# Dynamic Process over Networks:
## Representation, Modeling, Learning & Inference

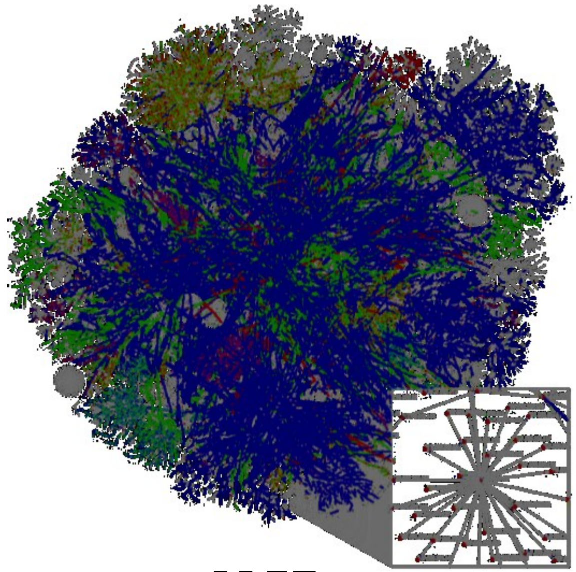
Le Song

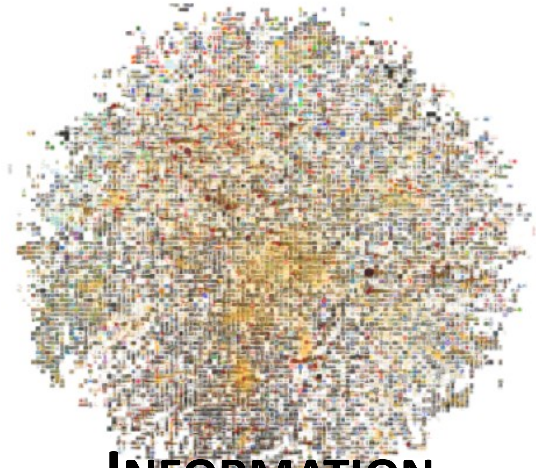
CSE, College of Computing

Georgia Institute of Technology

# Networks are everywhere

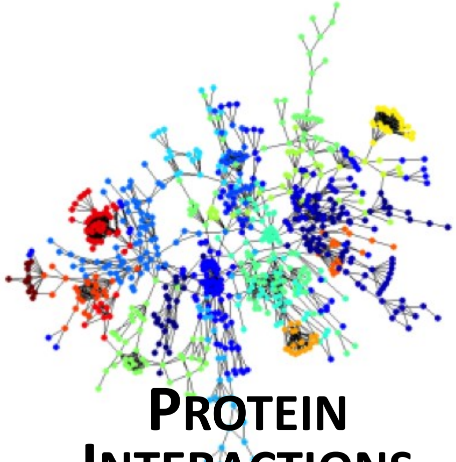
SOCIAL NETWORKS



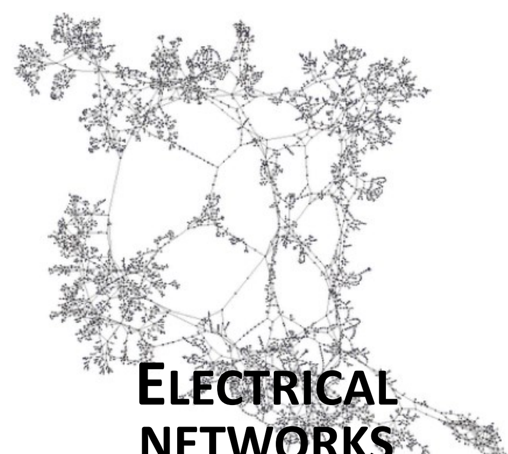


INFORMATION NETWORKS

TRANSPORTATION NETWORKS



PROTEIN INTERACTIONS



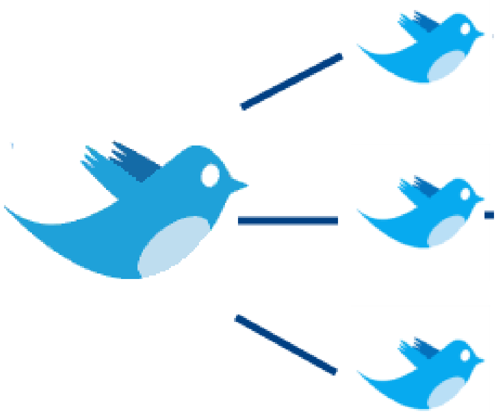
ELECTRICAL NETWORKS


Mostly discrete events in continuous time

Dynamics are essential

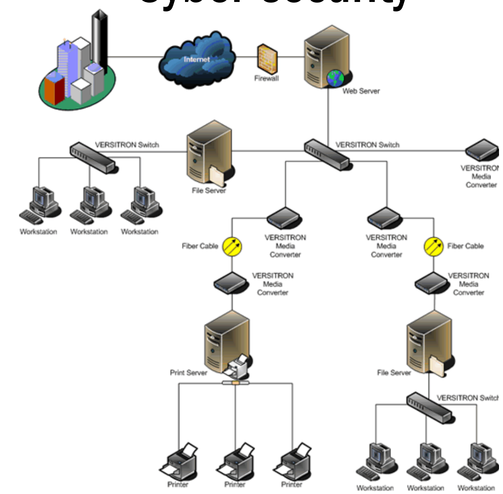
# Dynamics are essential to many applications

## Information spread



## Epidemiology

## Cyber-security




## Healthcare analytics


## Smart city

## Sustainability problems

# Networks for a purpose

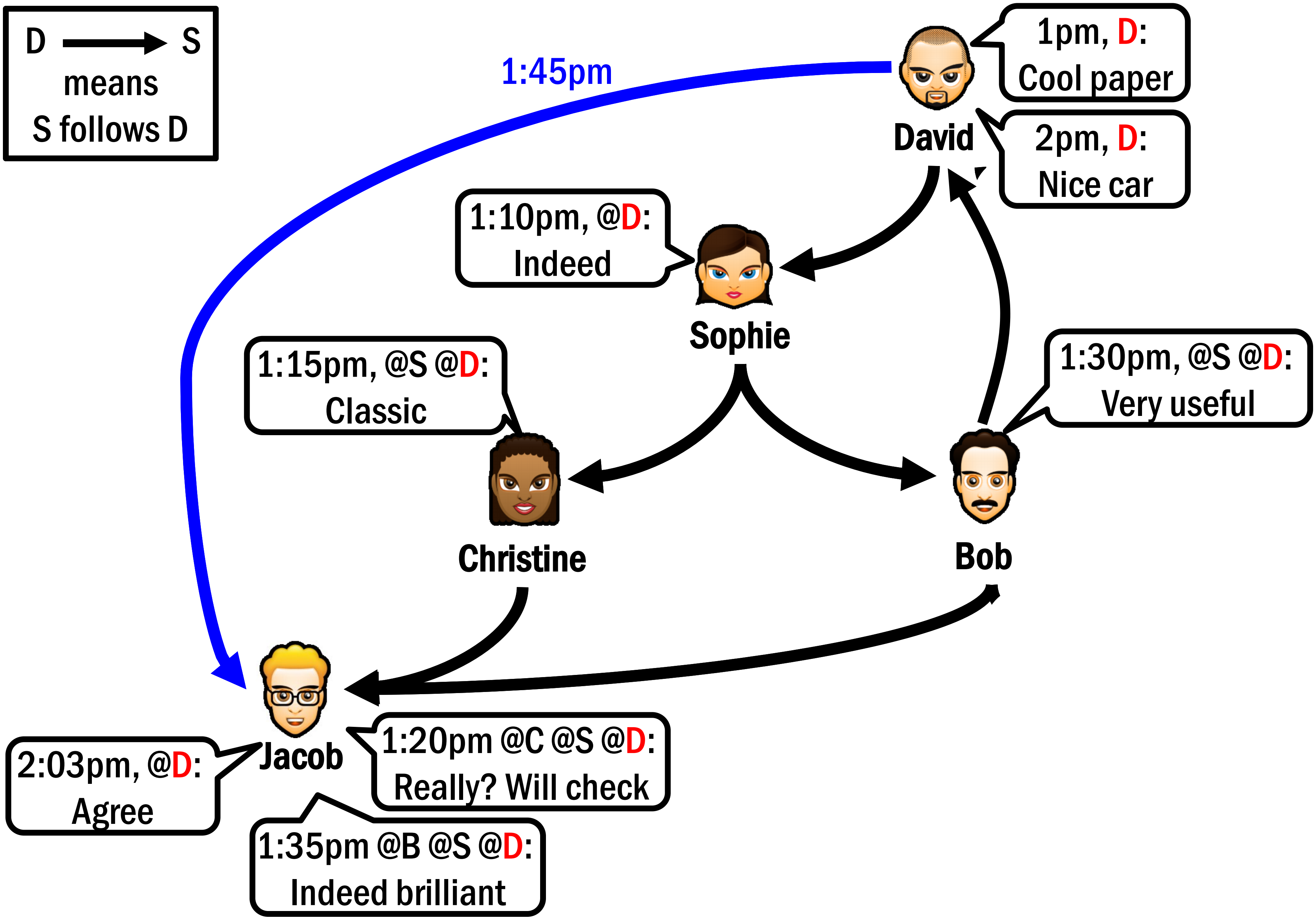
People follow others to receive interesting information

