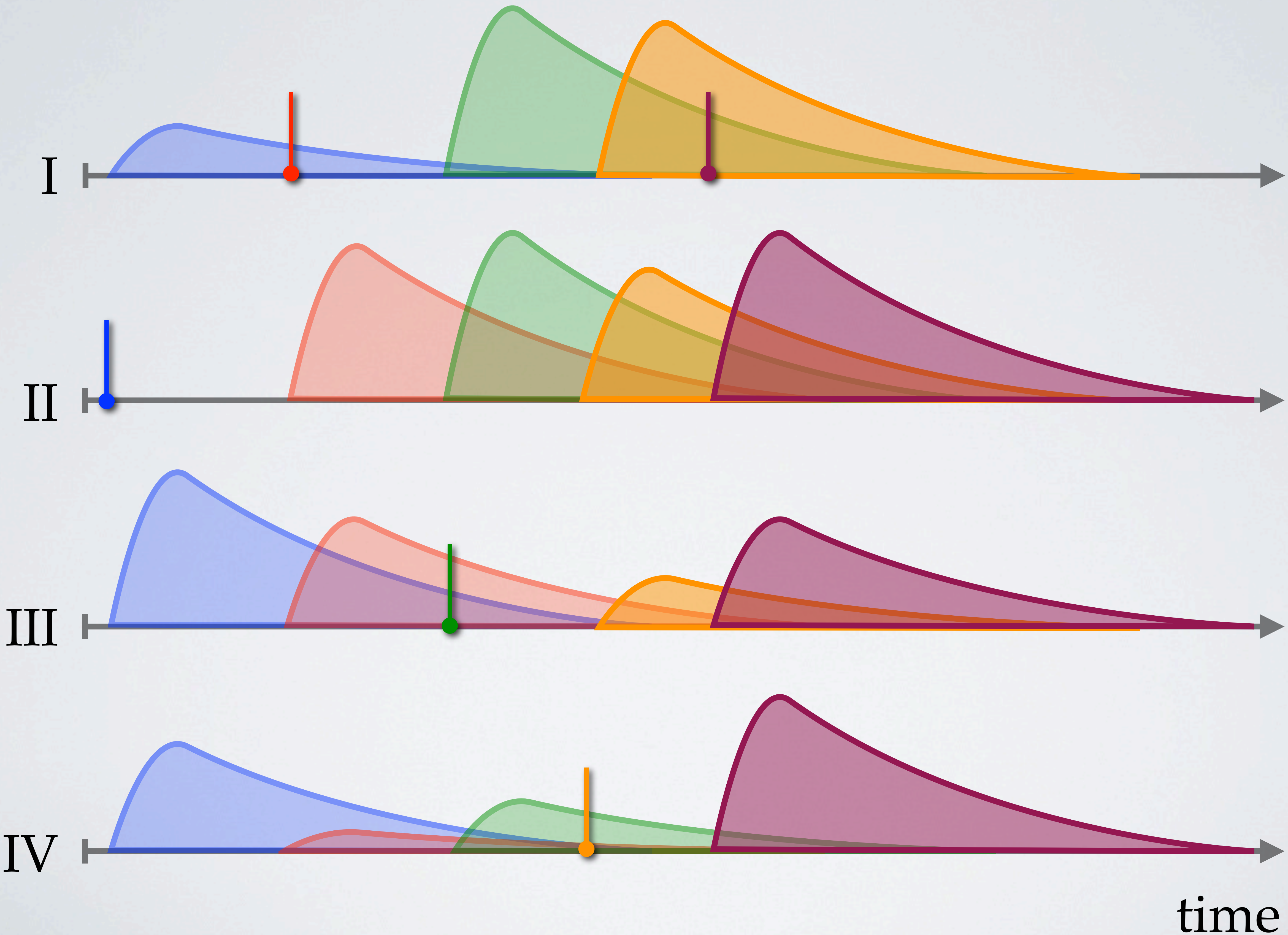
HAWKES PROCESS DYNAMICS



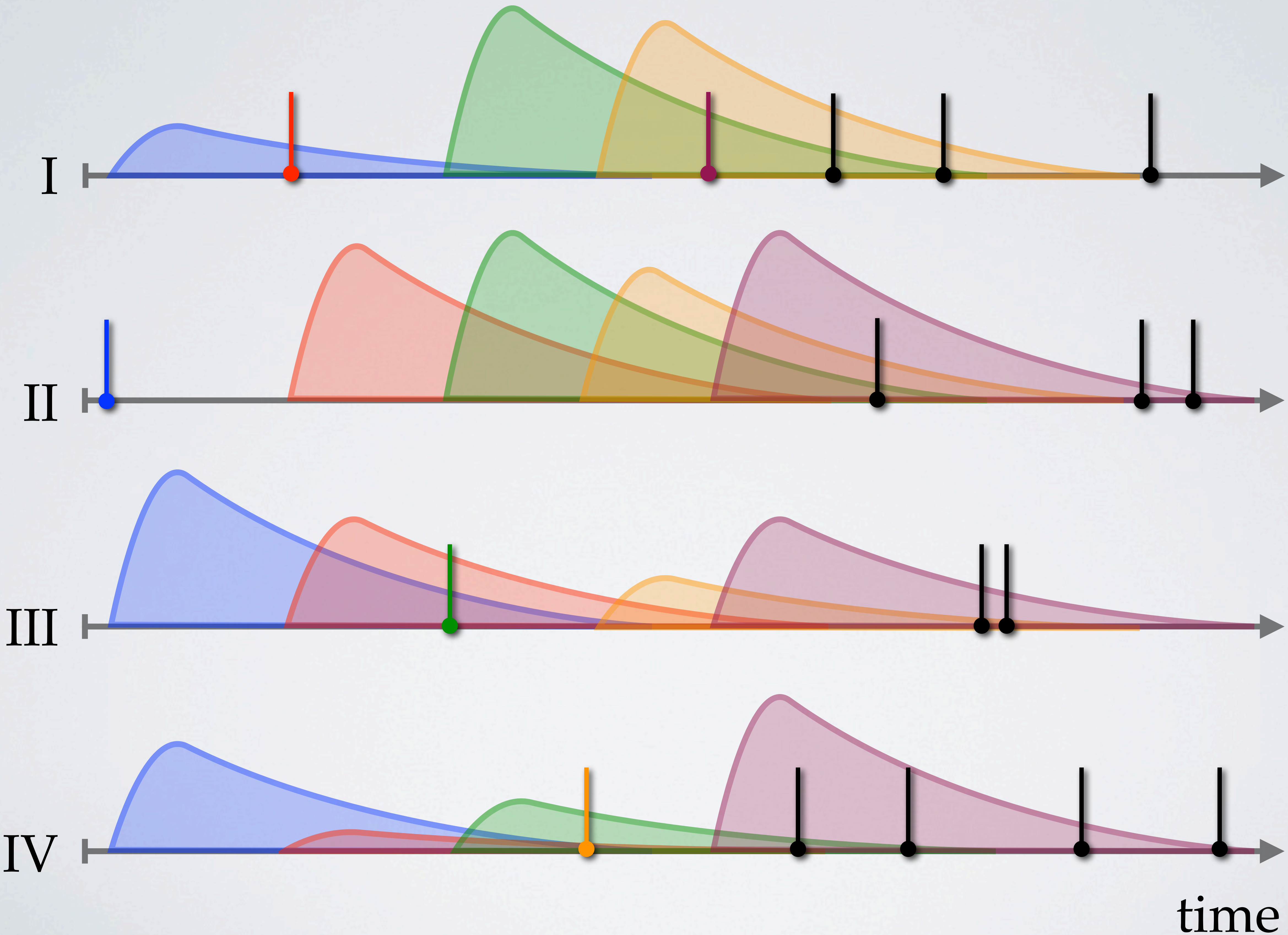
Hawkes Process Dynamics



Hawkes Process Dynamics



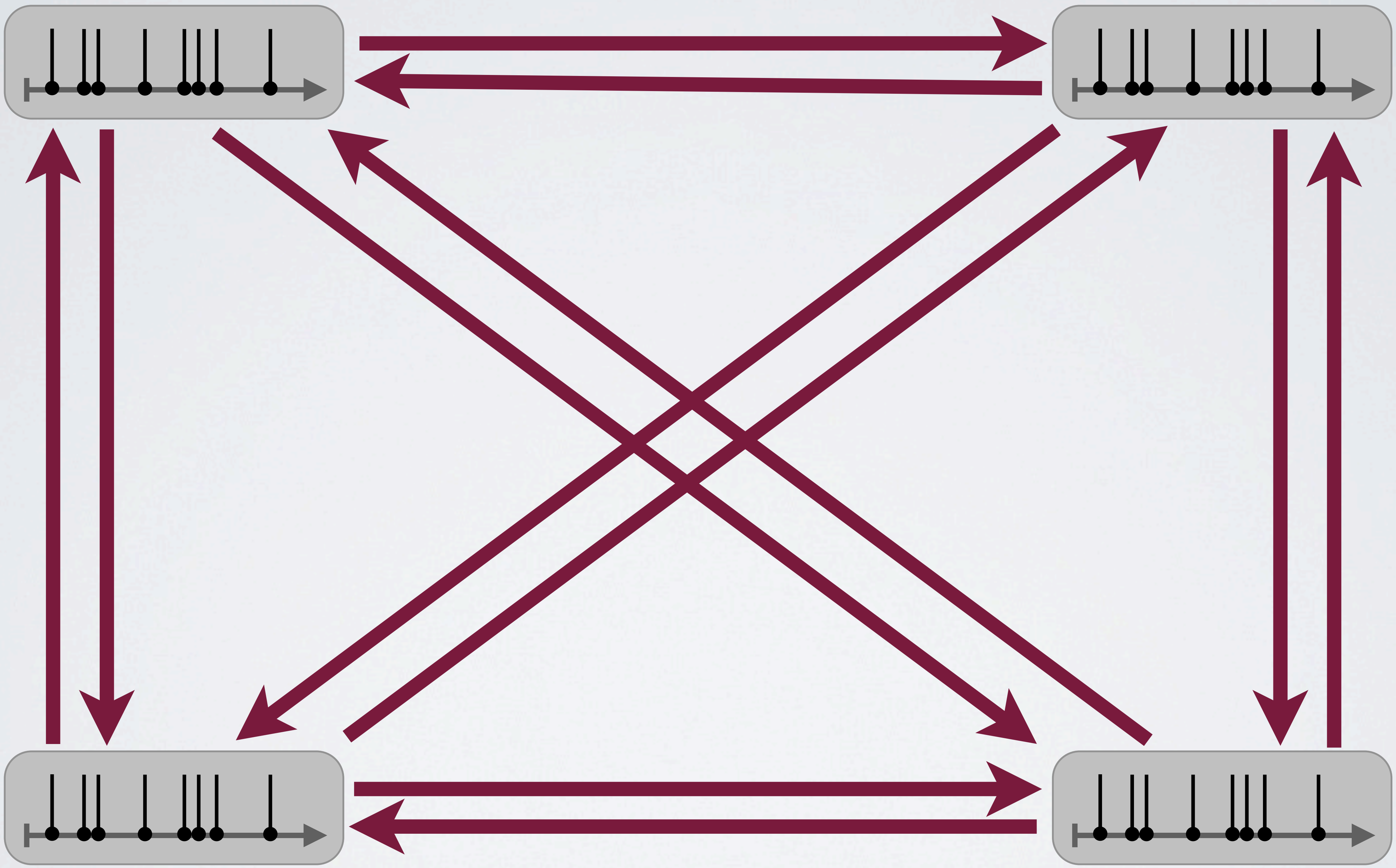
HAWKES PROCESS DYNAMICS



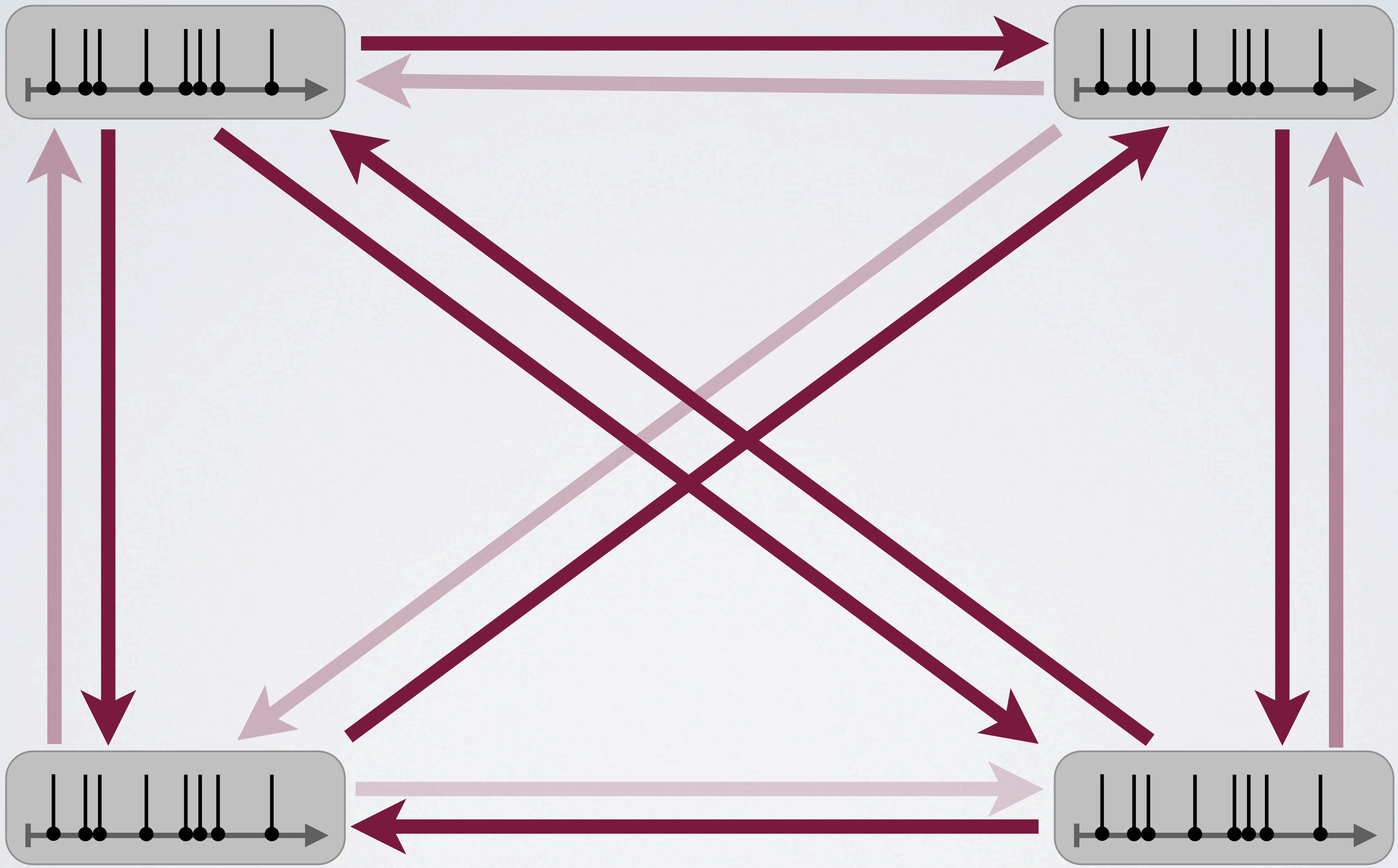
HAWKES PROCESS

- The Hawkes process specifies conditional Poisson dynamics in causal cascades.
  - Hawkes - Biometrika (1971), J. RSS-B (1971)
  - Hawkes - Stochastic Point Processes (1972)
  - Hawkes & Oakes - J. Applied Prob. (1974)
- 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.

