Information propagation

with Temporal Point Processes

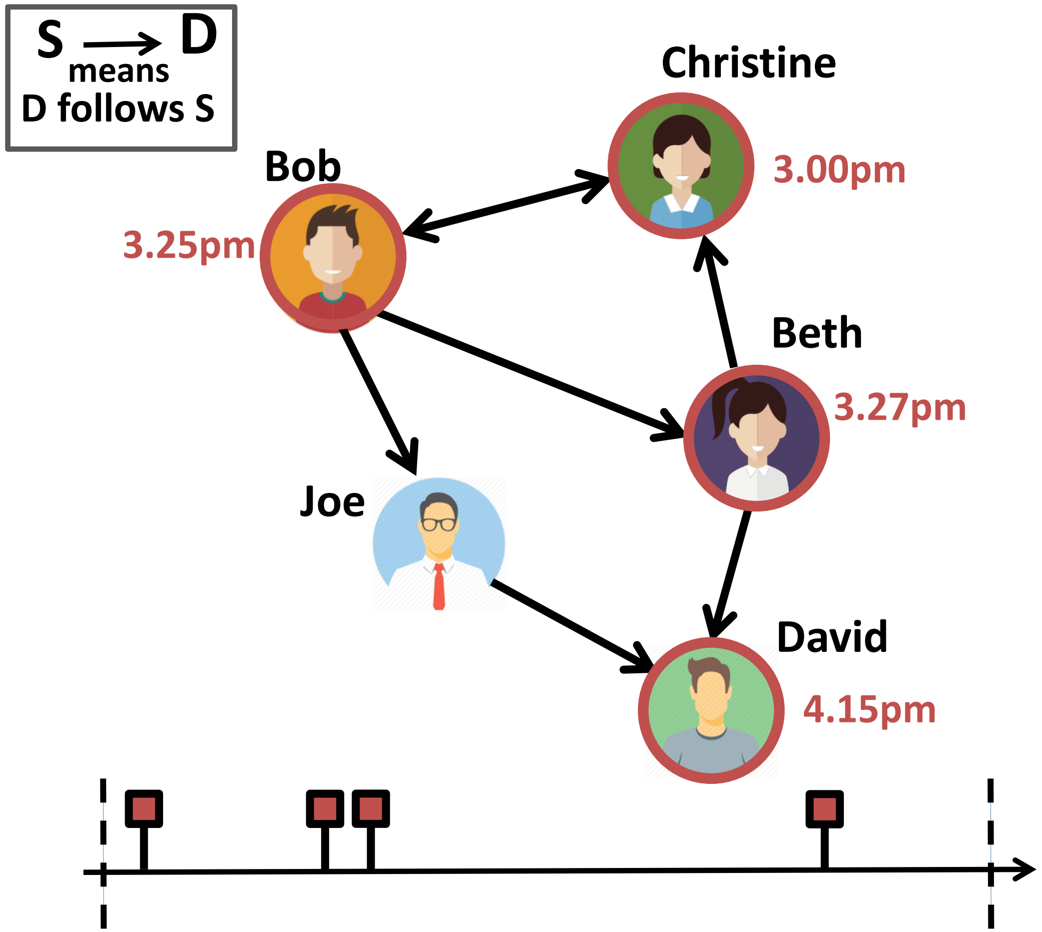
HUMAN-CENTERED MACHINE LEARNING

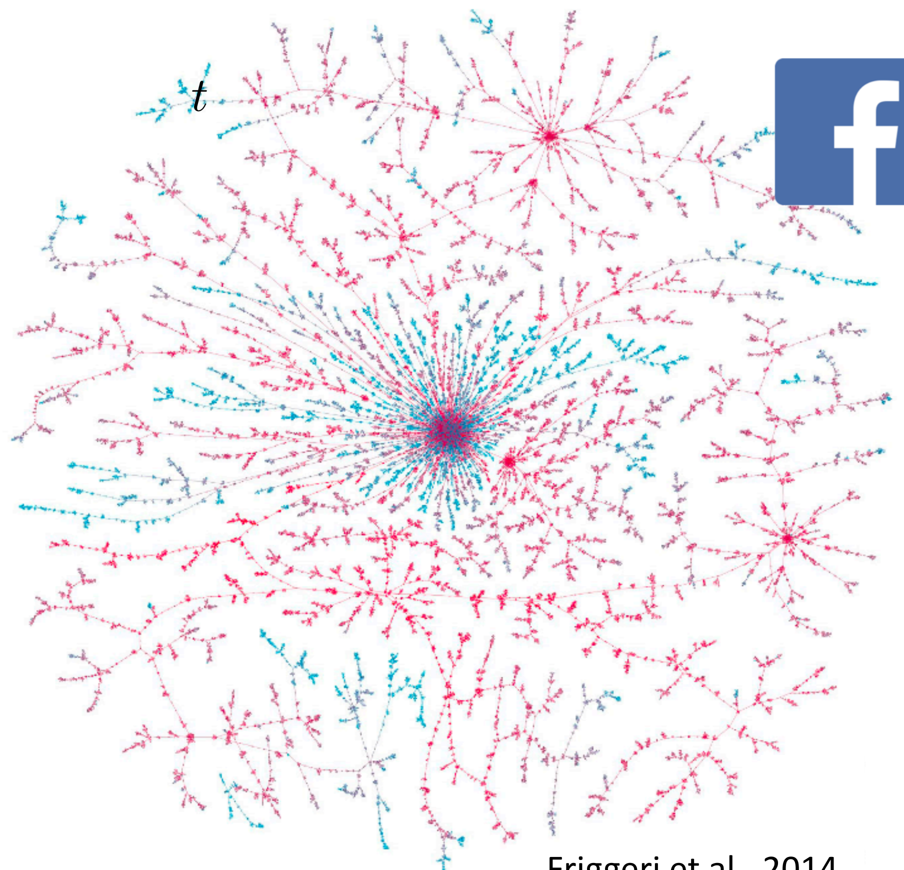
http://courses.mpi-sws.org/hcml-ws18/

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Information cascades:
Terminating point process models