People becomes friends to share joys of life

People follow topics to acquire or contribute knowledge

People link to each other to find job opportunities

# A running example: coevoluation

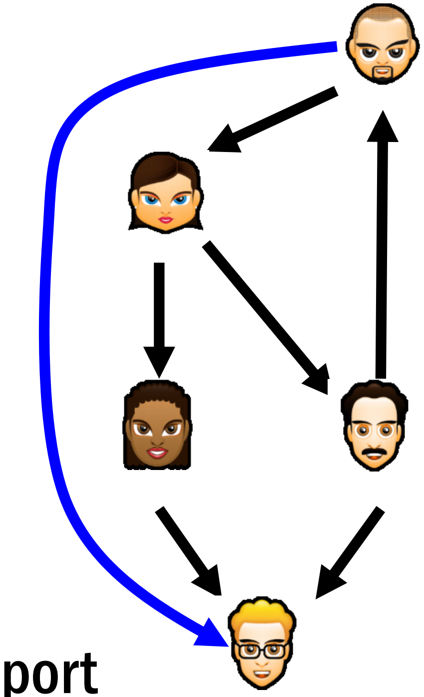


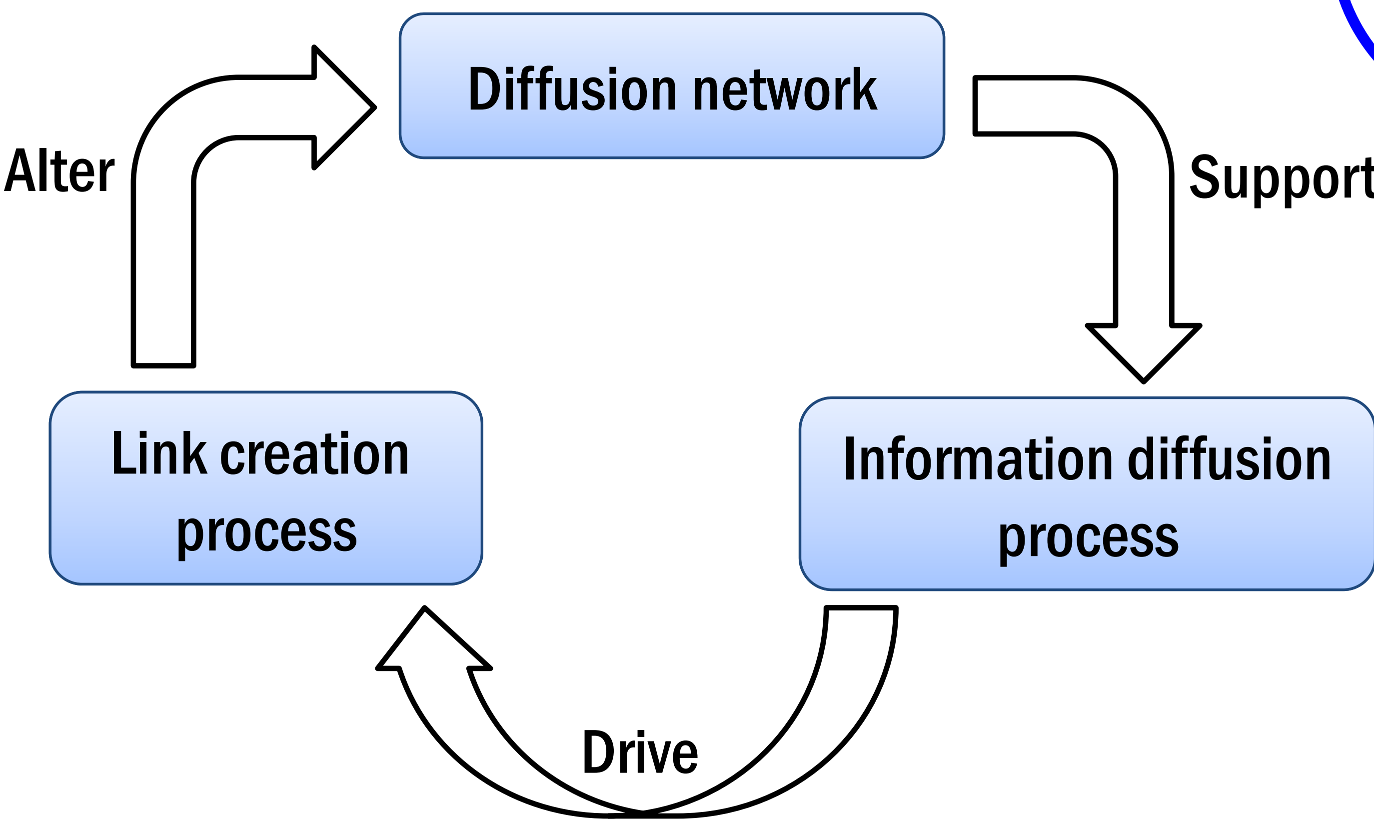
# Two interacting processes

Information diffusion over the network

Link creation driven by information diffusion

New link alters diffusion paths





# Previous network models

## Lots of network structure & network evolution models:

- Small world [Watts & Strogatz ‘98]
- Bowtie [Broder et al. ‘00]
- Preferential Attachment [Barabasi & Albert ‘99]
- Kronecker graphs [Leskovec et al. ‘10]

**Joint models of information diffusion and network coevolution missing!**

## Lots of information diffusion models:

- Discrete time independent cascade [Kemp et al. ’03]
- Continuous time independent cascade [Du et al. ‘13]
- Hawkes process [Zhou et al. ‘13]

## Empirical studies of effects of diffusion on network structure

- Traffic-based shortcut [Weng  et al. ‘13]
- Tweet-Retweet-Follow [Antoniades & Dovrolis, ‘13]
- Bursty dynamics [Myers & Leskovec ‘14]

# Coevolve:

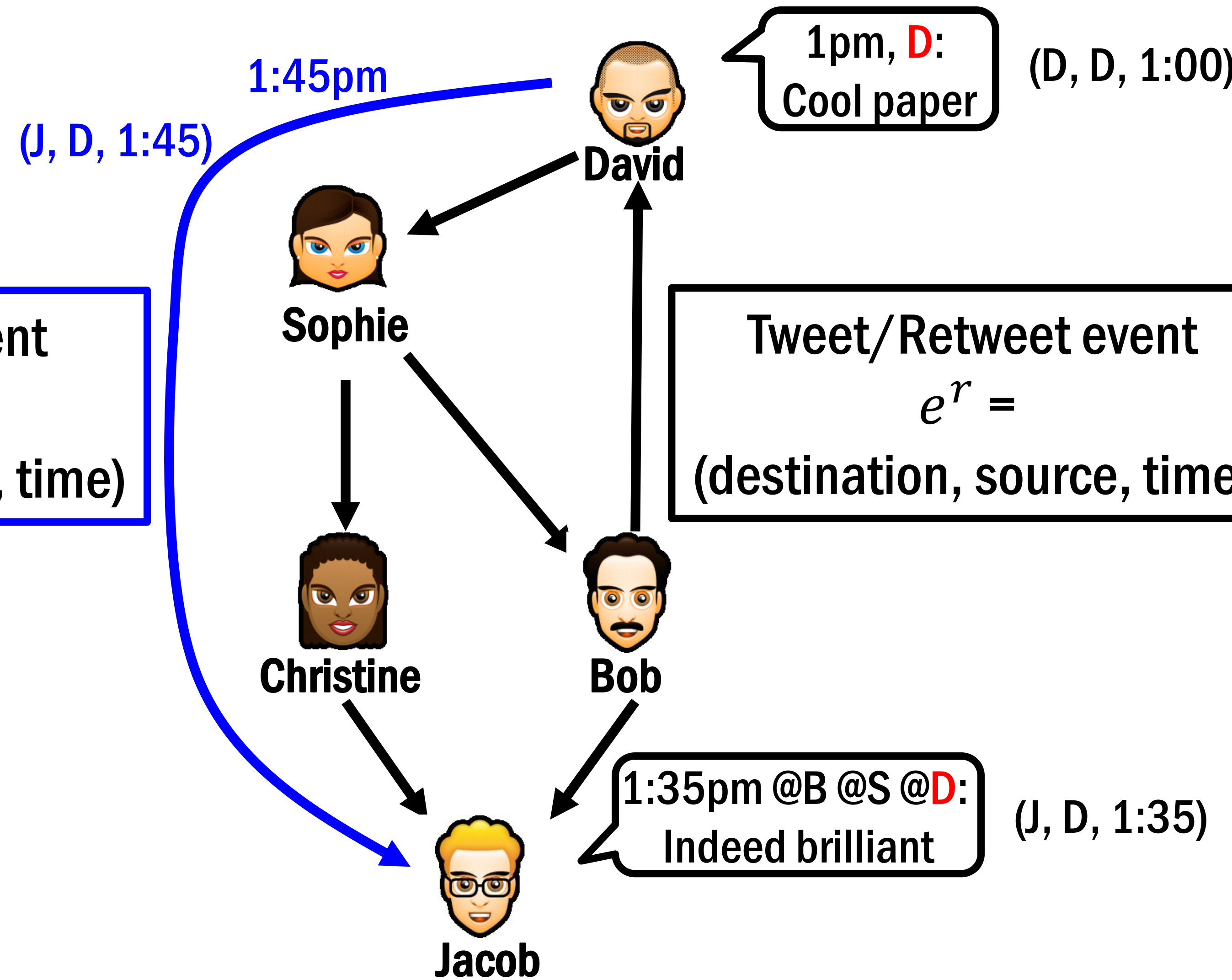
## A Model of Information Diffusion and Network Evolution