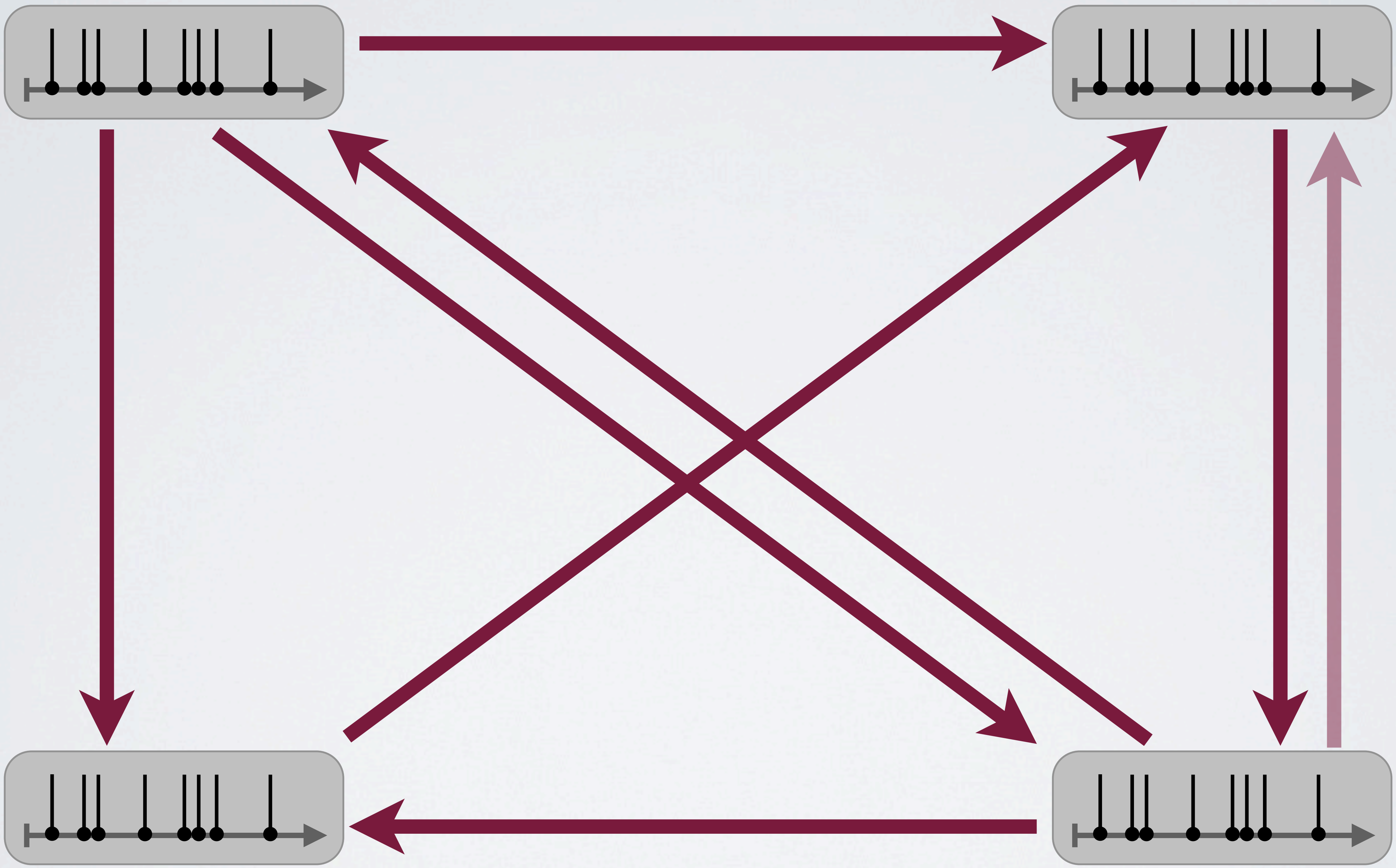
Graph Structure from Hawkes Dynamics



Graph Structure from Hawkes Dynamics



GRAPH STRUCTURE FROM HAWKES DYNAMICS



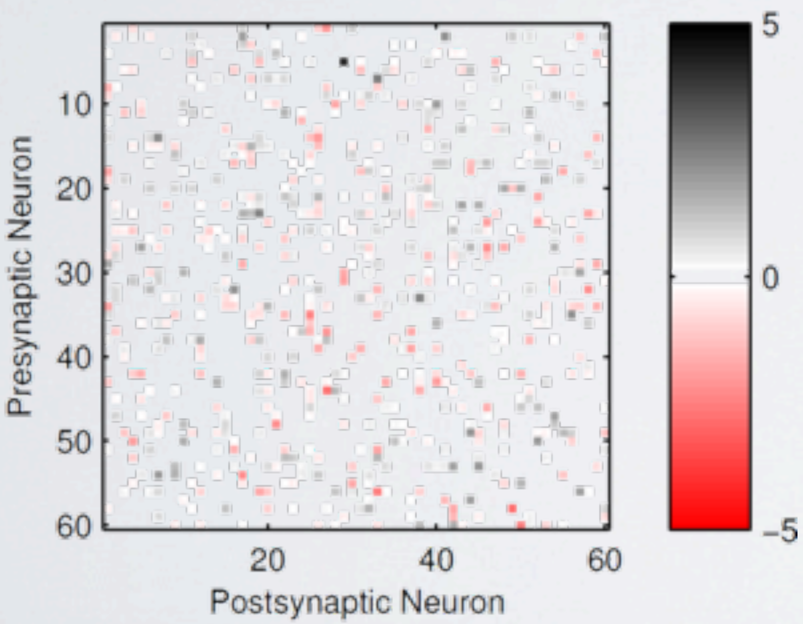
Bayesian Inference

- 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.

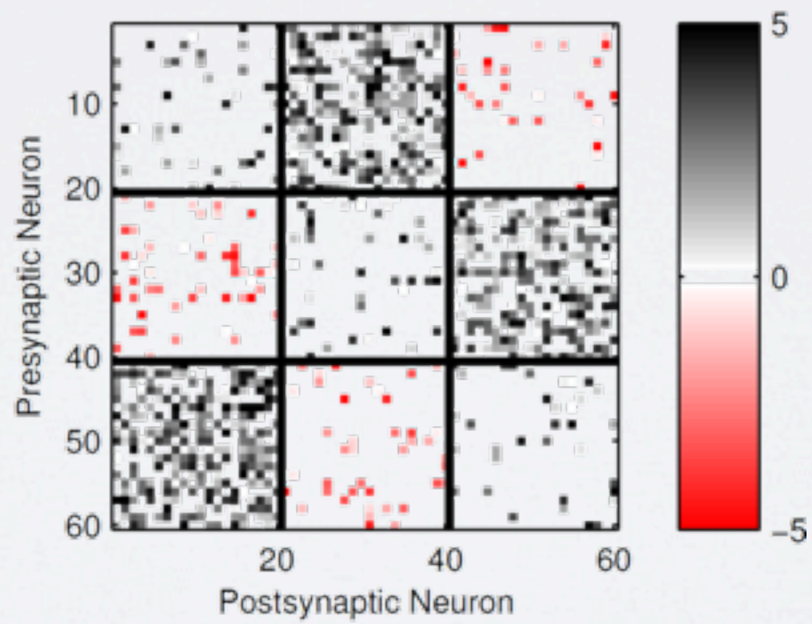
Exchangeable Random Graphs

- Recent work on exchangeable random graphs provides a rich set of priors for underlying networks, e.g., Diaconis & Janson (2007), Orbanz & Roy (2013).

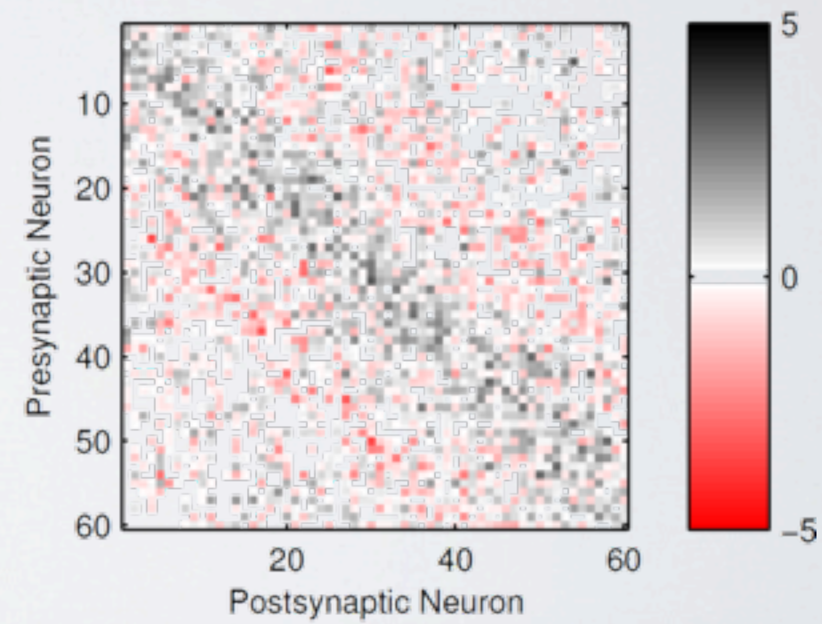
- Aldous-Hoover unifies many existing graph models:



Erdös-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 $\int_0^\infty g_\theta(t)\,\mathrm{d}t = 1$.

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

- 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 $\boldsymbol{Z} \in \{0,1\}^{N \times N}$ to represent these latent variables, where $Z_{n,n'} = 1$ means that event n induced event n'.

DATA-AUGMENTED LIKELIHOOD

$$p(\{s_n, c_n\}_{n=1}^{N}, \boldsymbol{Z} \mid \boldsymbol{A}, \boldsymbol{W}, \{\lambda_k^0(t)\}_{k=1}^{K}, \theta) =$$

$$\prod_{k=1}^{K} \exp\left\{-\int_0^T \lambda_k^0(\tau)\,\mathrm{d}\tau\right\} \prod_{n=1}^{N} \lambda_k^0(s_n)^{\mathbb{I}(c_n=k)\mathbb{I}(1-\sum_{n'} Z_{n',n})}$$

$$\times \prod_{k'=1}^{K} \left\{-\int_{s_n}^{T} A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n)\,\mathrm{d}\tau\right\}$$

$$\times \prod_{n'=1}^{N} \left(A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n)\right)^{Z_{n,n'}}$$

DATA-AUGMENTED LIKELIHOOD

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

Poisson process likelihood for events from the background process.

Data-Augmented Likelihood

$$p(\{s_n, c_n\}_{n=1}^{N}, \boldsymbol{Z} \mid \boldsymbol{A}, \boldsymbol{W}, \{\lambda_k^0(t)\}_{k=1}^{K}, \theta) =$$

$$\prod_{k=1}^{K} \exp\left\{-\int_0^T \lambda_k^0(\tau)\,\mathrm{d}\tau\right\} \prod_{n=1}^{N} \lambda_k^0(s_n)^{\mathbb{I}(c_n=k)\mathbb{I}(1-\sum_{n'} Z_{n',n})}$$

$$\boxed{\begin{aligned} &\times \prod_{k'=1}^{K} \left\{-\int_{s_n}^{T} A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n)\,\mathrm{d}\tau\right\} \\ &\times \prod_{n'=1}^{N} \left(A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n)\right)^{Z_{n,n'}} \end{aligned}}$$

Poisson process likelihood for events induced by previous events.

DATA-AUGMENTED LIKELIHOOD

$$p(\{s_n, c_n\}_{n=1}^N, \boldsymbol{Z} \mid \boldsymbol{A}, \boldsymbol{W}, \{\lambda_k^0(t)\}_{k=1}^K, \theta) =$$

$$\boxed{\prod_{k=1}^K \exp\left\{-\int_0^T \lambda_k^0(\tau)\,\mathrm{d}\tau\right\}} \prod_{n=1}^N \lambda_k^0(s_n)^{\mathbb{I}(c_n=k)\mathbb{I}(1-\sum_{n'} Z_{n',n})}$$

$$\boxed{\times \prod_{k'=1}^K \left\{-\int_{s_n}^T A_{c_n,k'} W_{c_n,k'} g_\theta(\tau - s_n)\,\mathrm{d}\tau\right\}}$$

$$\times \prod_{n'=1}^N \left(A_{c_n,k'} W_{c_n,k'} g_\theta(s_{n'} - s_n)\right)^{Z_{n,n'}}$$

Ugly looking, but just normalization constants.