An example: information cascade





Friggeri et al., 2014

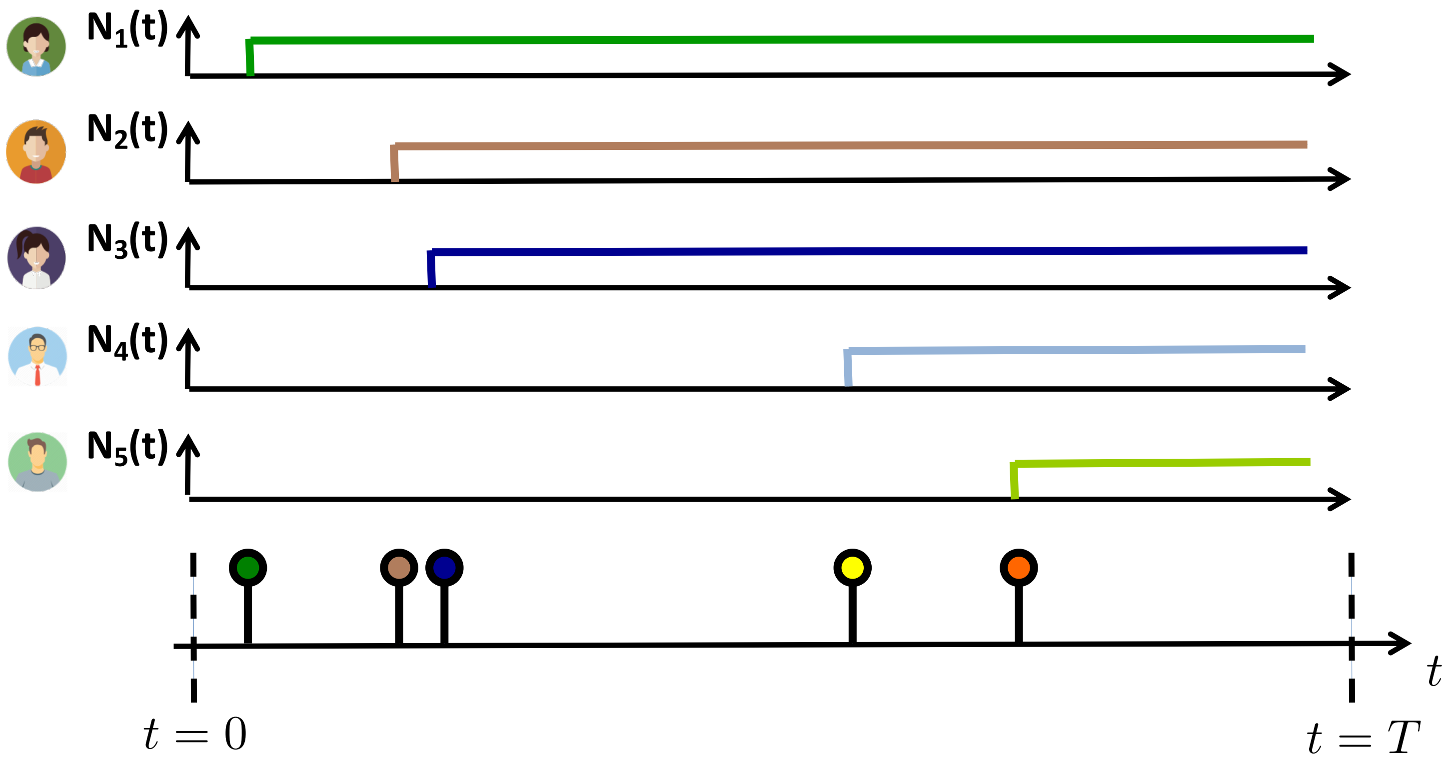
They can have an impact in the off-line world

theguardian

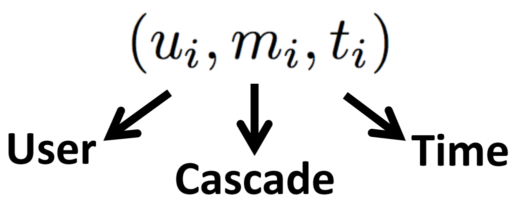
Click and elect: how fake news helped Donald Trump win a real election

Information cascade representation

We represent an information cascade using **terminating temporal point processes**:



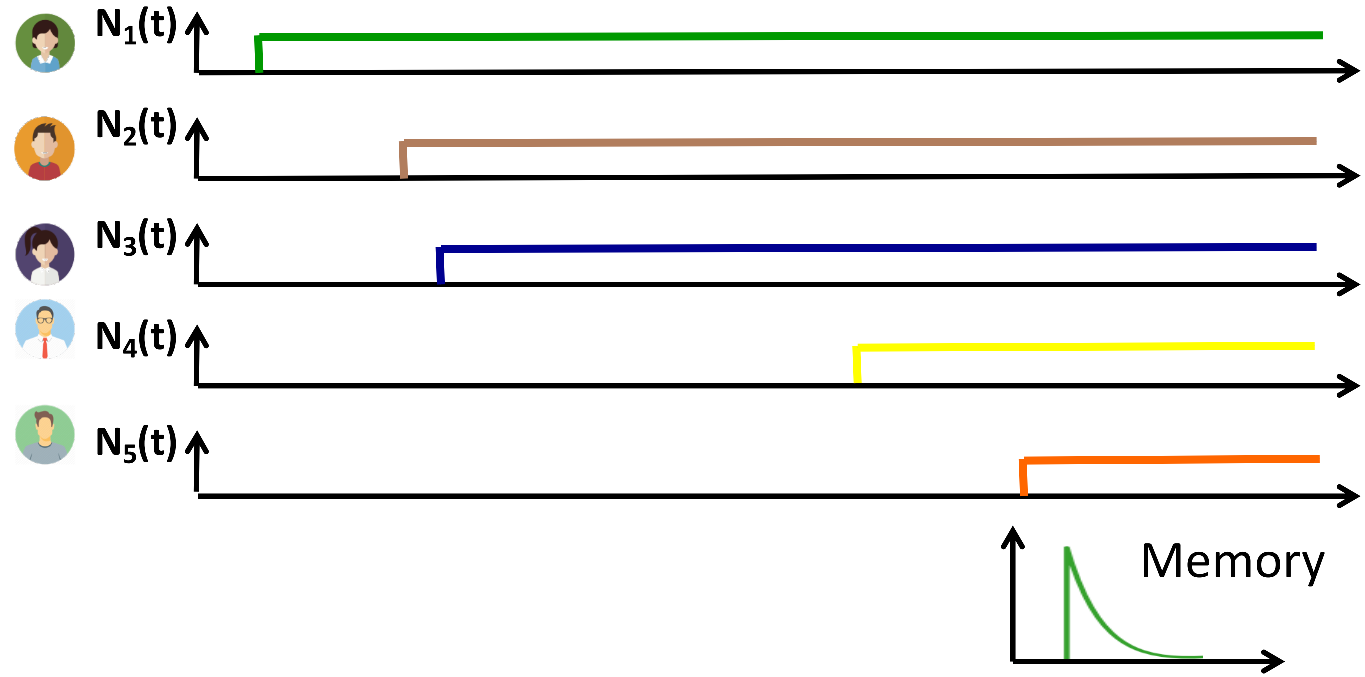
***Sharing* event:**



Information cascade intensity

Source
(given, not modeled)

Shares / follow-ups
(modeled)



$$\lambda^*_u(t) = \underbrace{(1 - N(t))}_{\text{Share only once}} \sum_{v \in [m]} b_{vu} \underbrace{\sum_{e_i \in \mathcal{H}_v(t)} \kappa(t - t_i)}_{\text{Previous message by user v}}$$

Influence from user v on user u

[Gomez-Rodriguez et al., ICML 2011]

Model inference from multiple cascades

Conditional intensities

$$\lambda_u^*(t)$$

Cascade log-likelihood

$$\mathfrak{L} = \sum_{u=1}^{n} \log \lambda_u^*(t_u) - \int_0^T \lambda_u^*(\tau)\, d\tau$$

Sum up log-likelihoods of multiple cascades!

Maximum likelihood approach to find model parameters!

Theorem. For any choice of parametric memory, the **maximum likelihood** problem is **convex in B**.

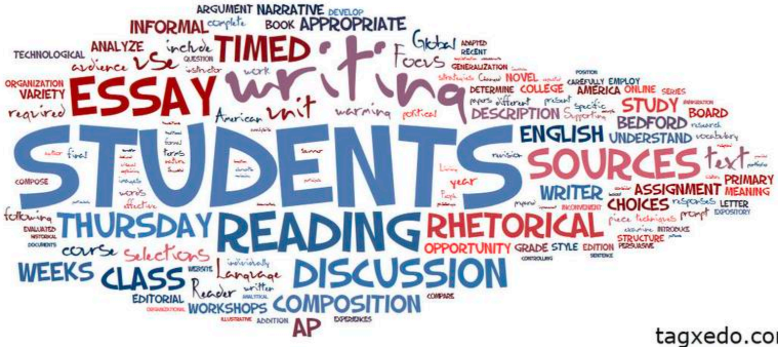
[Gomez-Rodriguez et al., ICML 2011]