**Model Formulation** → Efficient Estimation and Simulation → Model Properties → Real-Data Experiments

# Event representation

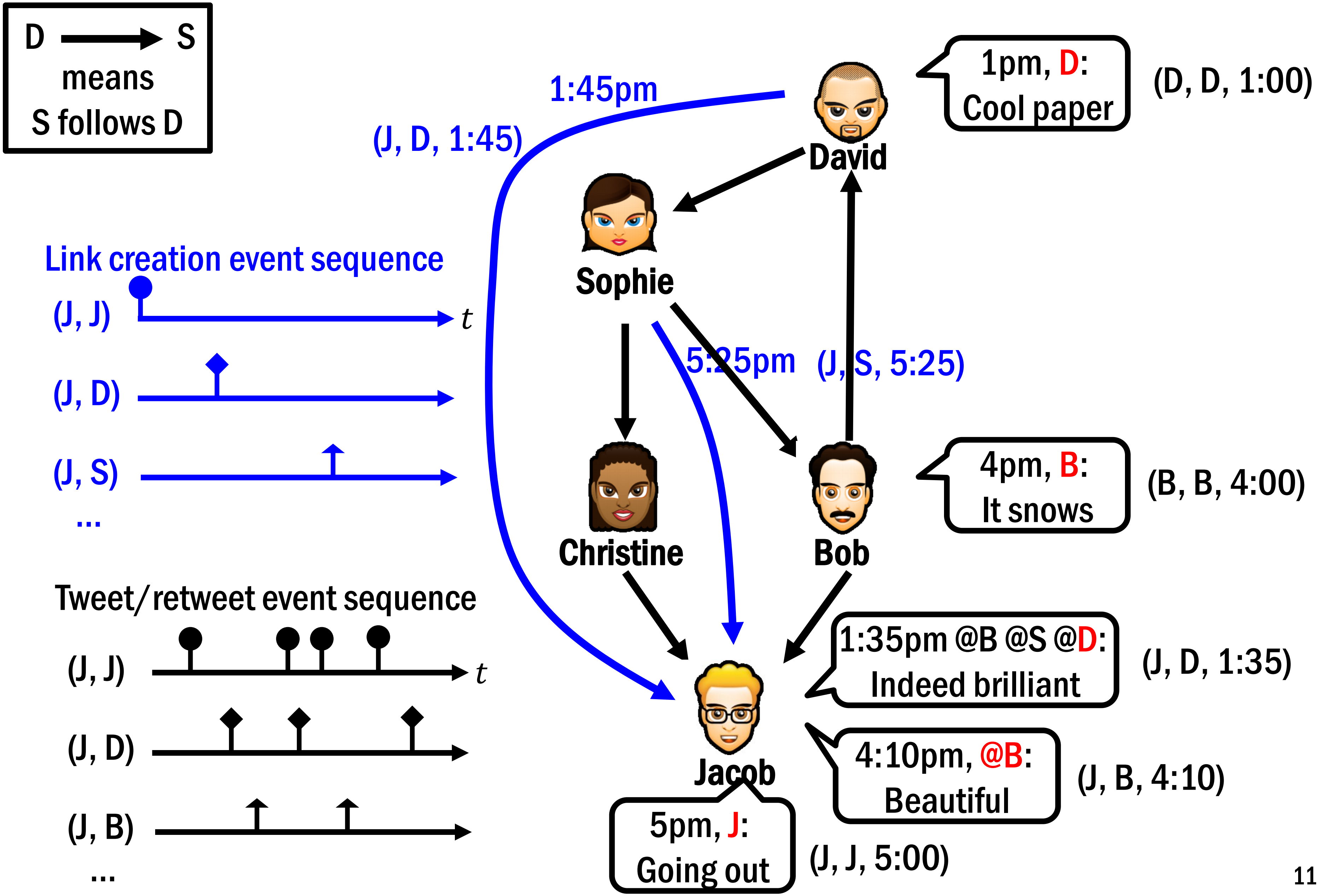
D → S means S follows D

Link creation event
$e^l$ =
(destination, source, time)




Tweet/Retweet event
$e^r$ =
(destination, source, time)

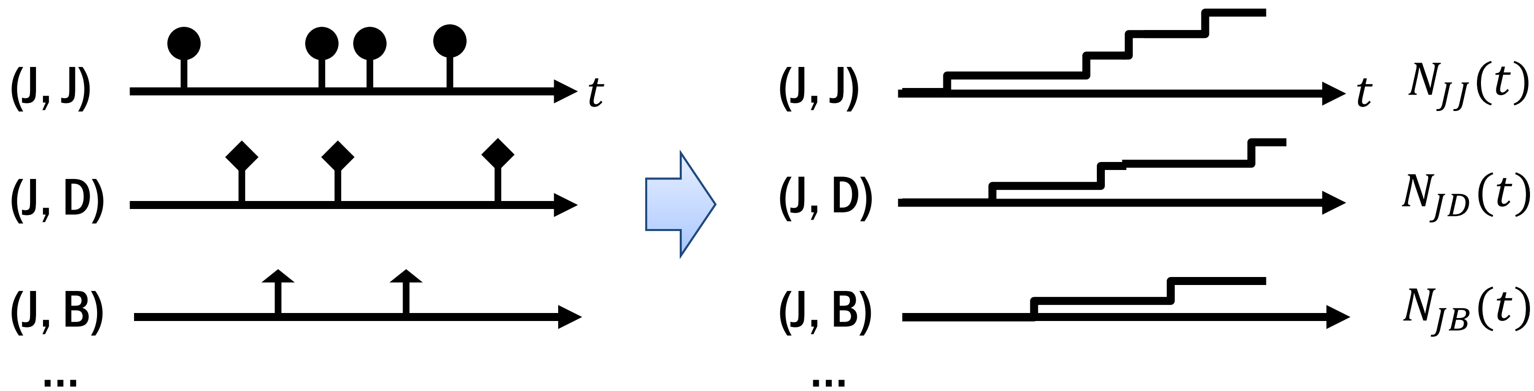
# Event sequence



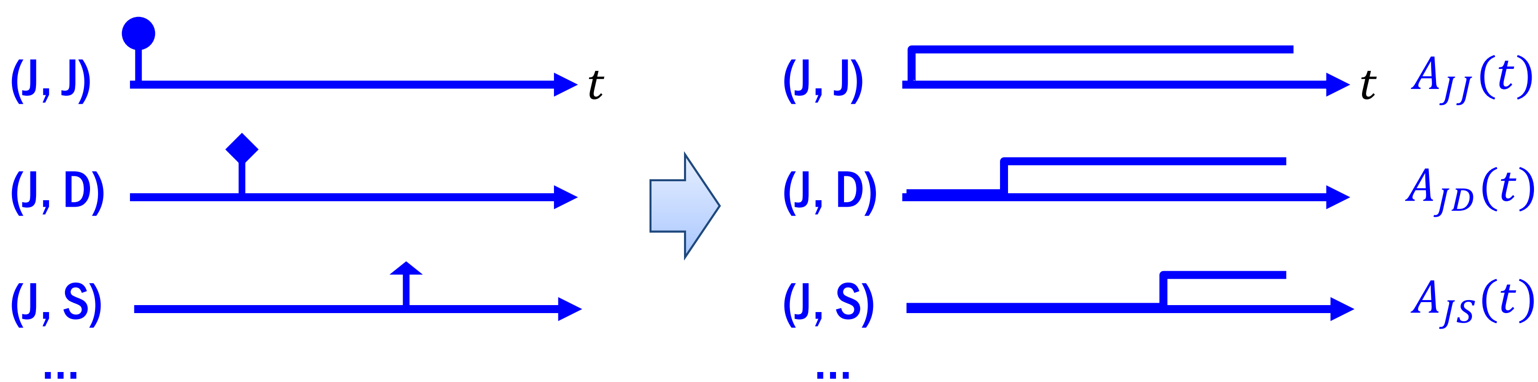
# Counting processes

For user J

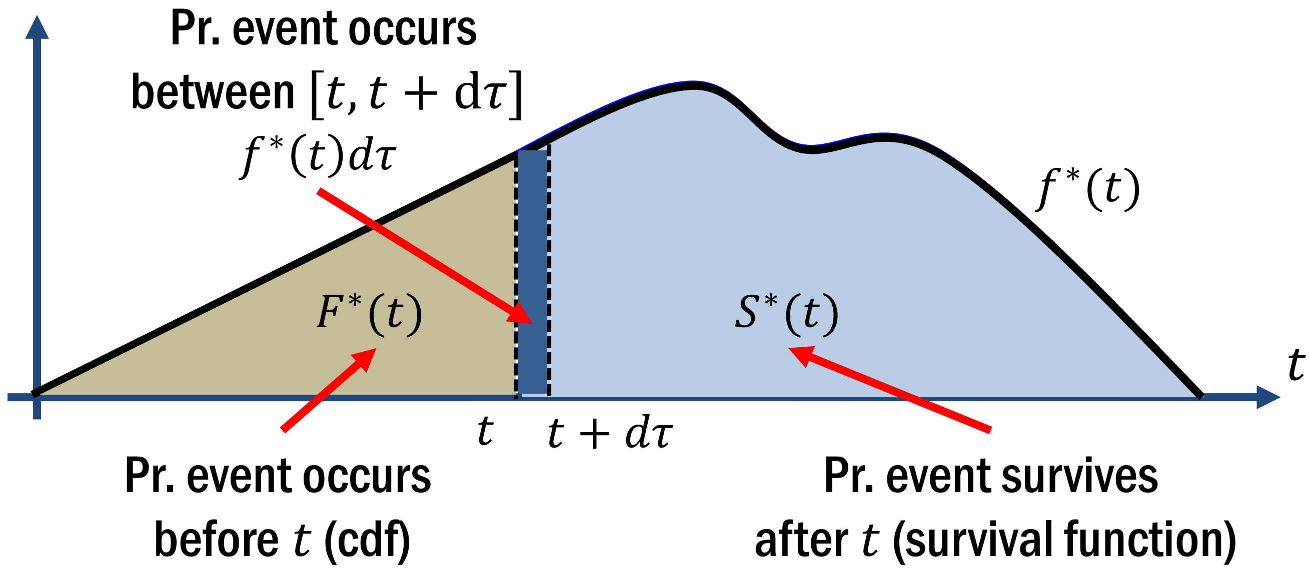
- “Identity Revealing” tweet/retweet processes $\boldsymbol{N}(t) \in \{0\} \cup Z^+$