MODELING DETAILS

- 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.*

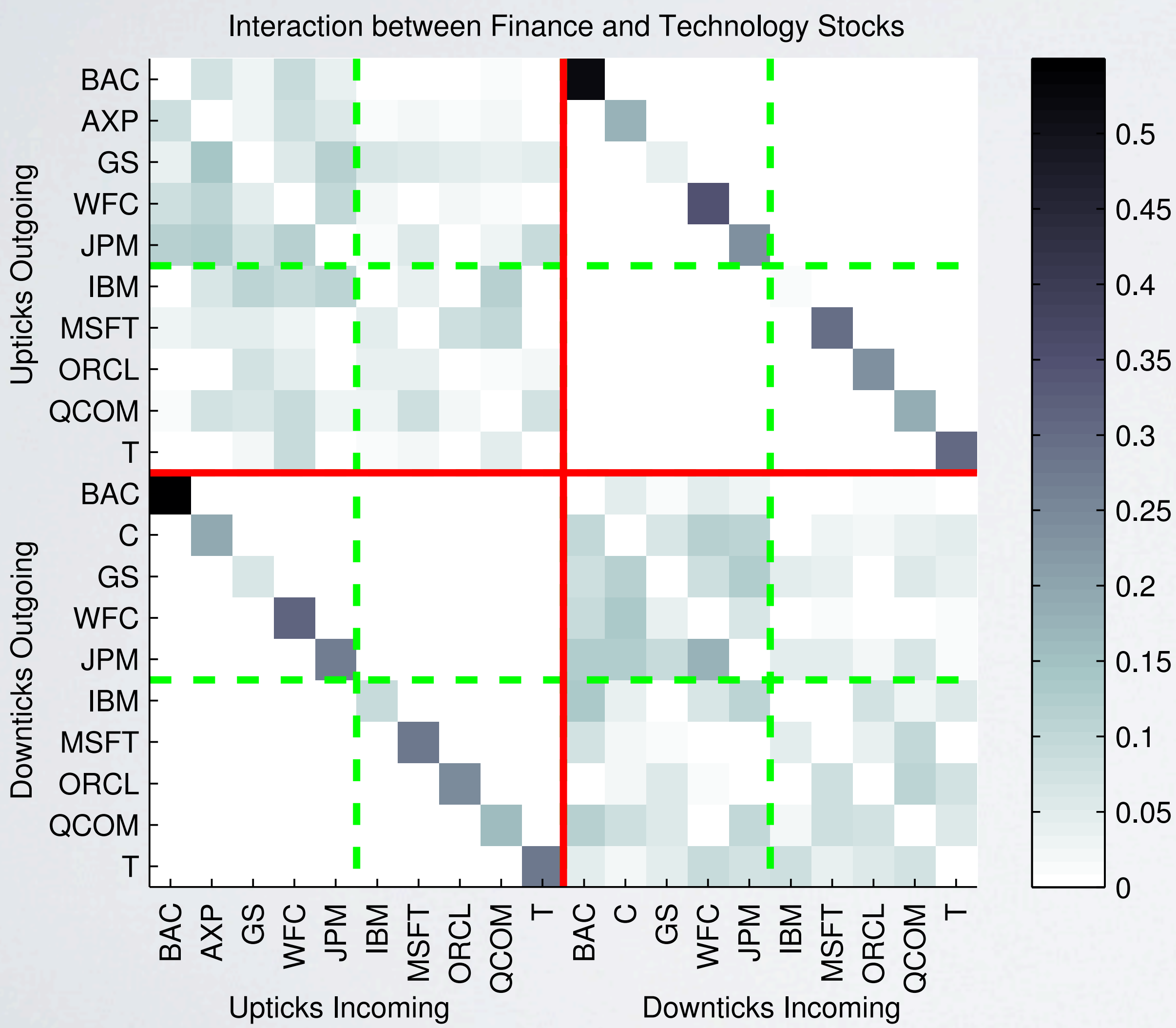
INFERENCE WITH MCMC

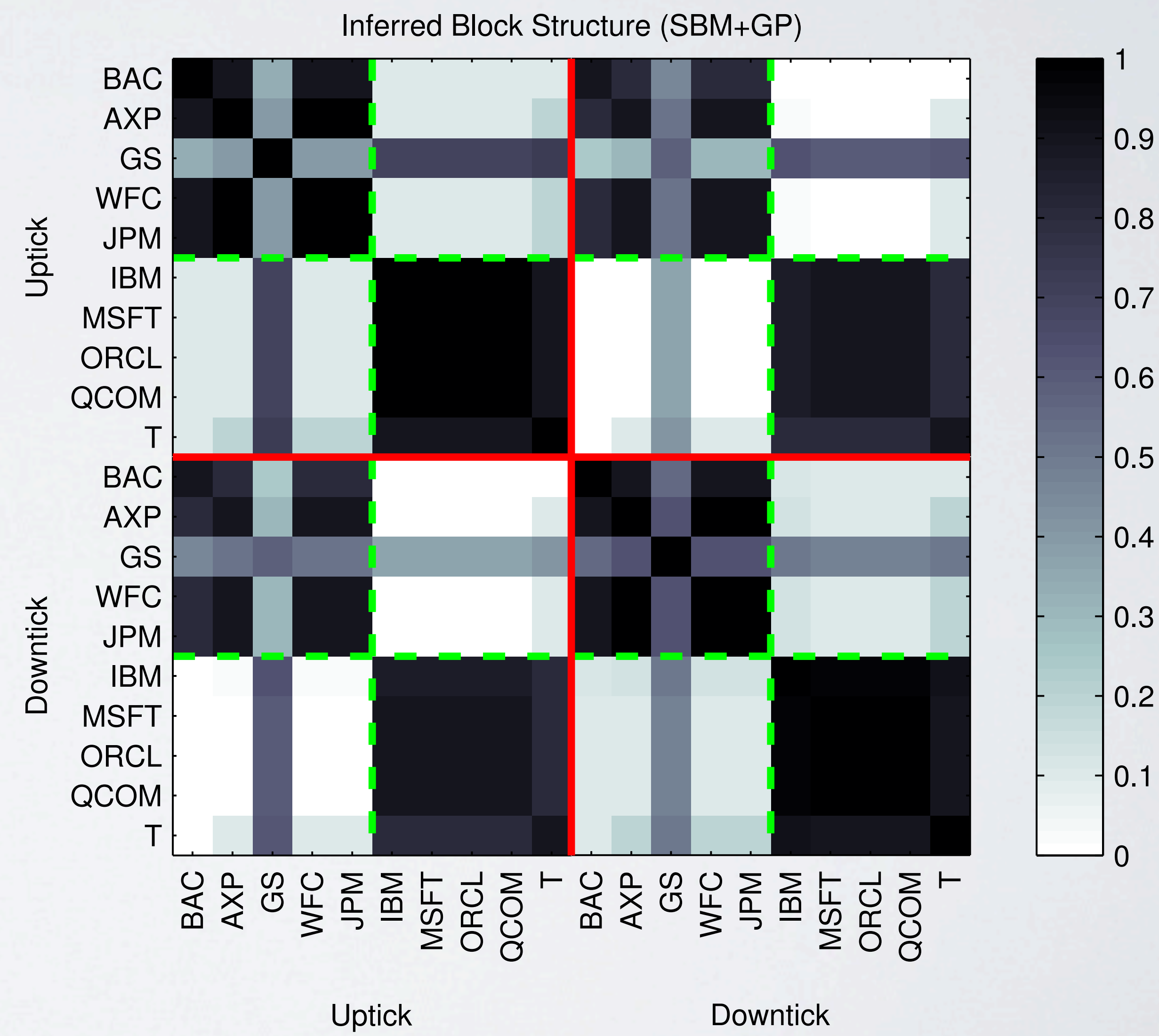
- Graph structure: collapsed block Gibbs
- Edge weights: Gibbs (conjugate gamma posterior)
- Latent parent explanations: parallel Gibbs
- Background rates: elliptical slice sampling
- Temporal kernels: Gibbs or slice sampling

- 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.