Topic-sensitive intensities

Topic-modulated influence:

$$b_{vu} = \sum_{l=1}^{K} b_{vu}^{l}\, m_l$$

LDA weight for topic l



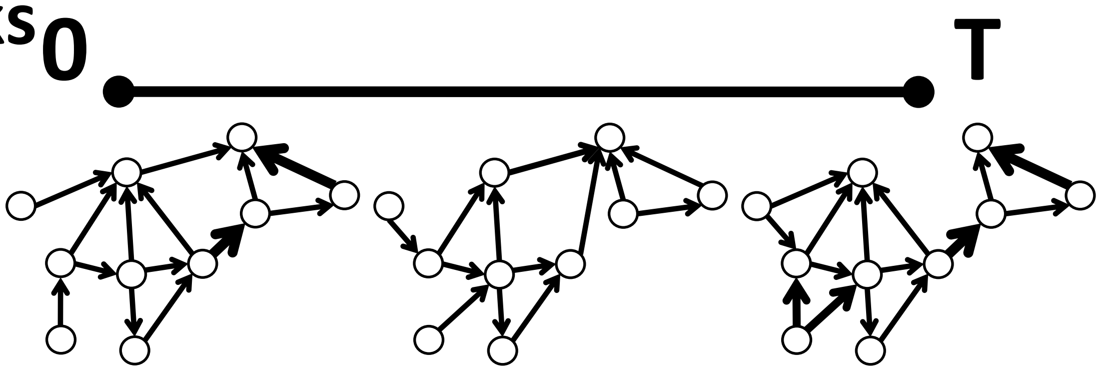
tagxedo.com

[Du et al., AISTATS 2013]

Dynamic influence

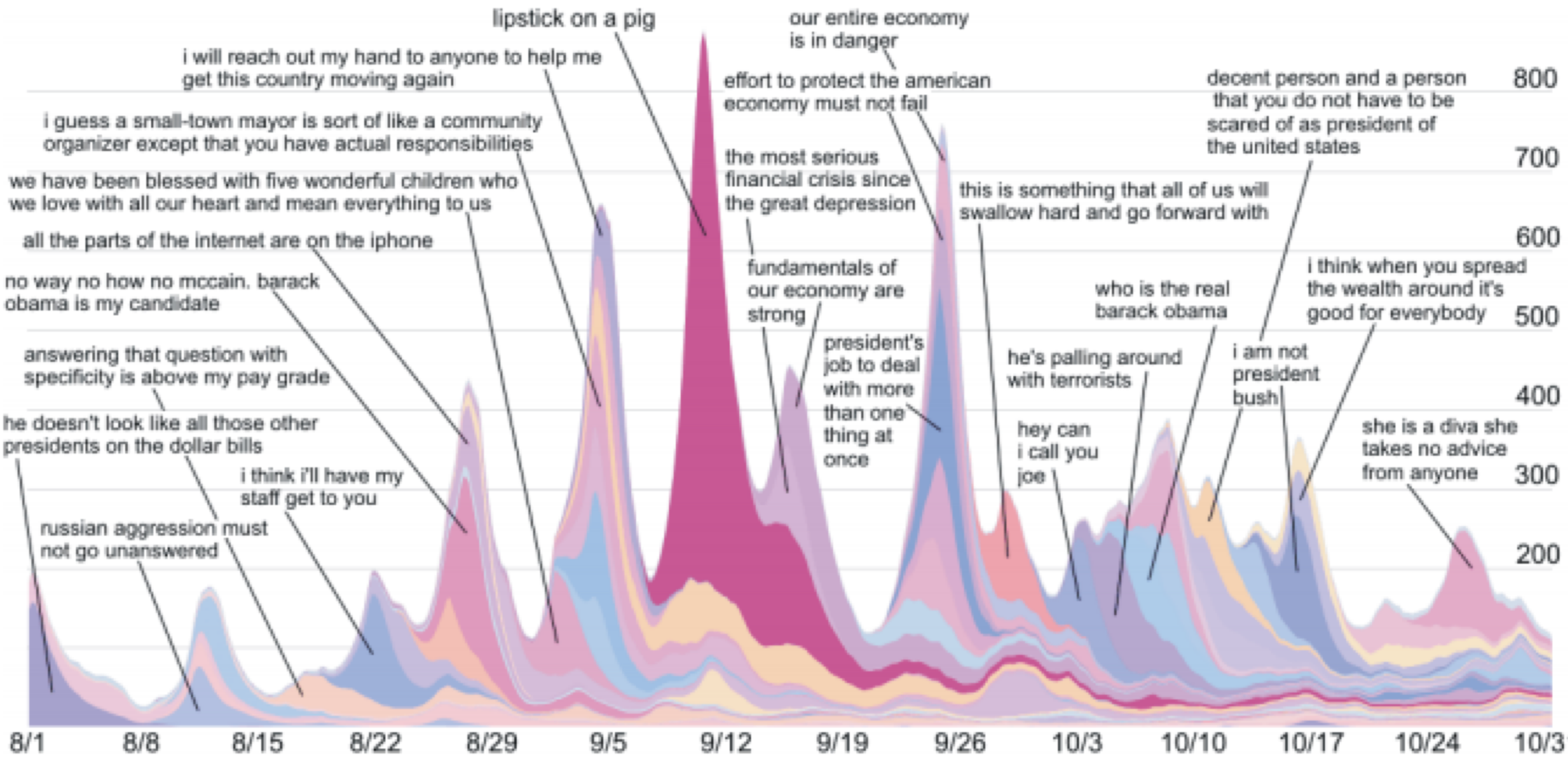
In some cases, influence change over time:

Propagation over networks with variable influence



[Gomez-Rodriguez et al., WSDM 2013]

Memetracker



[Leskovec et al., KDD '09]

Insights I: real world events

Youtube video: http://youtu.be/hBeaSfRCU4c

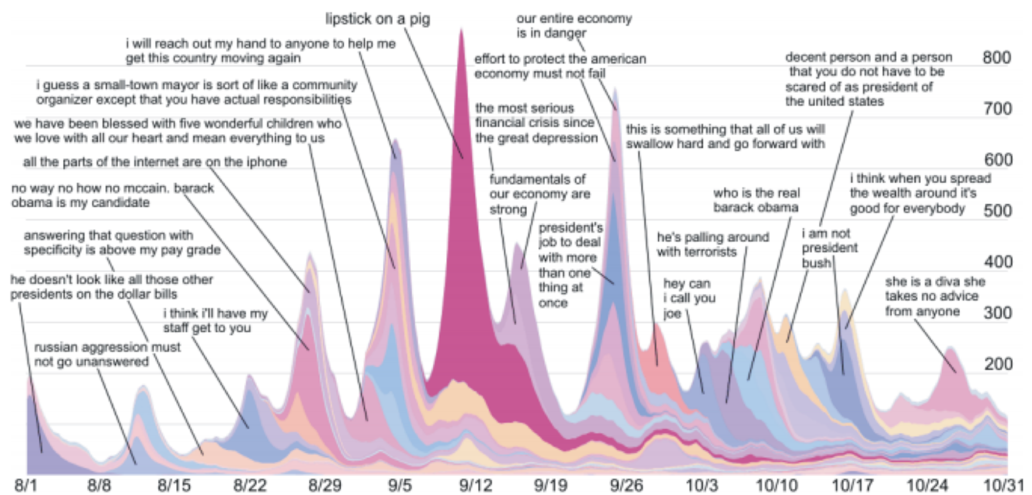
Insights II: dynamic clusters

Youtube video: http://youtu.be/hBeaSfRCU4c

Beyond information cascades:
Nonterminating point process models

Recurrent events: beyond cascades

Up to this point, we have assumed we can map each event to a cascade



In general, especially in social networks:

Difficult **to distinguish cascades** in event data

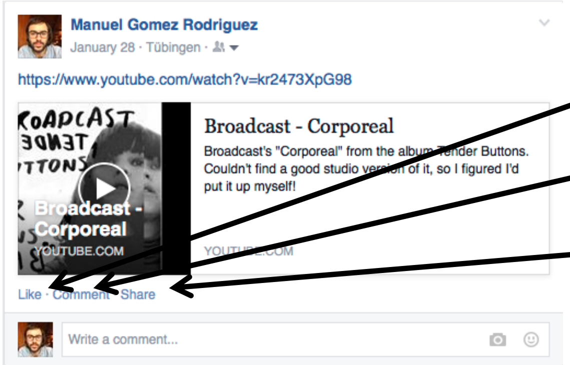
BUSINESS INSIDER

He has stuck to his decision so far; his recent Facebook status read, "I just killed a pig and a goat."