- “Information driven” link creation processes $\boldsymbol{A}(t) \in \{0,1\}$



# Intensity describes a counting process




**Intensity: Pr. event occurs between $[t, t + \mathrm{d}\tau]$ but not before $t$**

$$h^*(t)d\tau = \frac{f^*(t)d\tau}{S^*(t)} > 0$$

**Relation to counting process**

$$N(t) = \int_0^t h^*(\tau)d\tau + M(t)$$

# Example dynamics

**Poisson process** ⇐ Information diffusion and link creation

$$h^*(t) = \mu$$


**Hawkes process (self exciting process)** ⇐ Information diffusion

$$h^*(t) = \mu + \alpha \sum_{t_i < t} \exp(-|t - t_i|) = \mu + \alpha \exp(-|t|) \star dN(t)$$


**Survival process** ⇐ Link creation

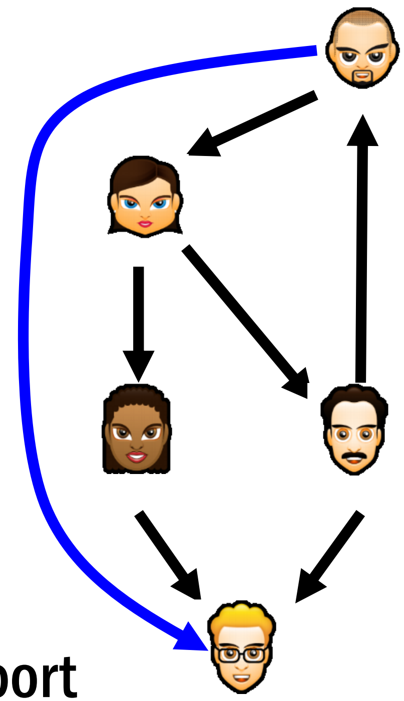
$$h^*(t) = \big(1 - N(t)\big)\, g^*(t)$$

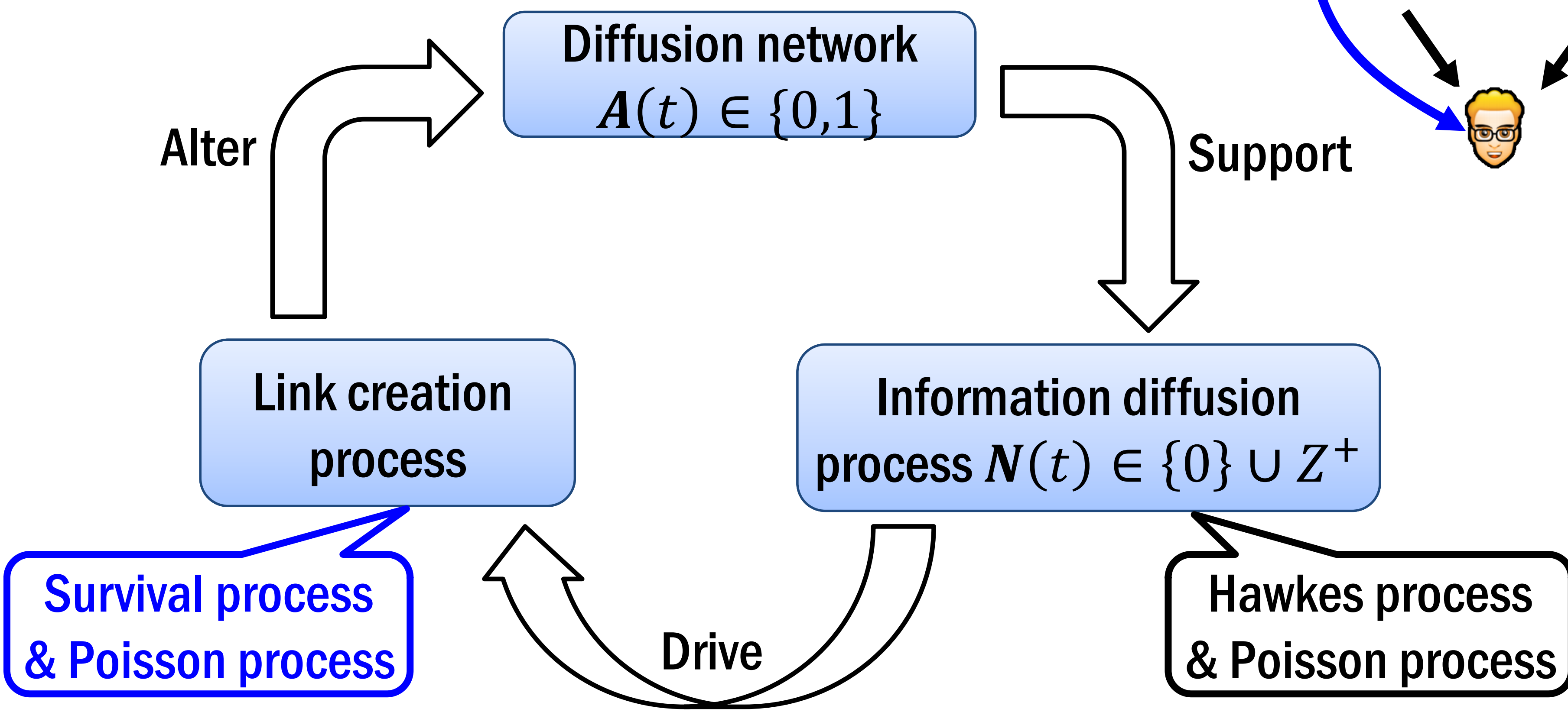
# Two interacting processes

Information diffusion over the network

Link creation driven by information diffusion

New link alters diffusion paths

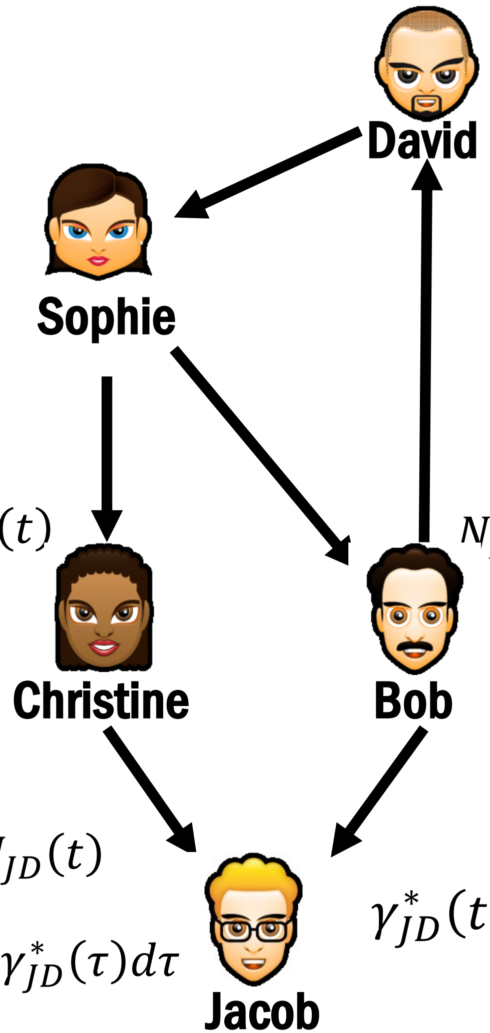




# Modeling information diffusion

Generate original tweet with Poisson

More exposures trigger more retweets




$N_{CD}(t)$ …

$N_{BD}(t)$ …

$$N_{DD}(t) = \int_0^t \gamma^*_{DD}(\tau)d\tau$$

$$\gamma^*_{DD}(t) = \overbrace{\eta_D}^{\substack{\text{Poisson process} \\ \text{User } D\text{'s own initiative}}}$$