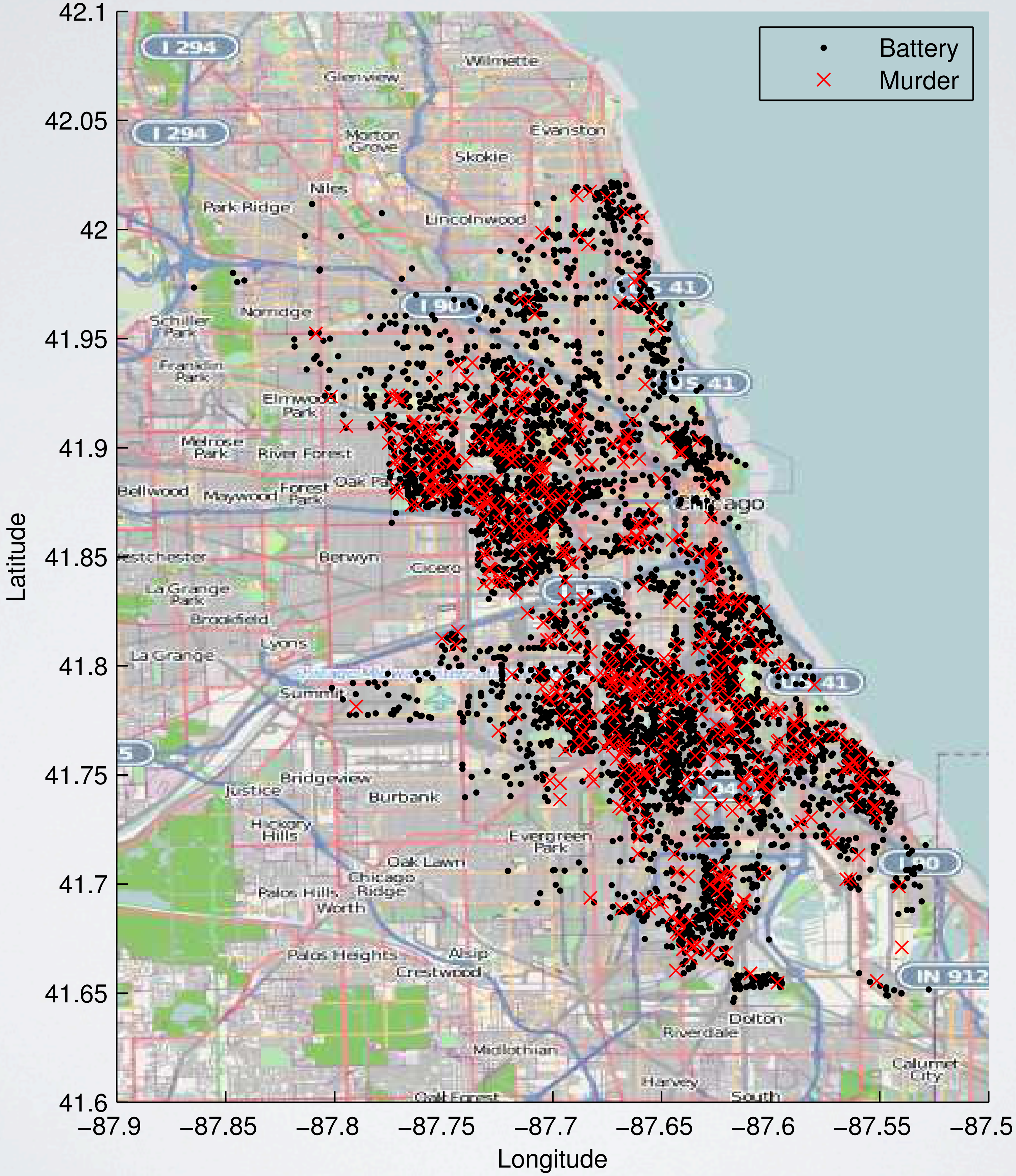


Weighted Edges

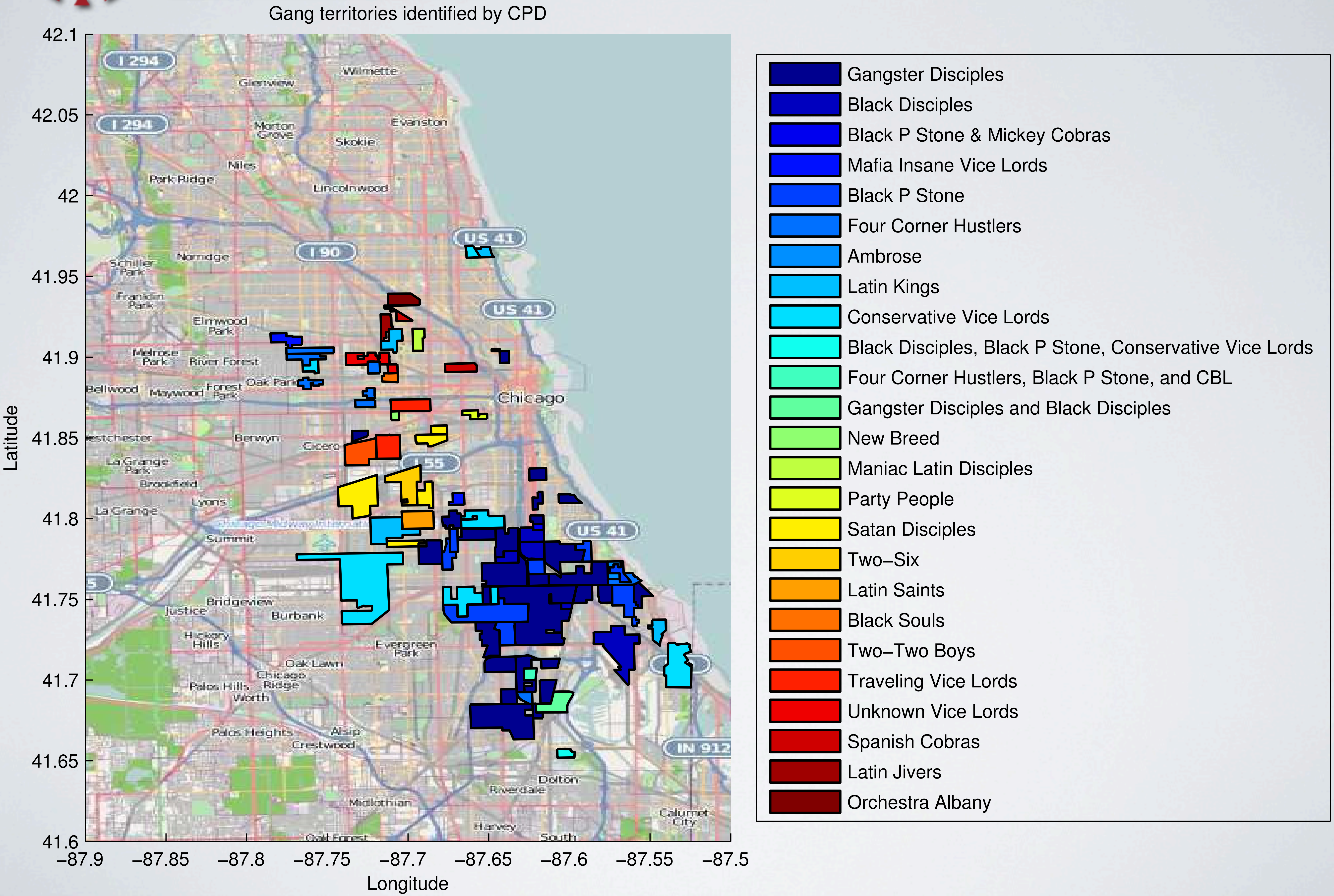
Block Membership

Homicides and Armed Batteries in Chicago

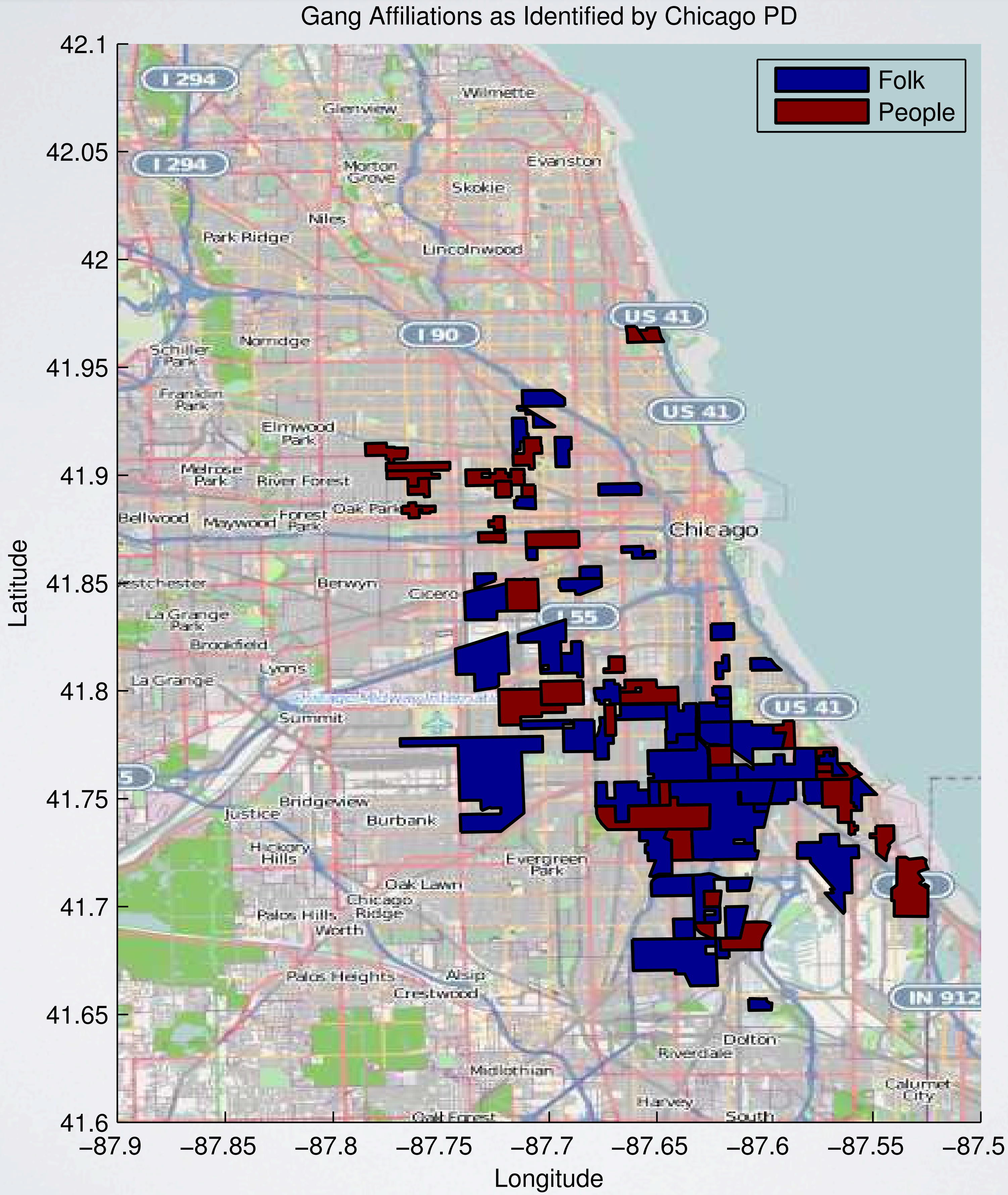
Murder and Battery in Chicago (2001–2012)