GAWKER

Mark Zuckerberg Is Killing Progressively Larger Animals

Most cascades are single nodes (or *forests*)



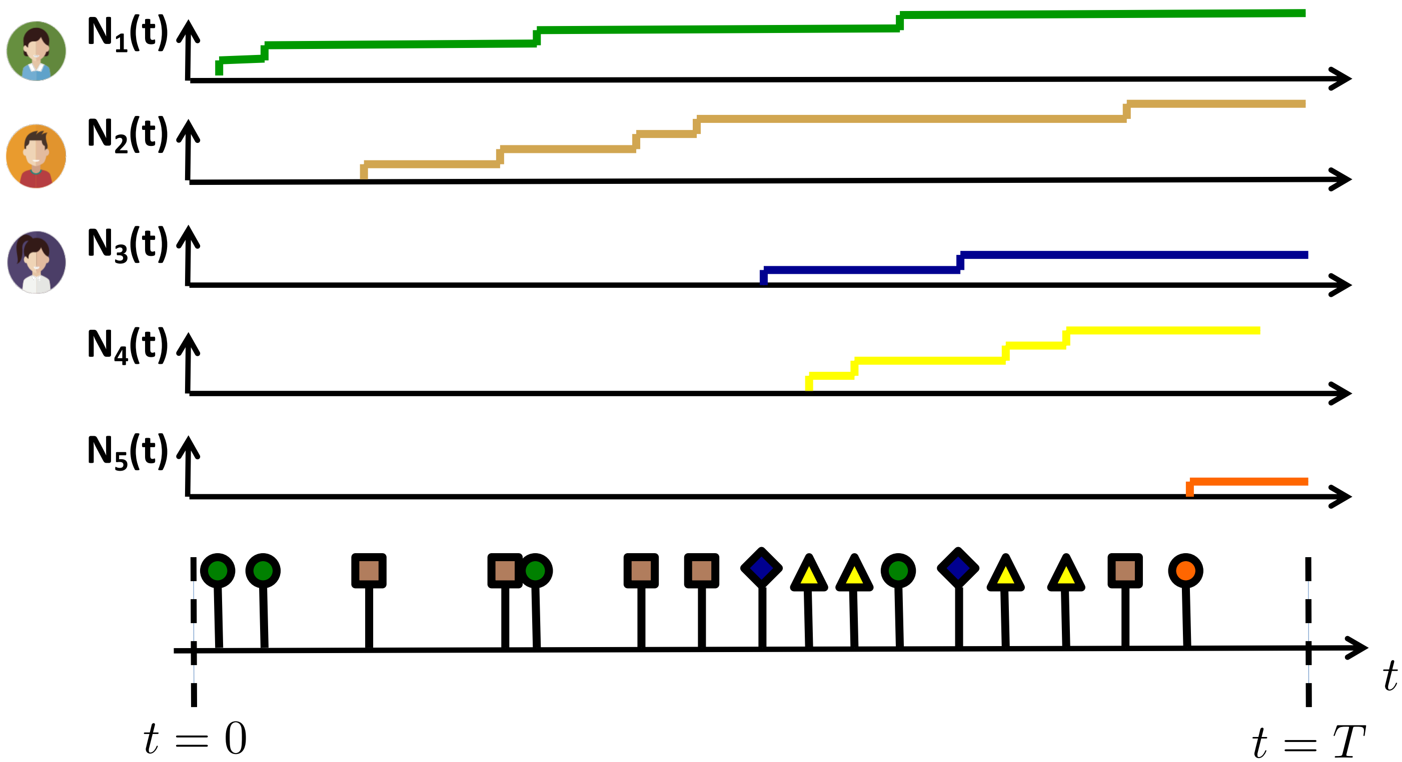
no likes

no comments

no shares

Recurrent events representation

We represent shares using **nonterminating temporal point processes**:

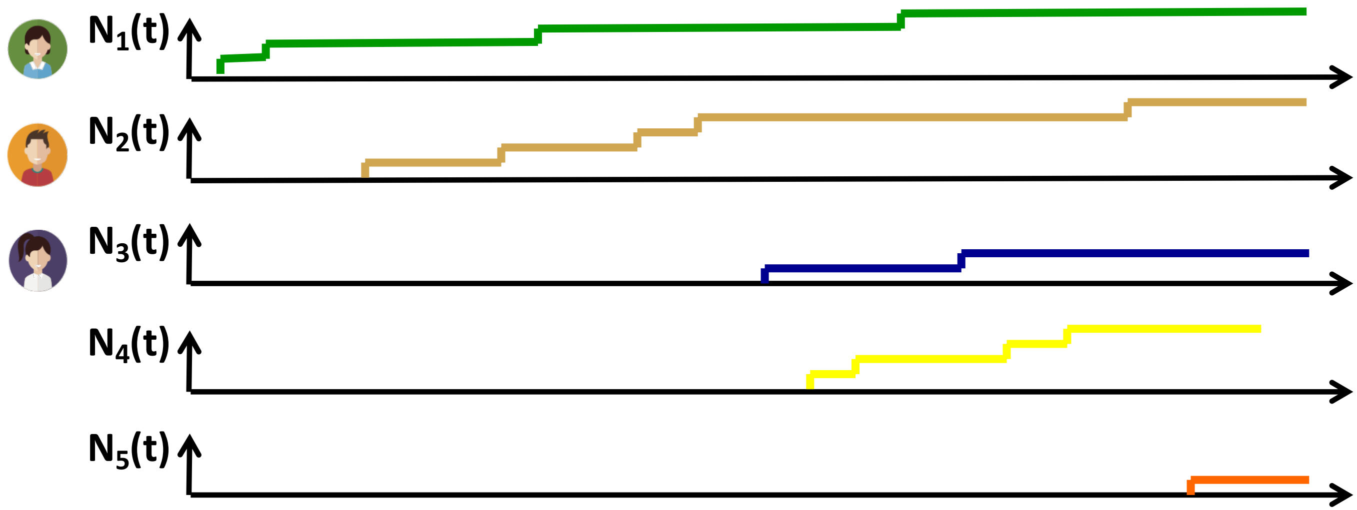


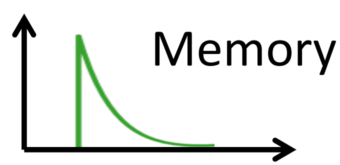
Recurrent event:

$$(u_i, t_i)$$

[Farajtabar et al., NIPS 2014]

Recurrent events intensity





Cascade sources!

$$\underbrace{\lambda_u^*(t)}_{\text{User's intensity}} = \underbrace{\mu_u}_{\text{Messages on her own initiative}} + \sum_{v\in[m]} b_{vu} \underbrace{\sum_{e_i\in\mathcal{H}_v(t)} \kappa(t-t_i)}_{\text{Previous messages by user v}}$$

Influence from user v on user u

Hawkes process

[De et al., NIPS 2016]

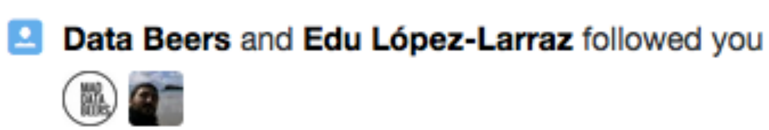
Information cascades and network evolution:
Nonterminating point process models