$$N_{JD}(t) = \int_0^t \gamma^*_{JD}(\tau)d\tau$$

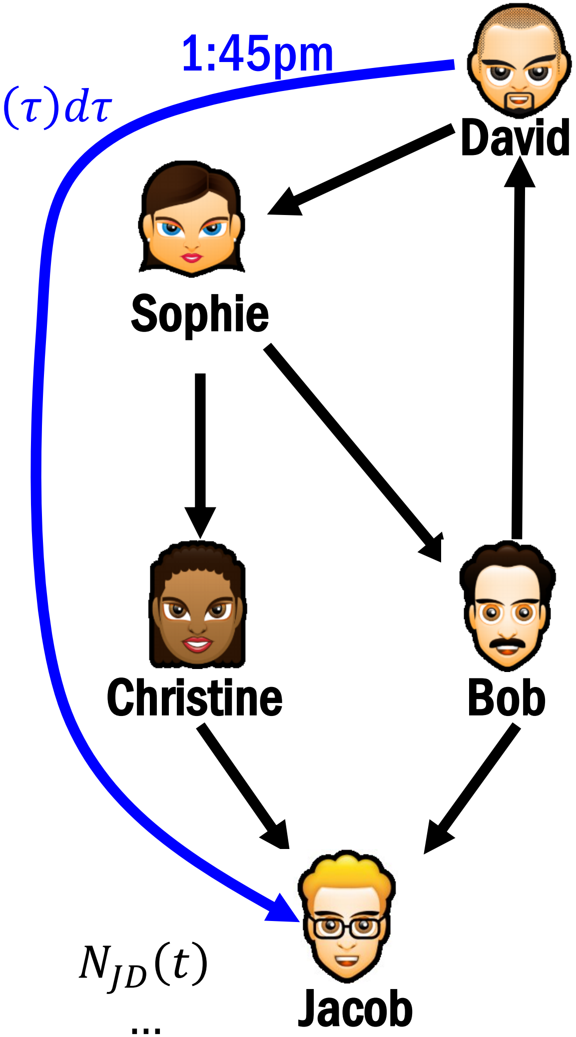
$$\gamma^*_{JD}(t) = \beta_D \exp(-|t|) \star \overbrace{\left(\overbrace{\overbrace{A_{JB}(t)dN_{BD}(t)}^{\substack{\text{Exposure due to} \\ \text{followee } B}} + \overbrace{A_{JC}(t)dN_{CD}(t)}^{\substack{\text{Exposure due to} \\ \text{followee } C}}}^{\substack{\text{Aggregate exposure} \\ \text{from all followees}}}\right)}^{\substack{\text{Hawkes process} \\ \text{High intensity with more exposure}}}$$

# Link creation process

## Retweet more often more likely to link directly

$$A_{JD}(t) = \int_0^t \lambda^*_{JD}(\tau)\, d\tau$$

$$\lambda^*_{JD}(t) = \underbrace{\underbrace{\left(1 - A_{JD}(t)\right)}_{\substack{\text{Check whether} \\ \text{the link already there}}} \cdot \Big(\overbrace{\mu_J}^{\substack{\text{Poisson process} \\ \text{User } J\text{'s own initiative}}} + \underbrace{\alpha_J \exp(-|t|) \star \underbrace{dN_{JD}(t)}_{\text{Retweet } D}}_{\text{Hawkes process}}\Big)}_{\substack{\text{High intensity when} \\ \text{no link and retweet often}}}$$



# Coevolve:

## A Model of Information Diffusion and Network Evolution

Model Formulation → Efficient Simulation and Estimation → Model Properties → Real-Data Experiments

# Simulation

# Estimate parameter via MLE

Given observation window $[0, T)$, a network of $m$ nodes, and a set of

- retweet events $\mathcal{E} = \{e_i^r = (u_i, s_i, t_i)\}$ and
- link creation events $\mathcal{A} = \{e_i^l = (u_i, s_i, t_i)\}$

Find optimal parameters by maximizing log-likelihood:

Concave in model parameters $\mu, \alpha, \eta, \beta$!

Decouple node-wise, parallelizable!

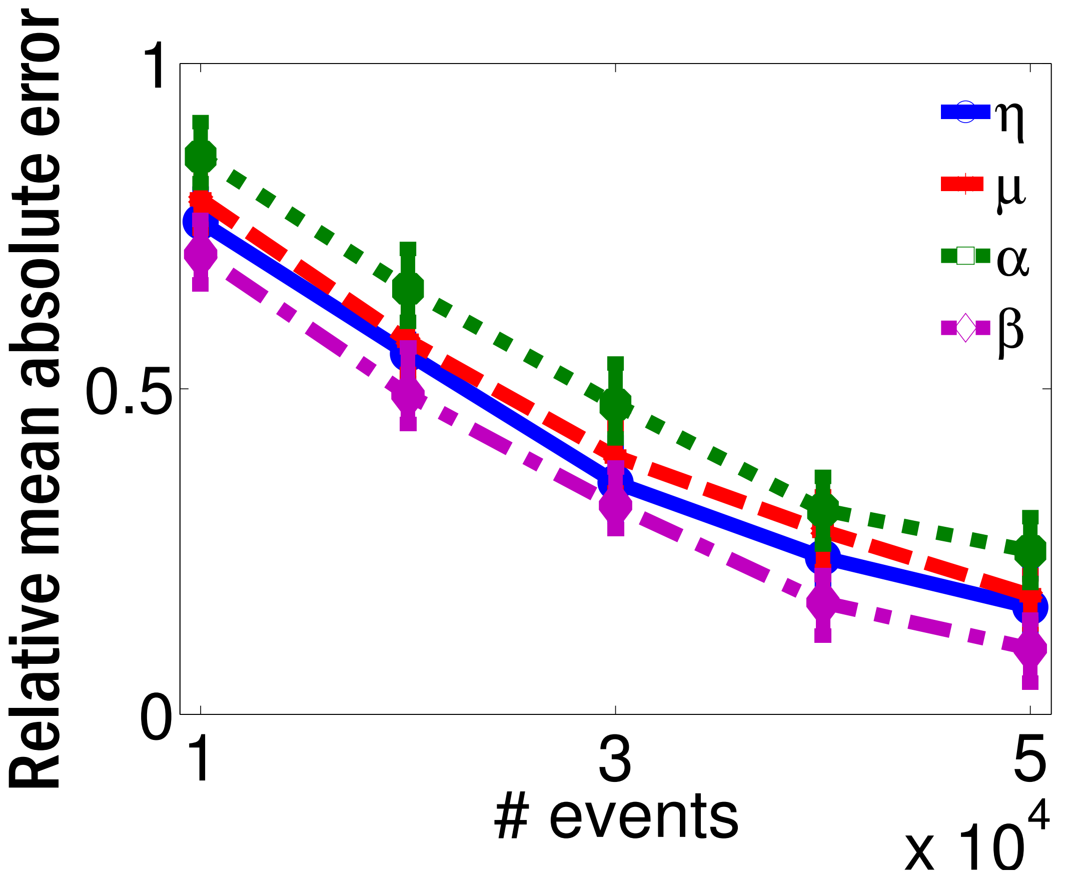
$$\mathcal{L}(\{\mu_u\}, \{\alpha_u\}, \{\eta_u\}, \{\beta_s\}) =$$

$$= \sum_{e_i^r \in \mathcal{E}} \log\left(\gamma_{u_i, s_i}^*(t_i)\right) - \sum_{u,s \in [m]} \int_0^T \gamma_{u,s}^*(\tau)\, d\tau \qquad \text{Tweet/Retweet}$$

$$+ \sum_{e_i^l \in \mathcal{A}} \log\left(\lambda_{u_i, s_i}^*(t)\right) - \sum_{u,s \in [m]} \int_0^T \lambda_{u,s}^*(\tau) d\tau \qquad \text{Link creation}$$

# Model estimation accuracy

## Parameter estimation improves with more events



# Coevolve:
# A Model of Information Diffusion and Network Evolution

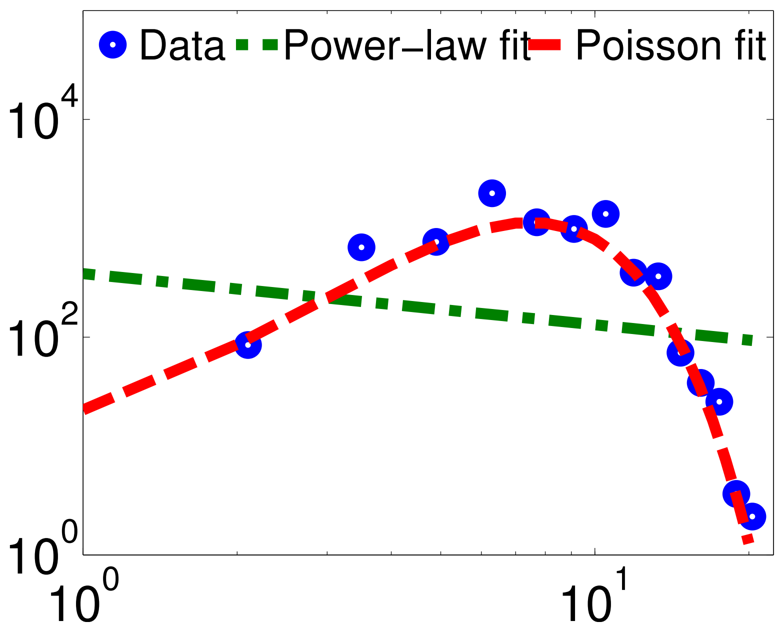
Model Formulation → Efficient Simulation and Estimation → **Model Properties** → Real-Data Experiments