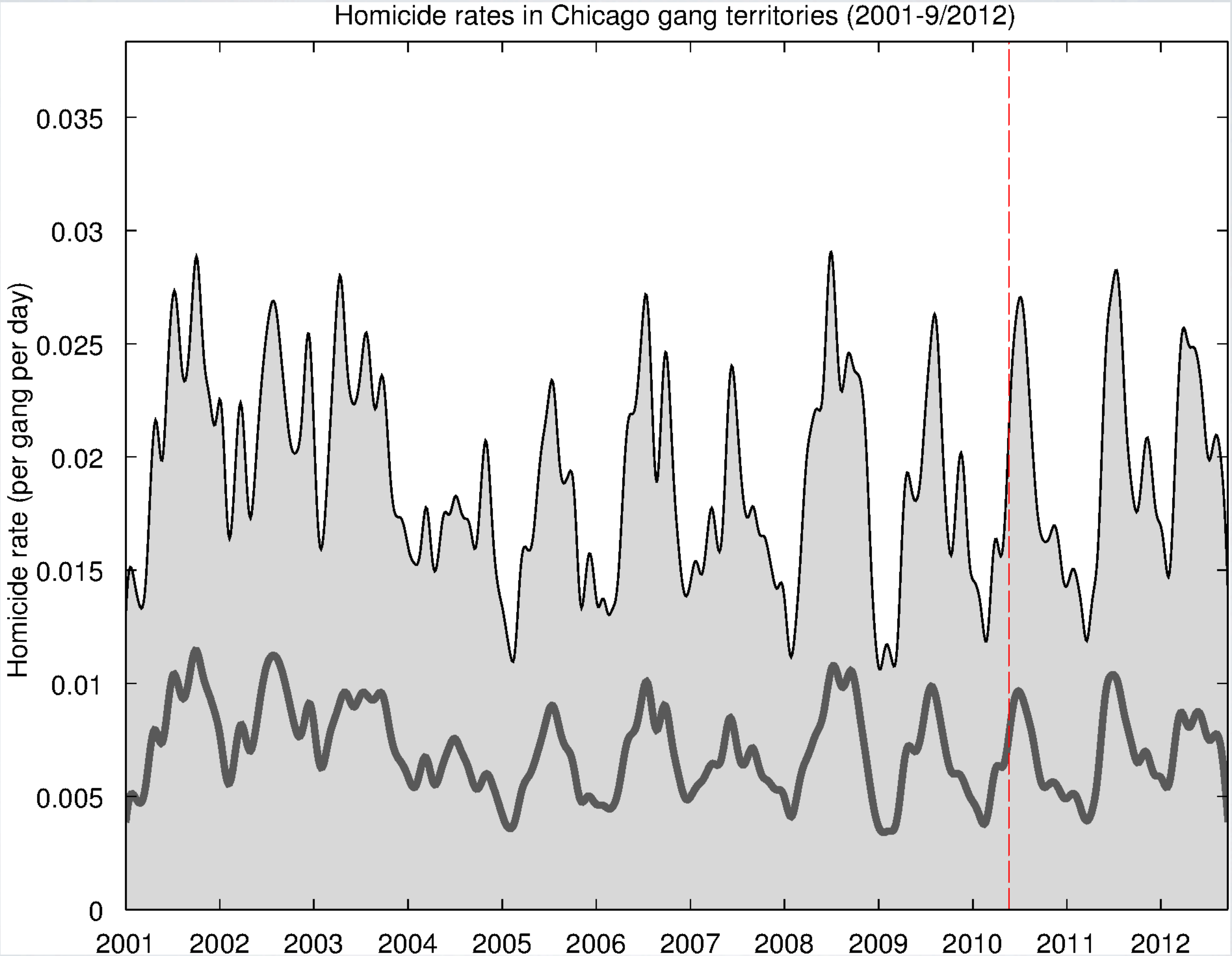
CPD-IDENTIFIED GANNG BOUNDARIES



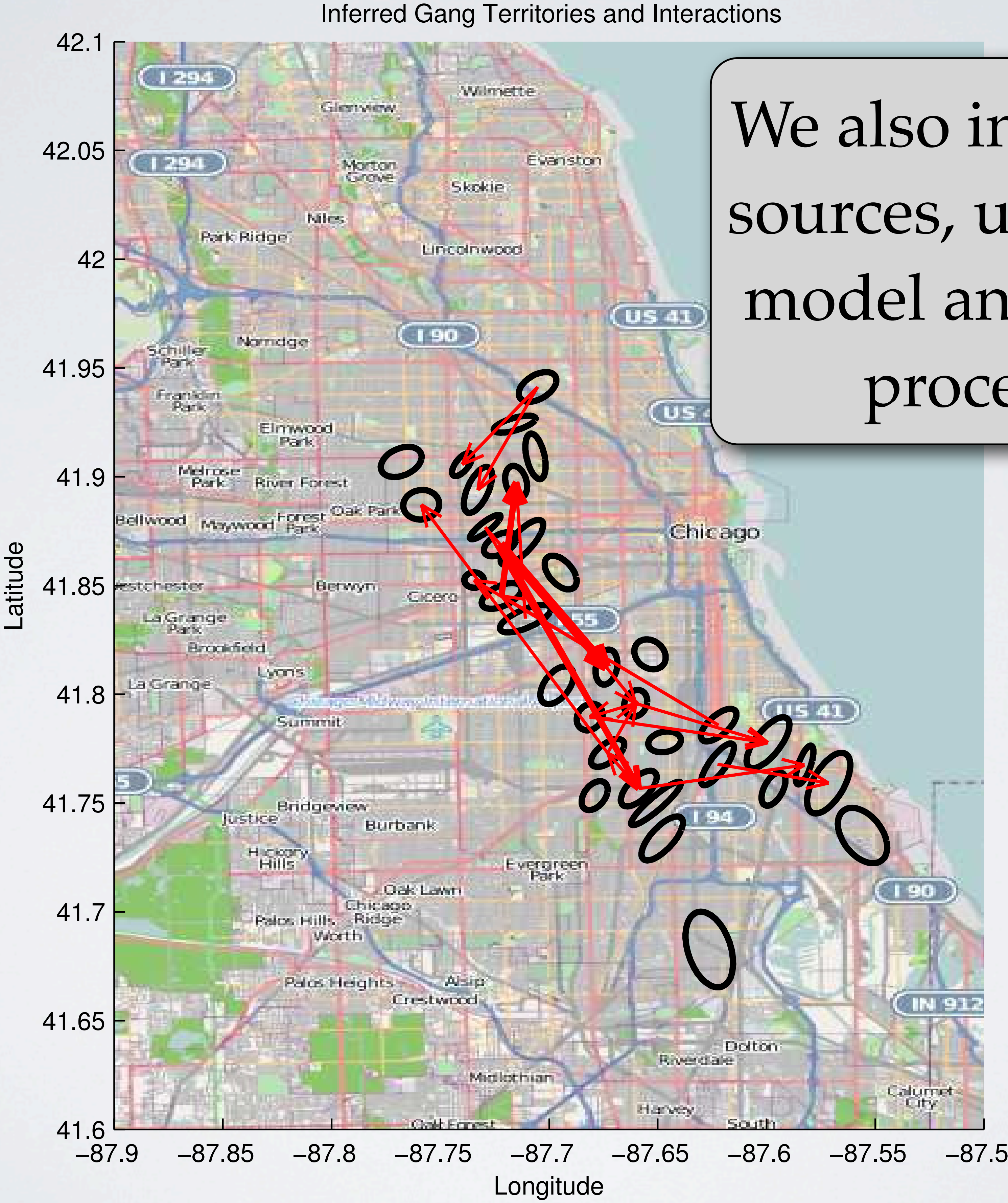
Alliances: Folk Nation vs. People Nation