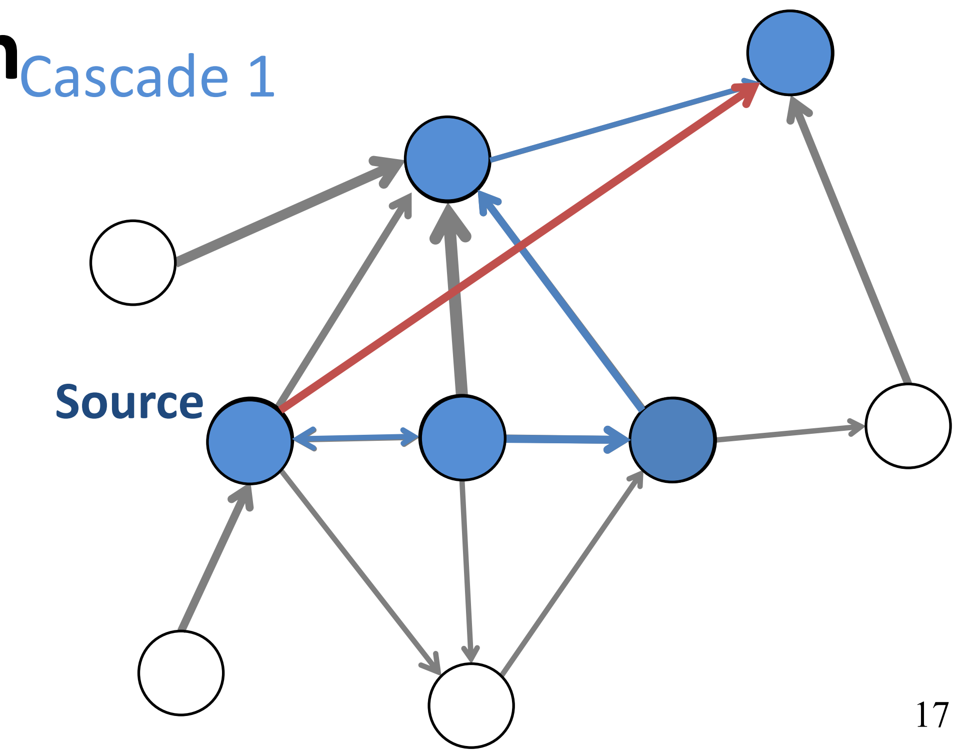
Beyond information cascades (II)

Recent empirical studies [Antoniades and Dovrolis, Myers & Leskovec] show that **information cascades also change the structure of social networks:**

Information propagation triggers new links

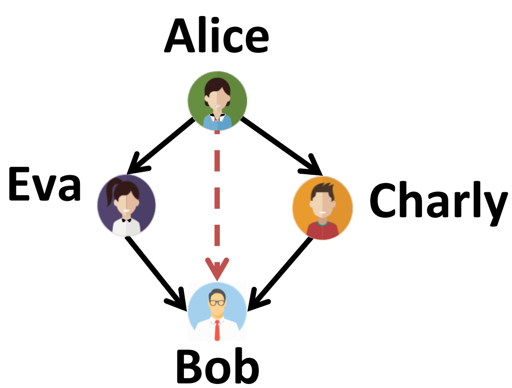
1 hour later





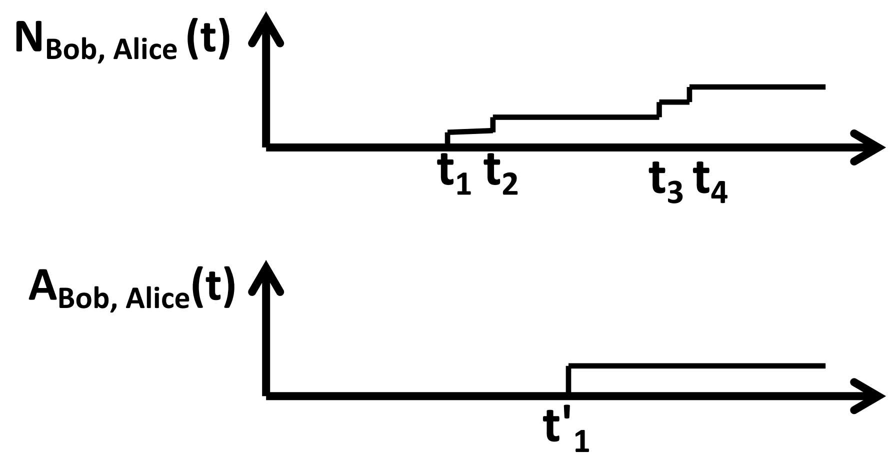
Co-evolution as interwoven point processes (I)

We model user's *retweet* and *link* events as nonterminating and terminating counting processes:



Bob *retweets (is exposed to)* Alice

Bob follows Alice



Key idea

Both counting processes have **memory** and **depend on each other**

Co-evolution as interwoven point processes (II)

We characterize retweet and link counting processes using their respective conditional intensities:

$$\mathbb{E}[\overbrace{d\boldsymbol{N}(t)}^{\text{Changes on retweets in [t, t+dt]}} \,|\, \underbrace{\mathcal{H}^r(t) \cup \mathcal{H}^l(t)}_{\text{History of retweets and links up to t}}] = \underbrace{\boldsymbol{\Gamma}^*(t)}_{\text{Instantaneous rates or intensities}}\, dt$$

$$\mathbb{E}[\underbrace{d\boldsymbol{A}(t)}_{\text{Changes on links in [t, t+dt]}} \,|\, \overbrace{\mathcal{H}^r(t) \cup \mathcal{H}^l(t)}] = \overbrace{\boldsymbol{\Lambda}^*(t)}\, dt$$

They are coupled through the histories

$$\mathcal{H}^r(t) \cup \mathcal{H}^l(t)$$

Intensity for information propagation

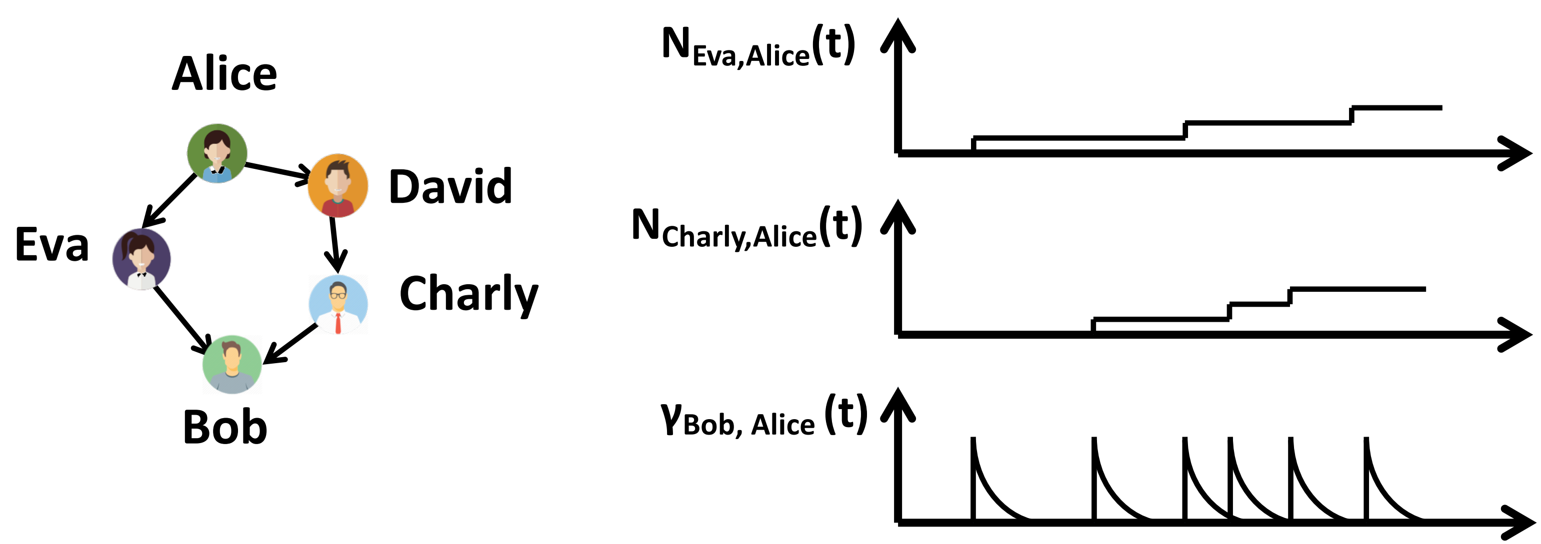
$$N_{us}(t) \updownarrow$$

Tweets on her own initiative

Time-varying network topology

$$\gamma^*_{us}(t) = \begin{cases} \eta_u & u = s \\ \underbrace{\beta_s \sum_{v \in \mathcal{F}_u(t)} \kappa_{\omega_1}(t) \star \left(A_{uv}(t)\; dN_{vs}(t) \right)}_{\text{Propagation of peer influence over the network}} & u \neq s \end{cases}$$

Node u does not need to follow s!