# Degree distribution

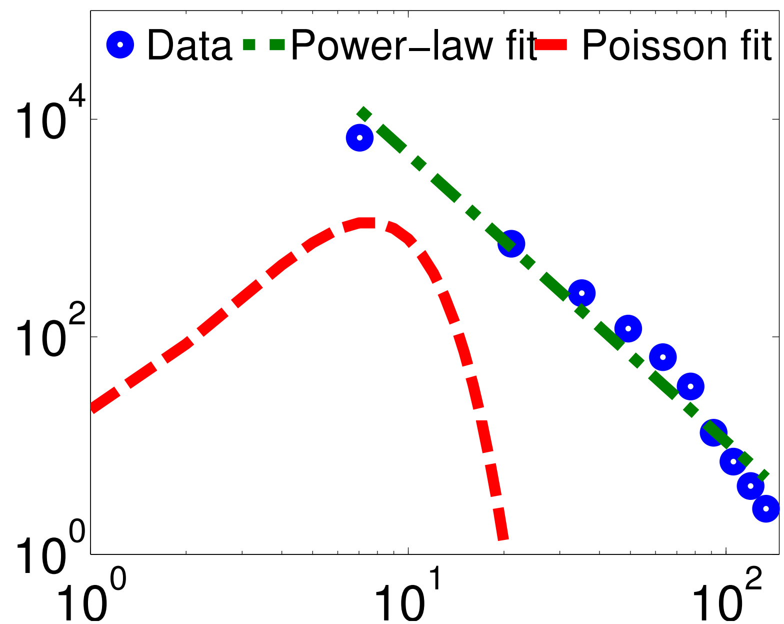
The higher

- the information driven link creation parameter $\alpha$
- or the retweet excitation parameter $\beta$,