SEASONAL VARIATION

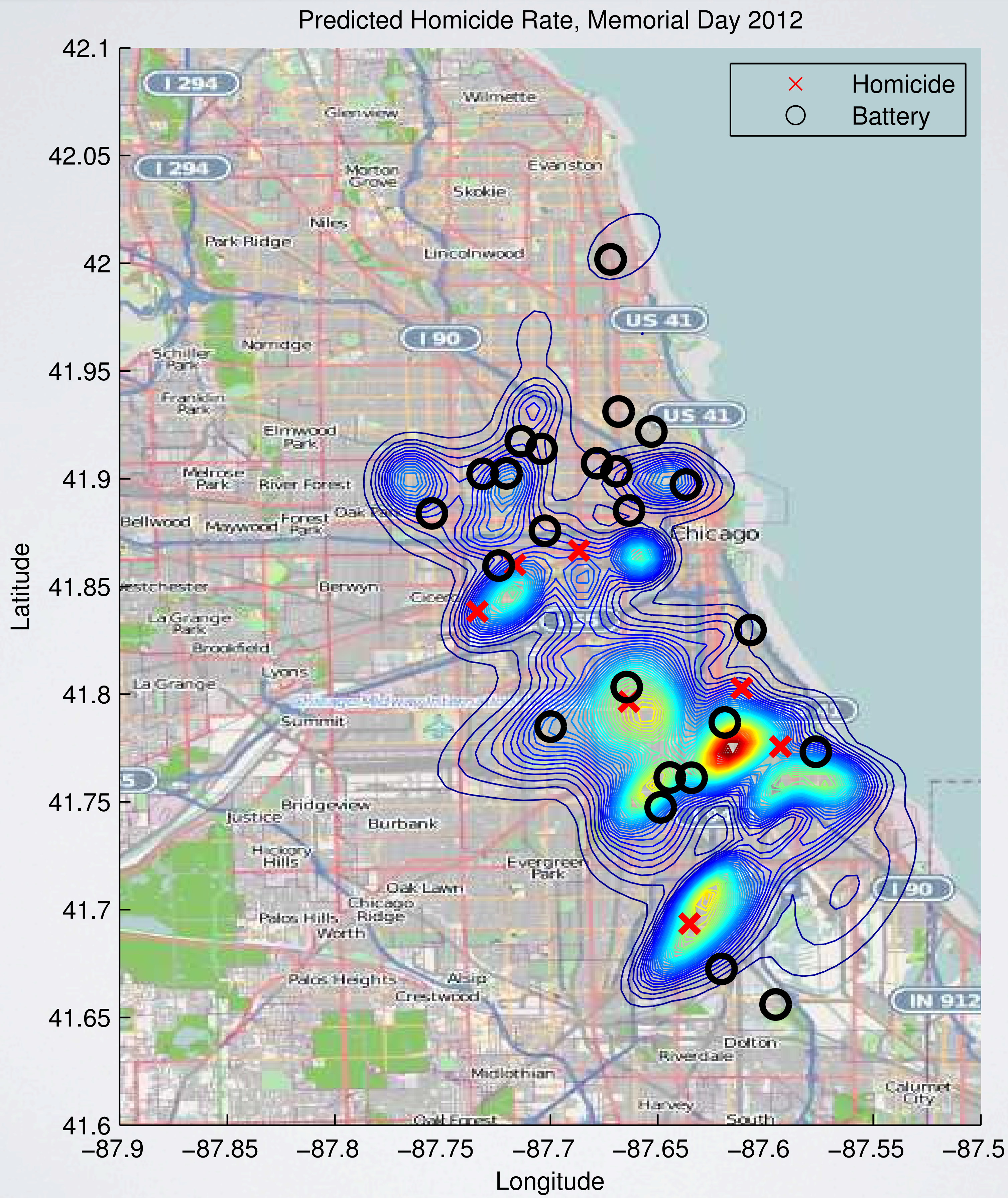


INFERRED INTERACTING TERRITORIES



We also infer the event sources, using a spatial model and a Dirichlet process prior.

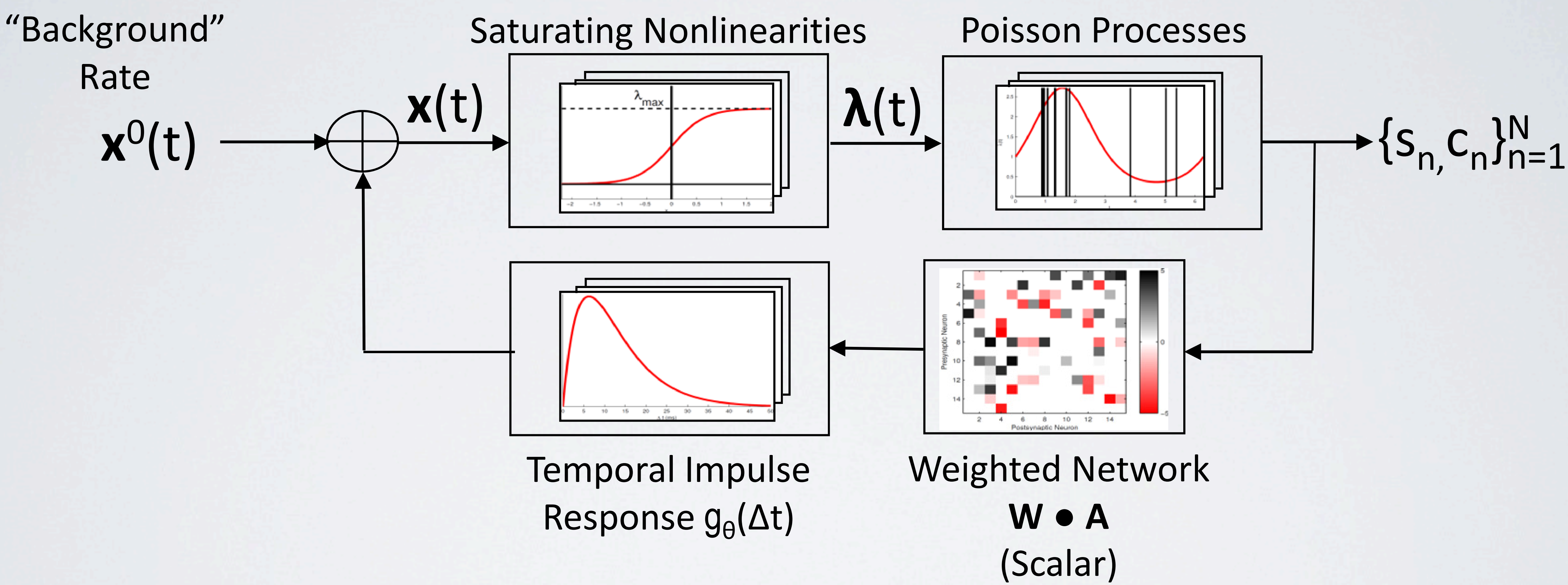
Trying to Predict Hotspots



NEURAL MODELING WITH INHIBITION

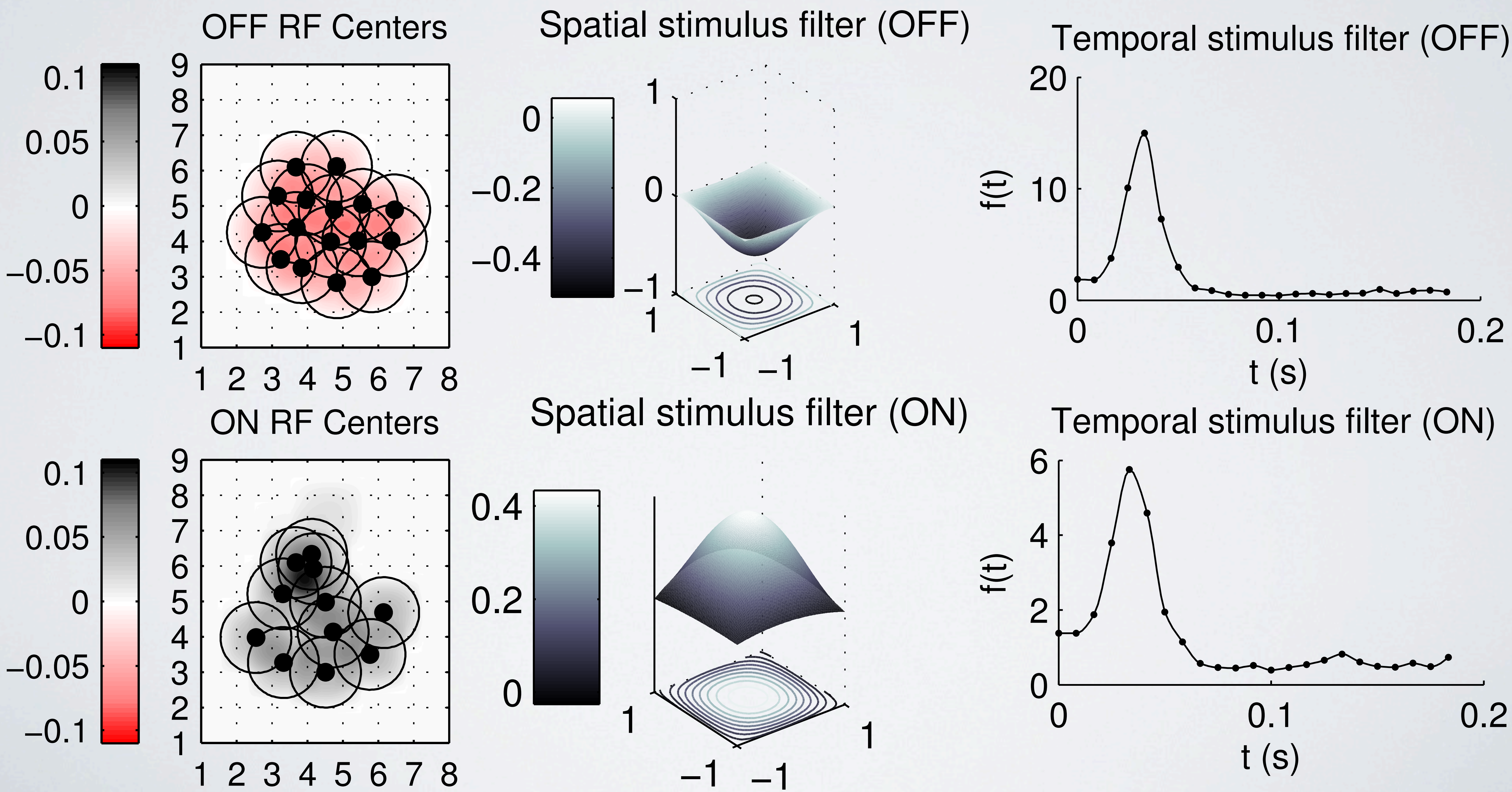
- Neural functional connectivity involves both **excitation** and **inhibition**, so the pure Hawkes is inappropriate.
- We extend the model to allow for negative weights and include a saturating nonlinearity.
- This effective becomes a Bayesian variant of the popular generalized linear model (GLM) from the computational neuroscience literature.
- We can leverage graph priors within this framework to discover latent neural properties and connectivity.