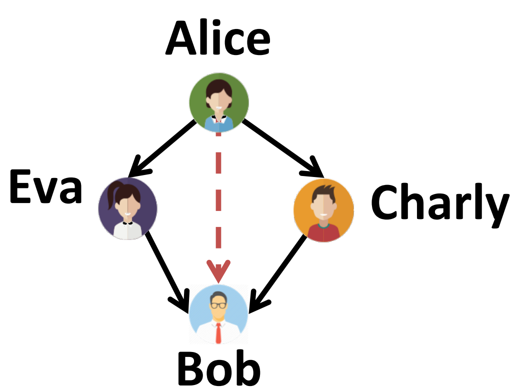
Intensity for network evolution

$$\lambda^*_{us}(t) = \underbrace{(1 - A_{us}(t))}_{\text{Ensures a link is created only once}}(\mu_u + \underbrace{\alpha_u\, \kappa_{\omega_2}(t) \star dN_{us}(t)}_{\text{Influence of retweet intensity on the link creation}})$$

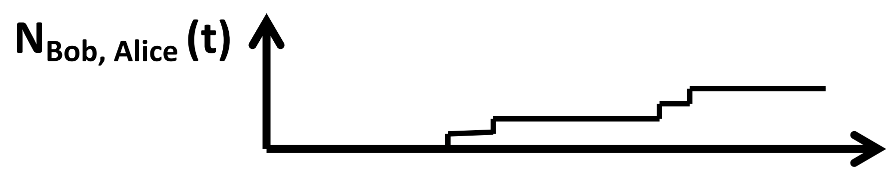
$$\Updownarrow$$

$$A_{us}(t)$$

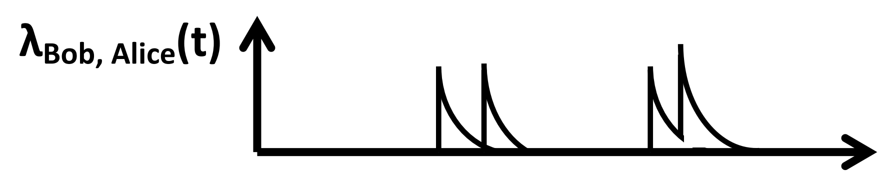
Links on her own initiative



Bob *retweets* *(is exposed to)* Alice



Bob's *risk* of following Alice



Model inference from historical data

Find *optimal* parameters using **maximum likelihood estimation (MLE)**:

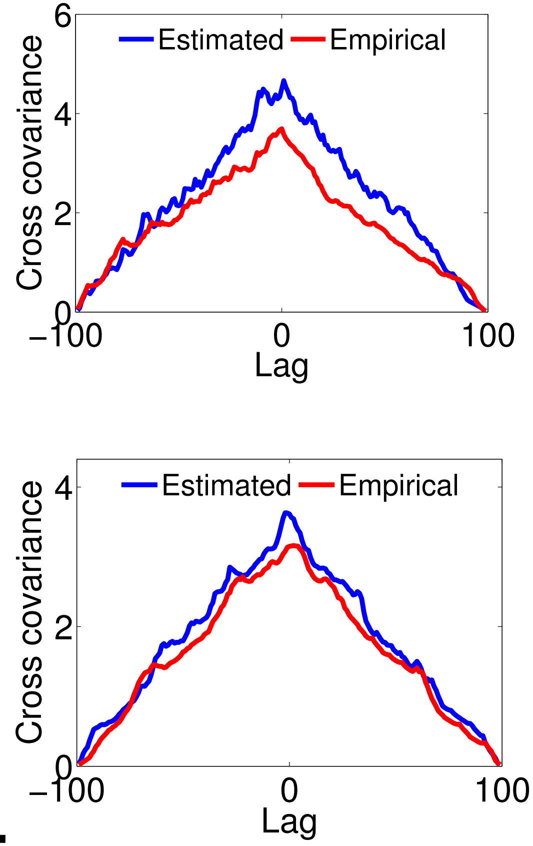
$$\mathfrak{L}(\{\mu_u\},\{\alpha_u\},\{\eta_u\},\{\beta_s\}) = \underbrace{\sum_{e_i^r \in \mathcal{E}} \log\left(\gamma^*_{u_i s_i}(t_i)\right) - \sum_{u,s \in [m]} \int_0^T \gamma^*_{us}(\tau)\, d\tau}_{\text{tweet / retweet}} +$$

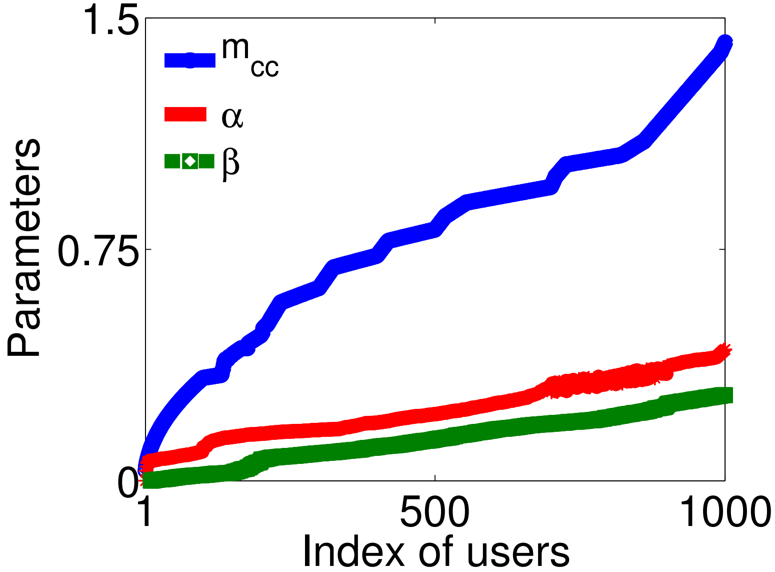
$$\underbrace{\sum_{e_i^l \in \mathcal{A}} \log\left(\lambda^*_{u_i s_i}(t_i)\right) - \sum_{u,s \in [m]} \int_0^T \lambda^*_{us}(\tau)\, d\tau}_{\text{links}}$$

For the choice of information propagation and link intensities, the MLE problem above is **parallelizable & convex**.