the closer to a power-law



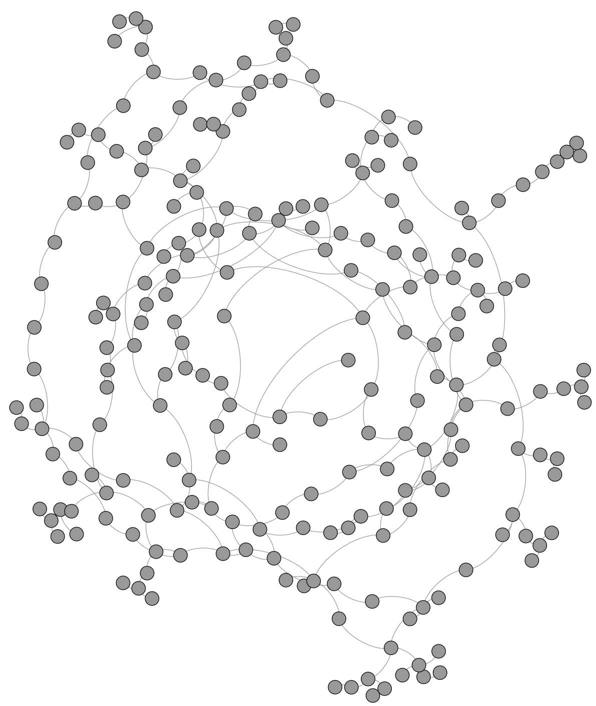

**Poisson, α = 0**



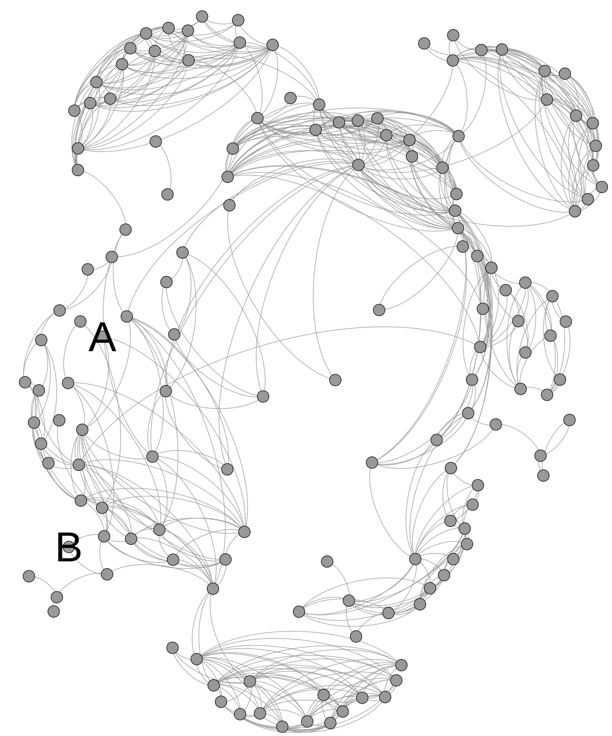

**Power-law, α = 0.2**

# Different types of networks

## Generate networks with very different structure



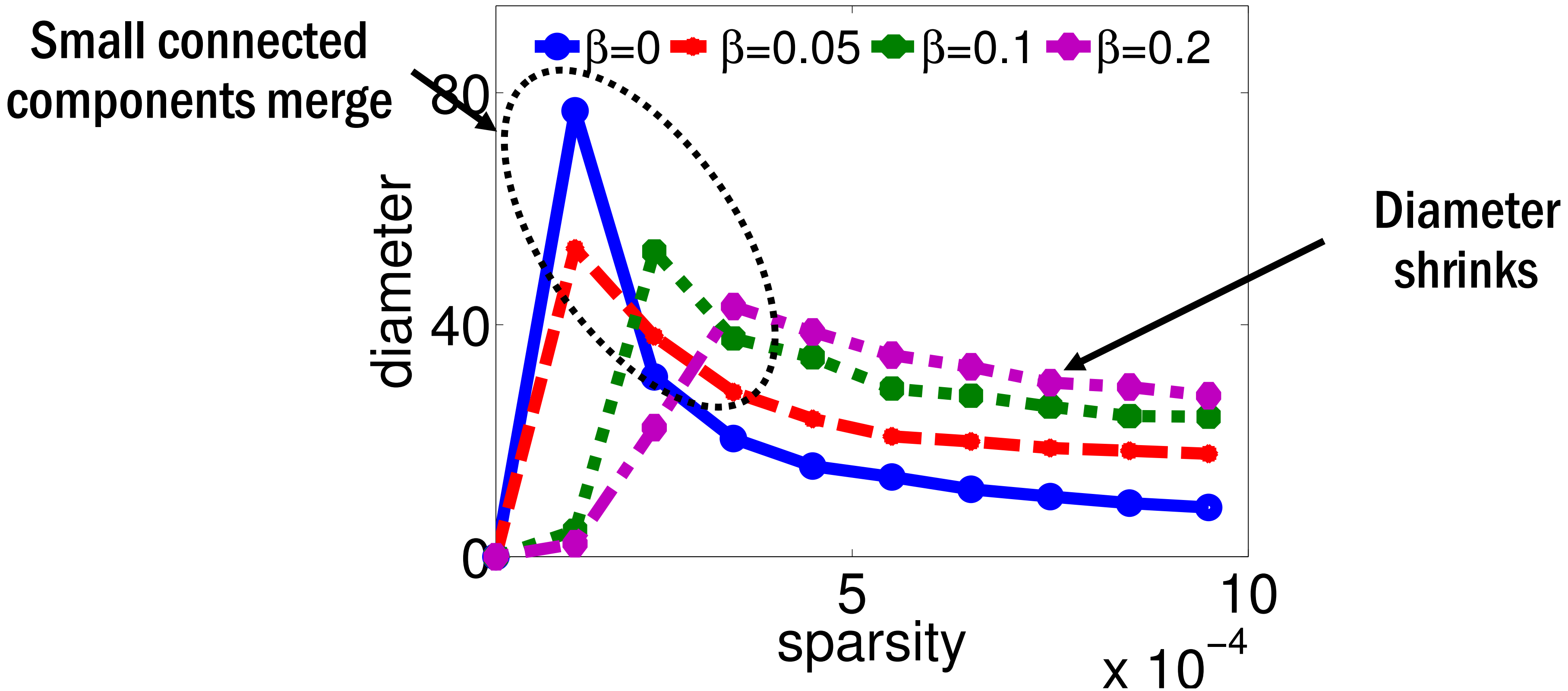
Erdos-Renyi




Scale-free network

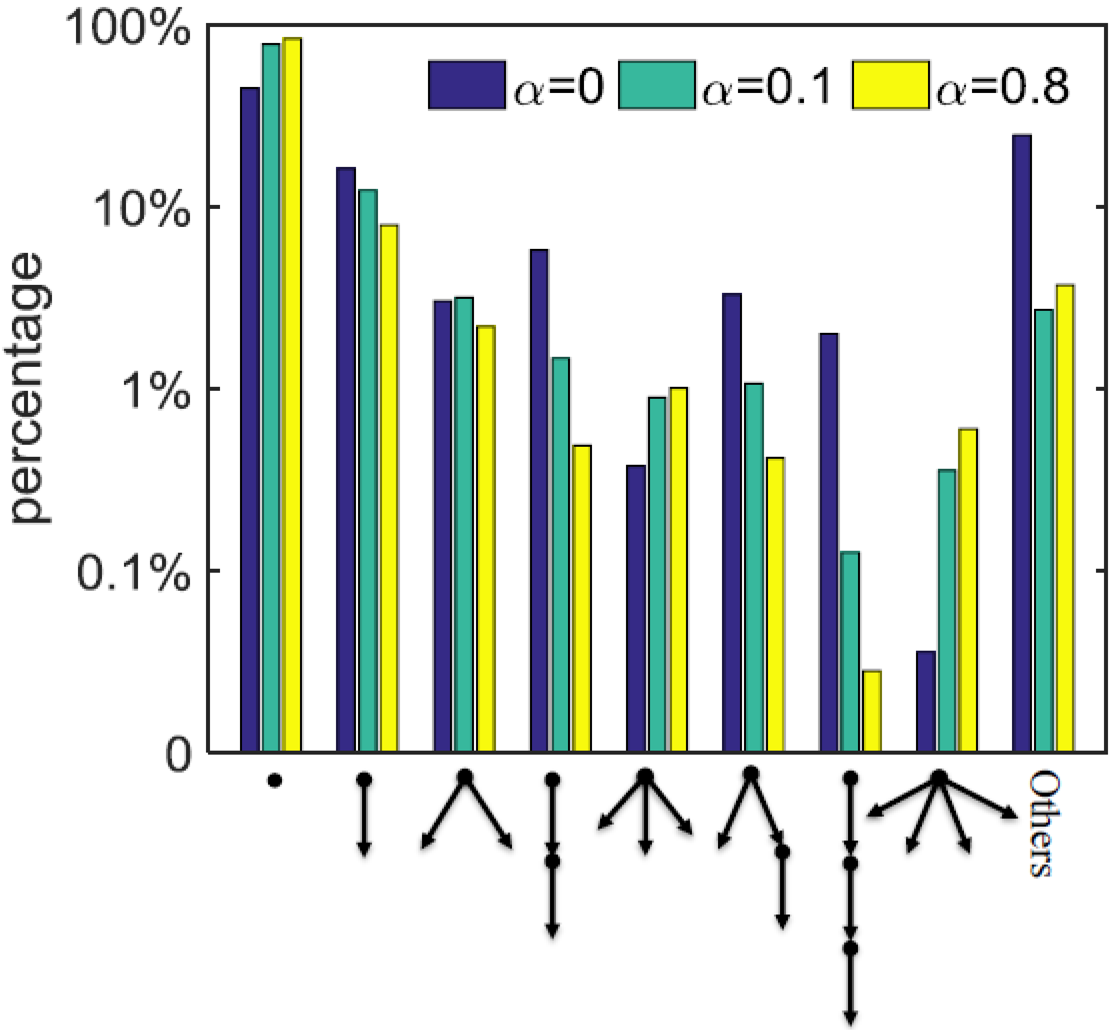
# Network diameters

Generate networks with small shrinking diameter



# Cascade patterns: structure

Generate short and fat cascades as $\alpha$ increases




$\beta = 0.2$

# Coevolve:

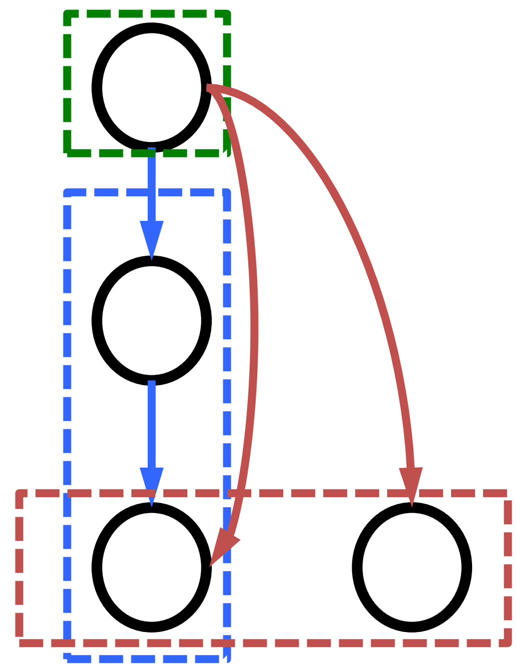
## A Model of Information Diffusion and Network Evolution

Model Formulation → Efficient Simulation and Estimation → Model Properties → Real-Data Experiments

# Links, tweets and retweets

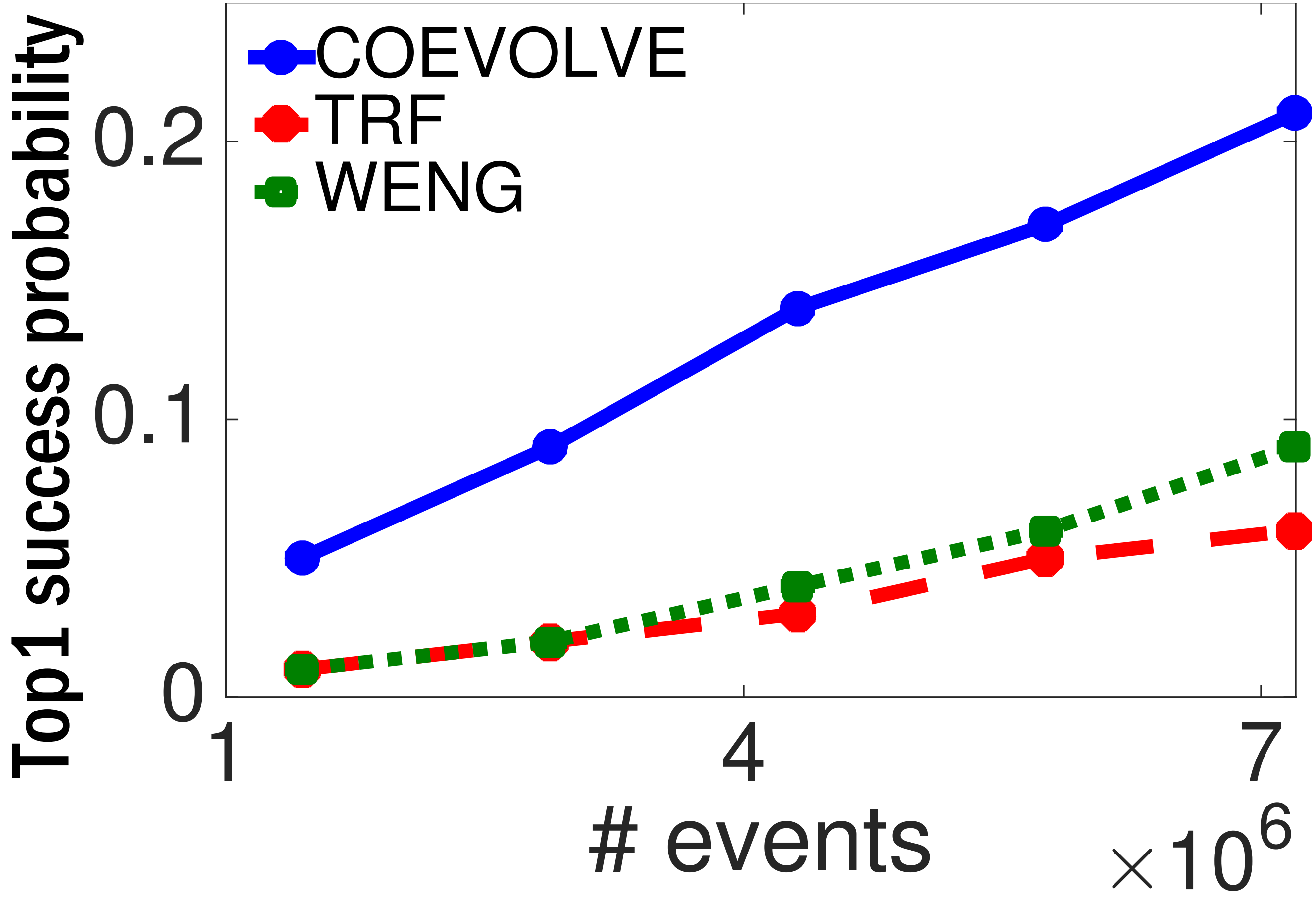
Evaluate with a Twitter dataset from [Antoniades and Dovrolis '13]

- ~371K tweets
  from 8,779 source nodes
- ~130K retweets
  from 77,200 users
- ~7M new links to source
  nodes by ~6M users



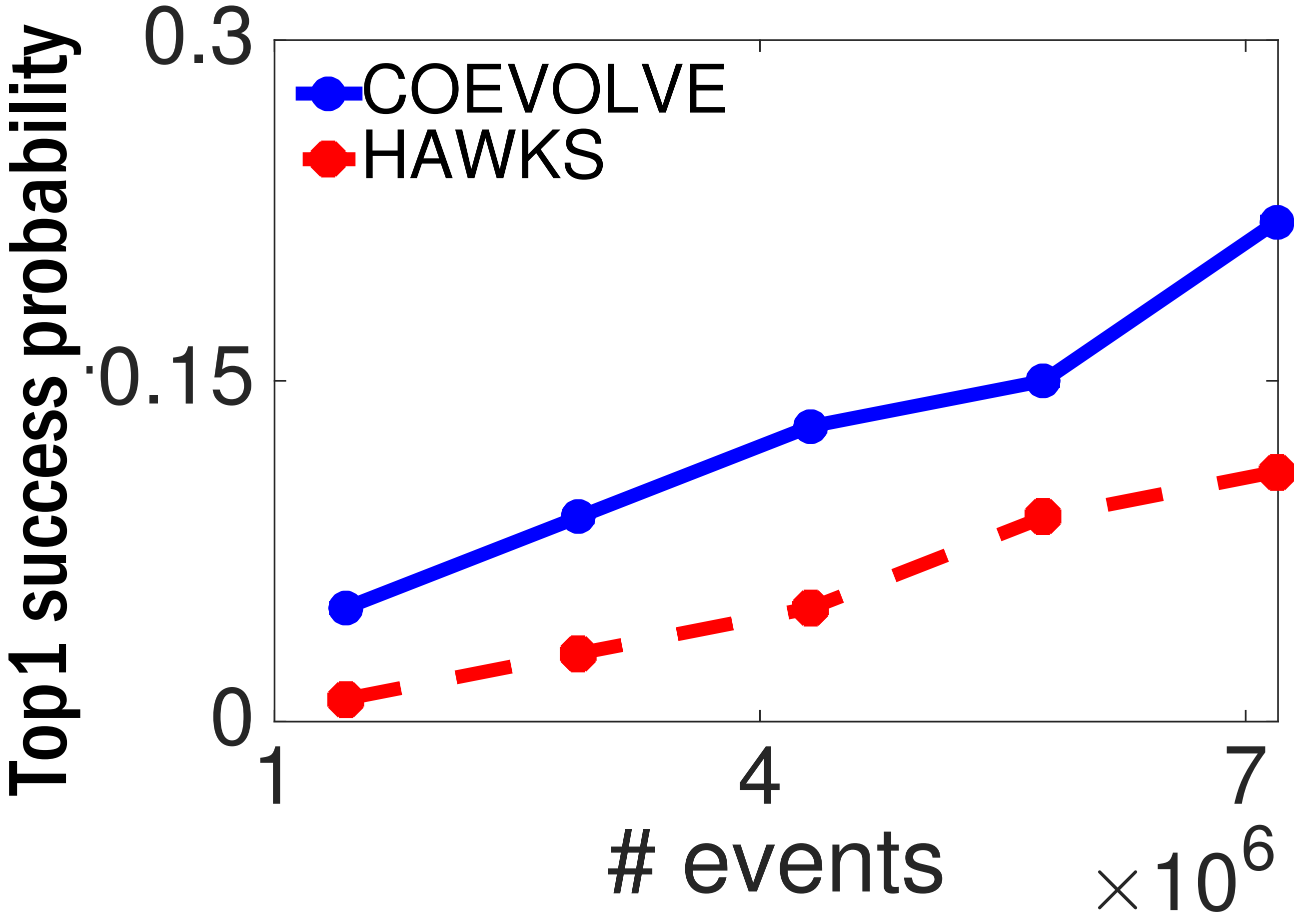
# Link creation prediction

Outperform the predictions given by two state-of-the-art [Weng et al '13, Myers & Leskovec '14]