BAYESIAN GLM FOR NEURAL DATA

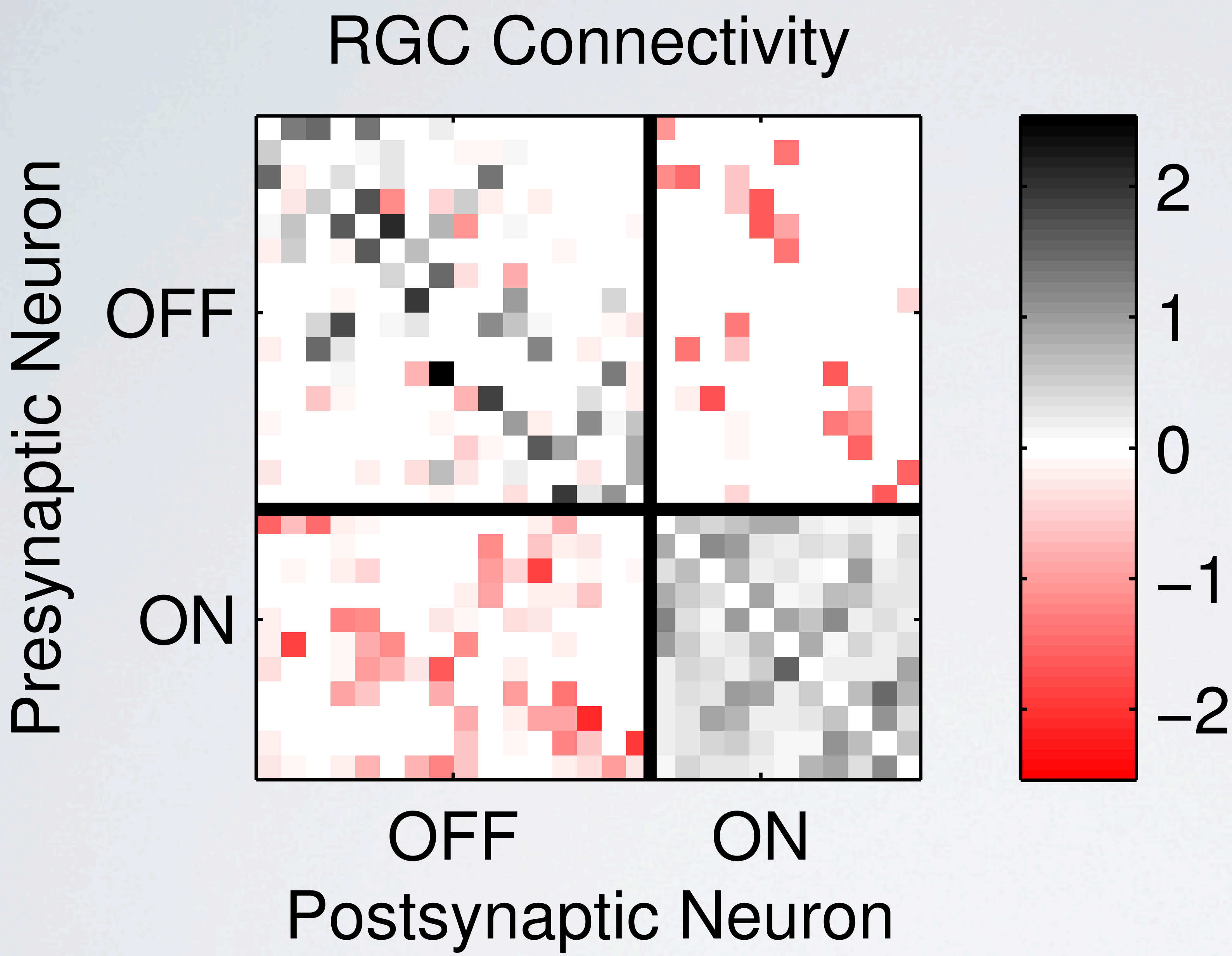


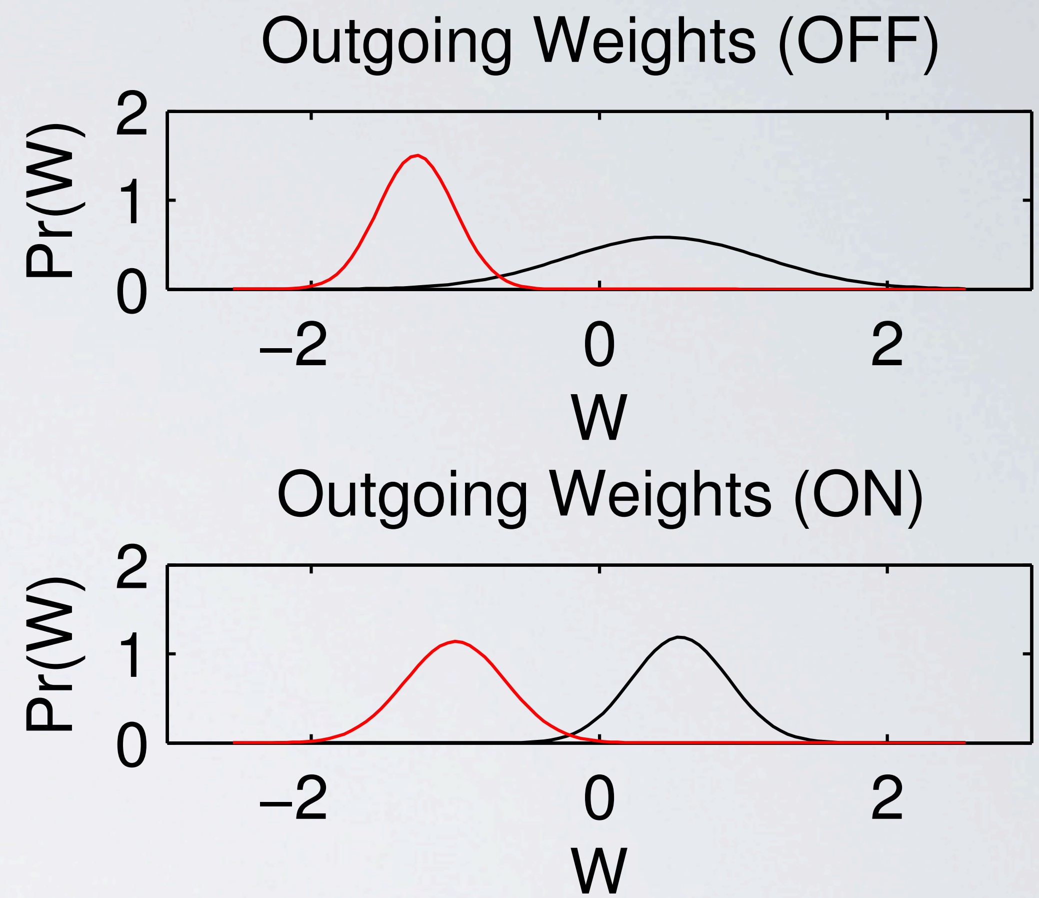
Retinal Ganglion Cells

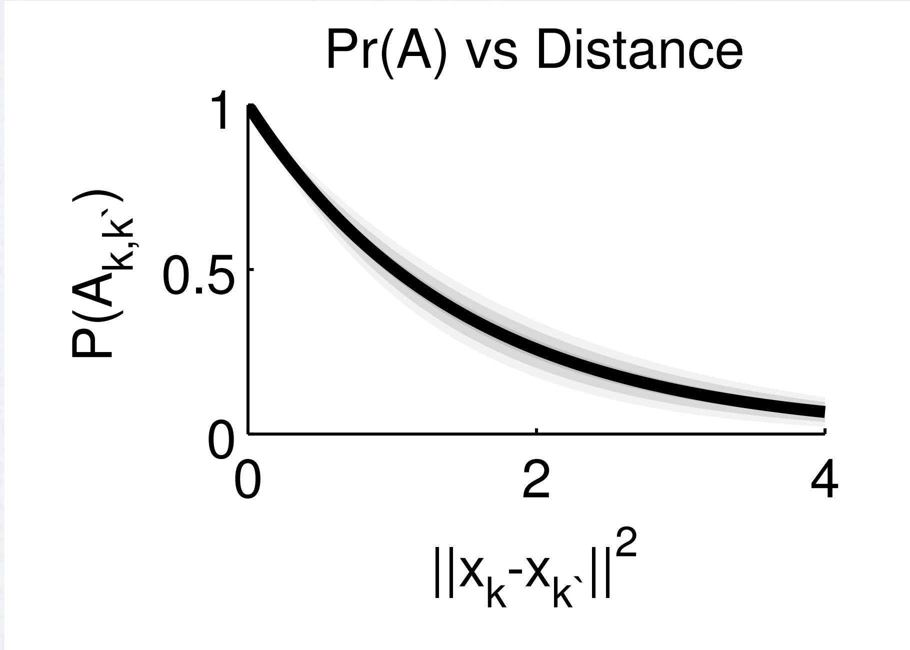
27 neurons, 50K spikes, macaque retina
(Data from Pillow and Chichilnisky)



Inferred Network Properties







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.

Scott Linderman