Retweet and link coevolution

Cross-covariance for two users



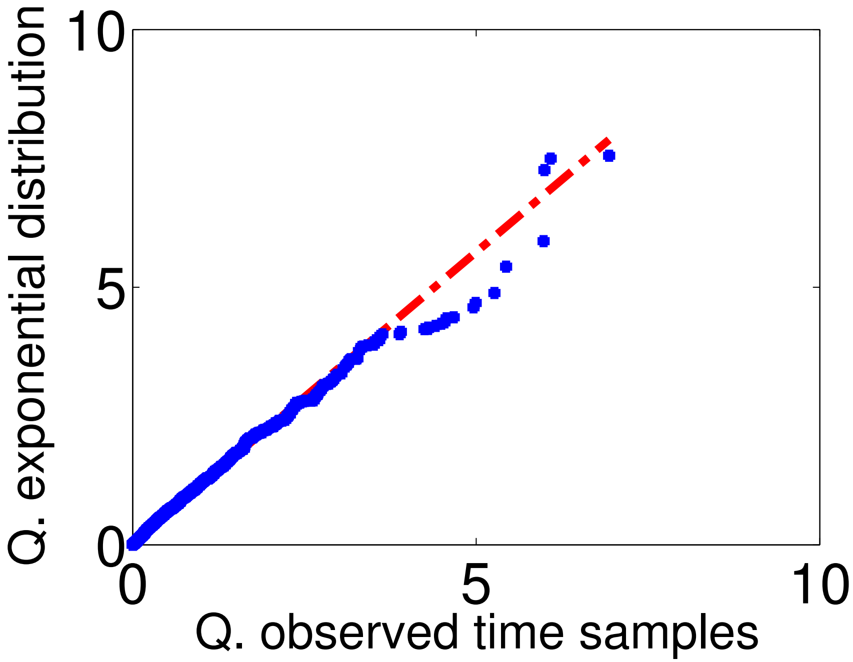


Average cross-covariance vs model parameters

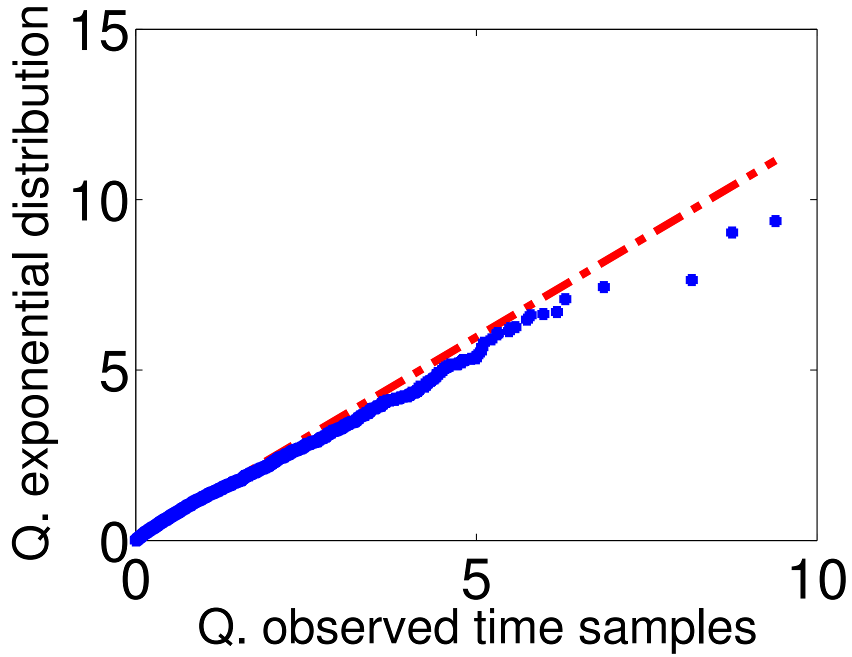
The fitted model generate **link and information diffusion events that coevolve similarly** (in terms of cross-covariance) **as real events.**

Model checking

https://en.wikipedia.org/wiki/Q-Q_plot



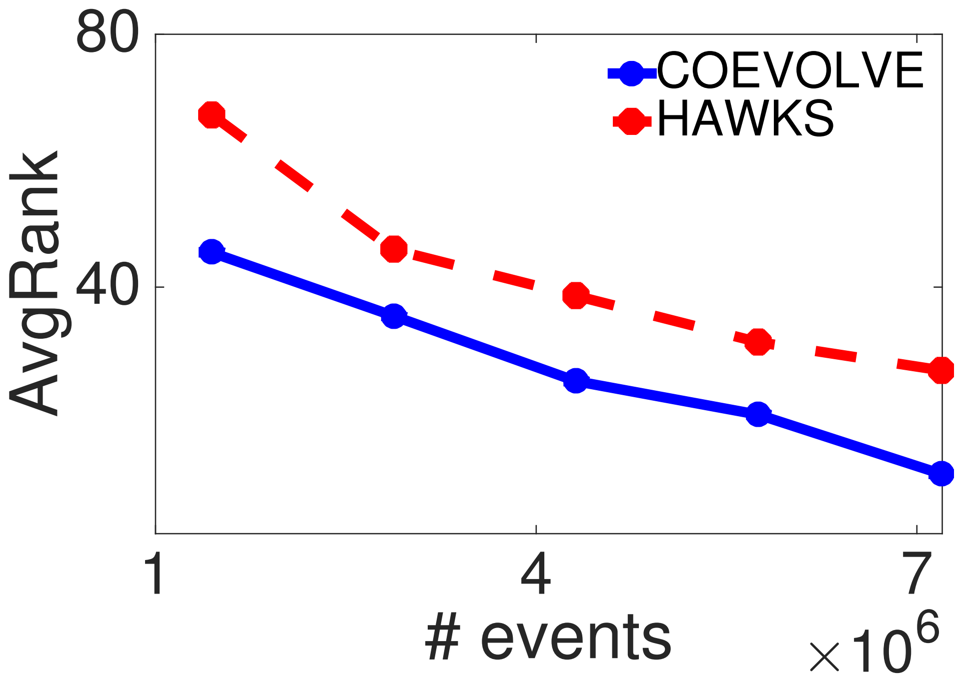
Link events



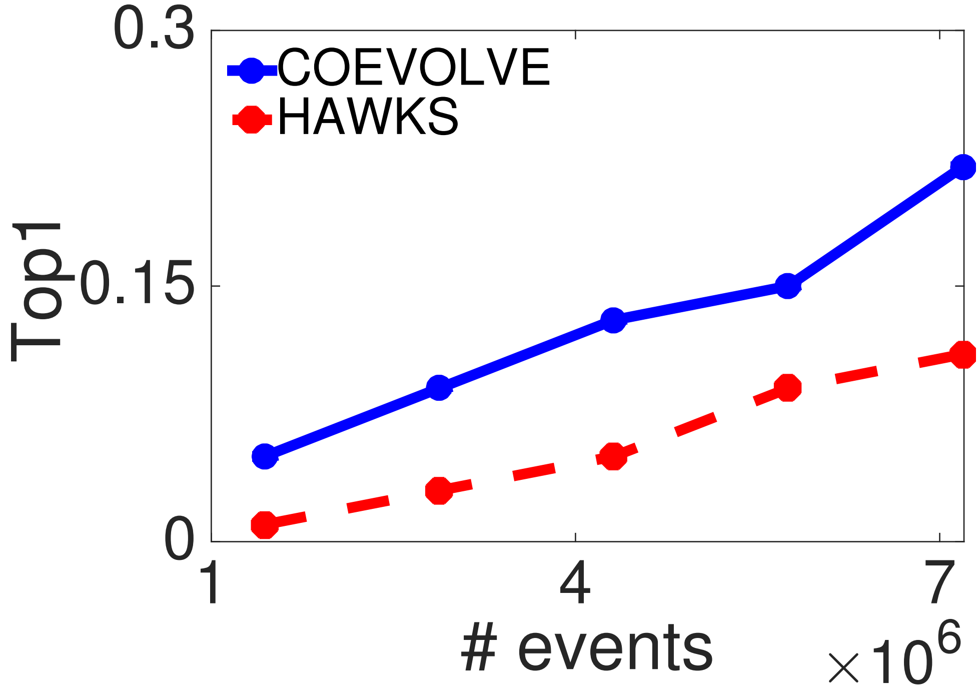
Retweet events

The quantiles of the intensity integrals $\int_{t_i}^{t_{i+1}} \lambda(t)\, dt$ computed using the fitted intensities match the quantiles of the unit-rate exponential distribution