# Information diffusion prediction

Beats the predictions given by a standard Hawkes process

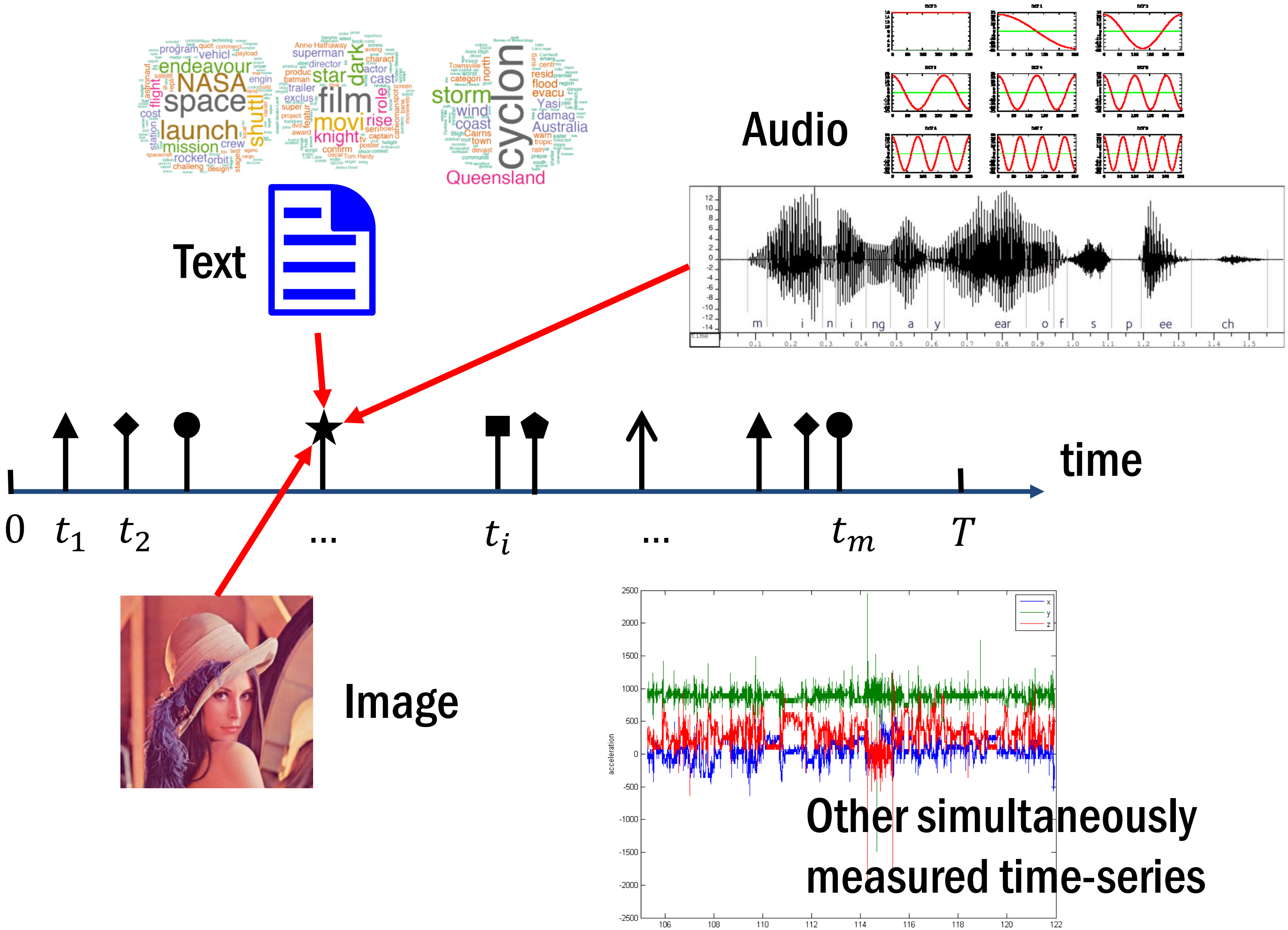


# Conclusion

- A data-driven joint point process model of information diffusion and network coevolution
  - Simple generative model
  - Efficient simulation
  - Convex estimation
  - Microscopic: more accurate link and event prediction
  - Macroscopic: realistic network properties and information diffusion properties

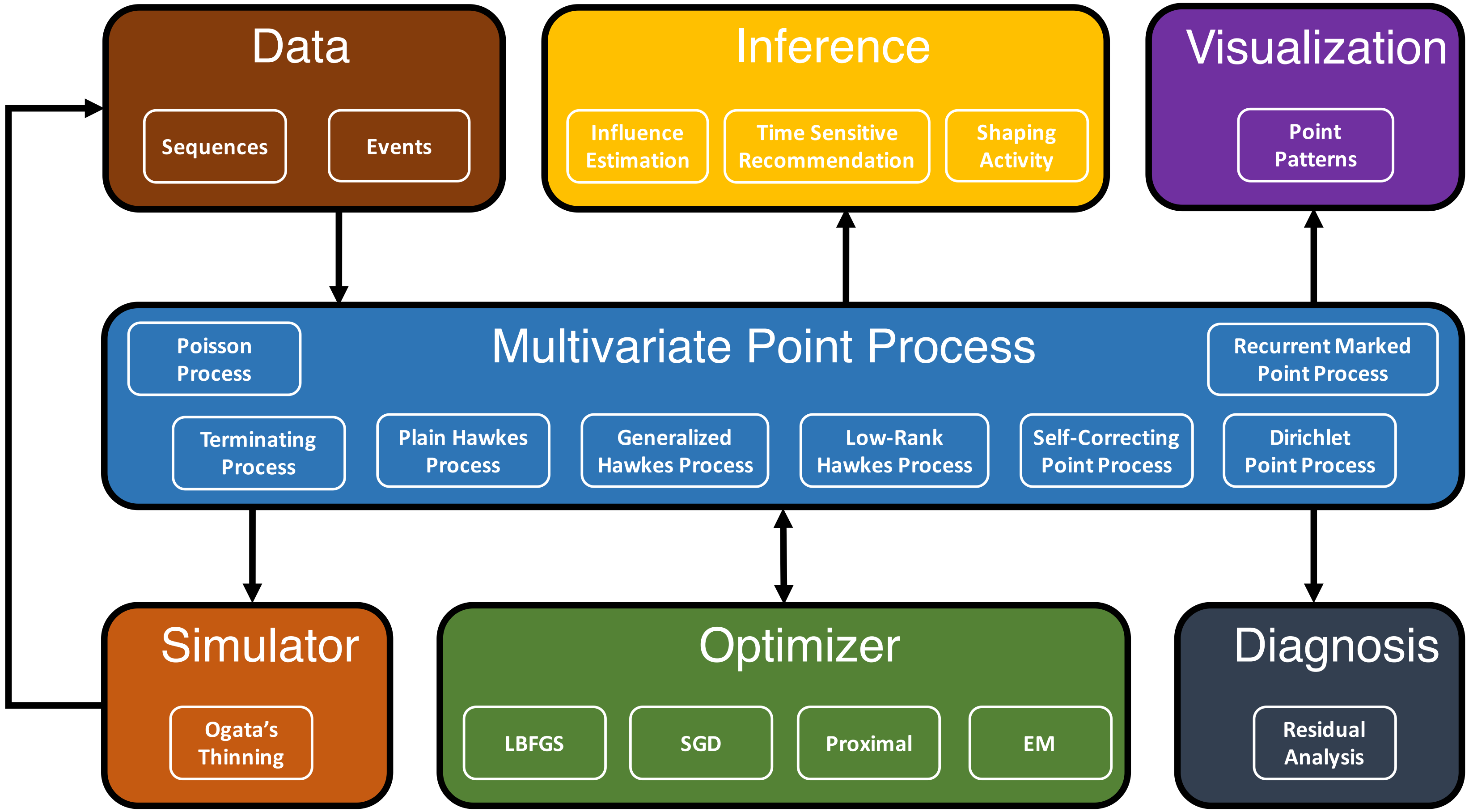
- Many possible extensions, such as
  - Node birth and death
  - Link deletion
  - Incorporate node attributes and tweet contents
  - Deep learning for the intensity function

# Joint models with rich context




Nan et al. AISTATS 2013
Nan et al. KDD 2015

# PtPack: C++ point process package




https://github.com/dunan/MultiVariatePointProcess

PtPack