Information diffusion prediction



Average rank



Success probability

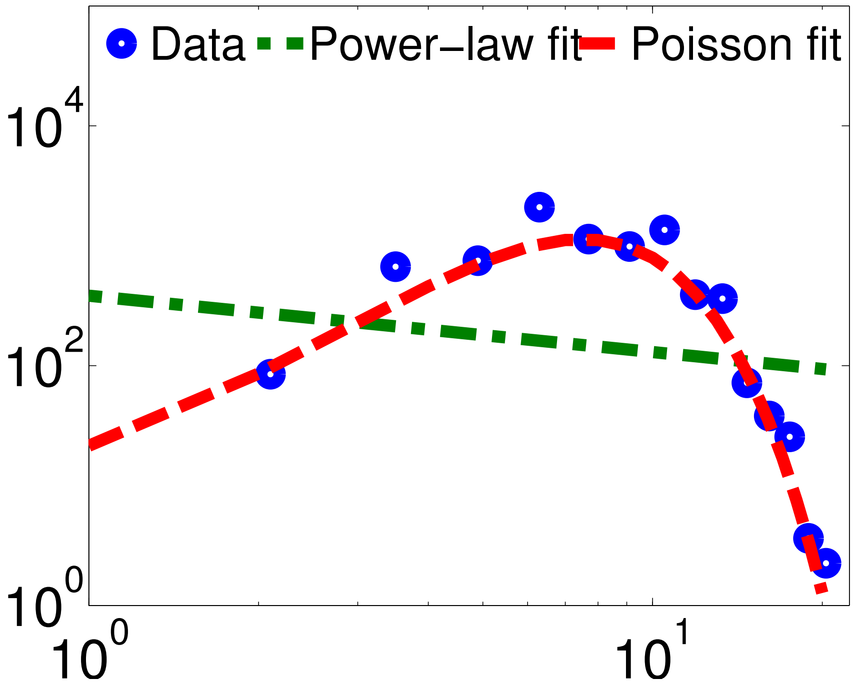
The model beats the predictions given by a standard Hawkes process

Network properties & cascade patterns

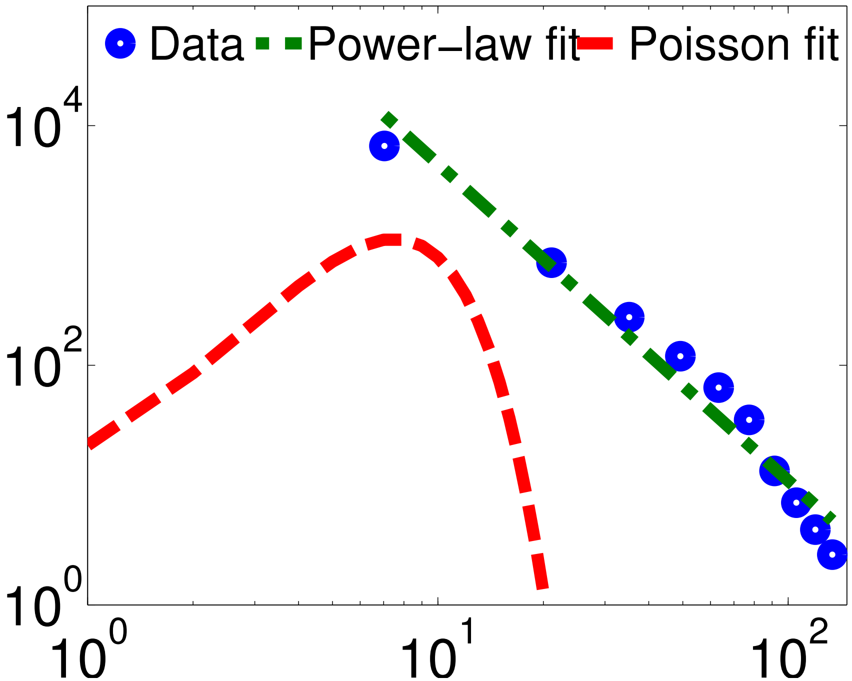
Can the model generate **realistic macroscopic static and temporal network patterns and information cascades?**

Network	Cascades
Degree distributions	**Cascade size distribution**
Network diameter	**Cascade depth distribution**
Level of triadic closure	**Cascade structure**

Degree distributions



Poisson, α = 0

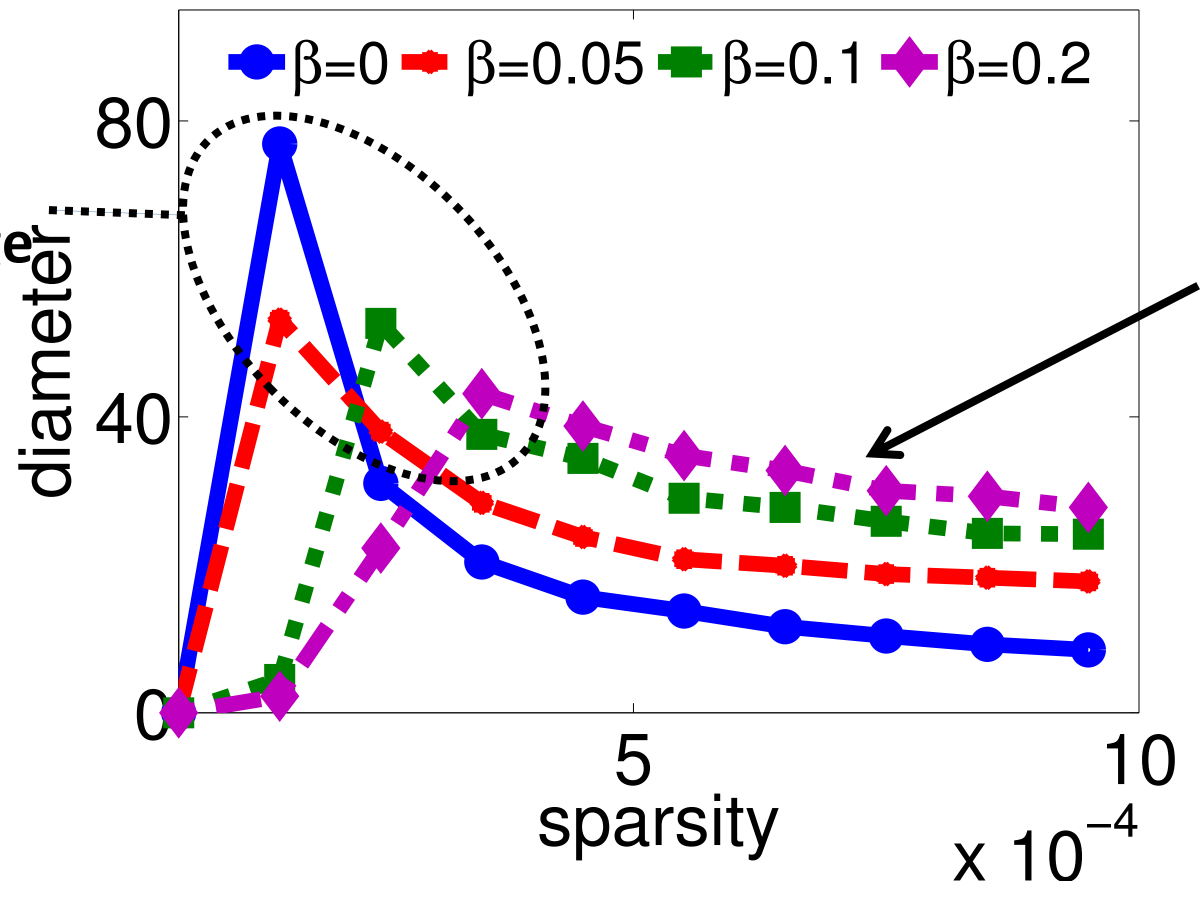


Power-law, α = 0.2

The higher the parameter α (or β), the closer the degree distribution is to a power-law

Small (shrinking) diameters

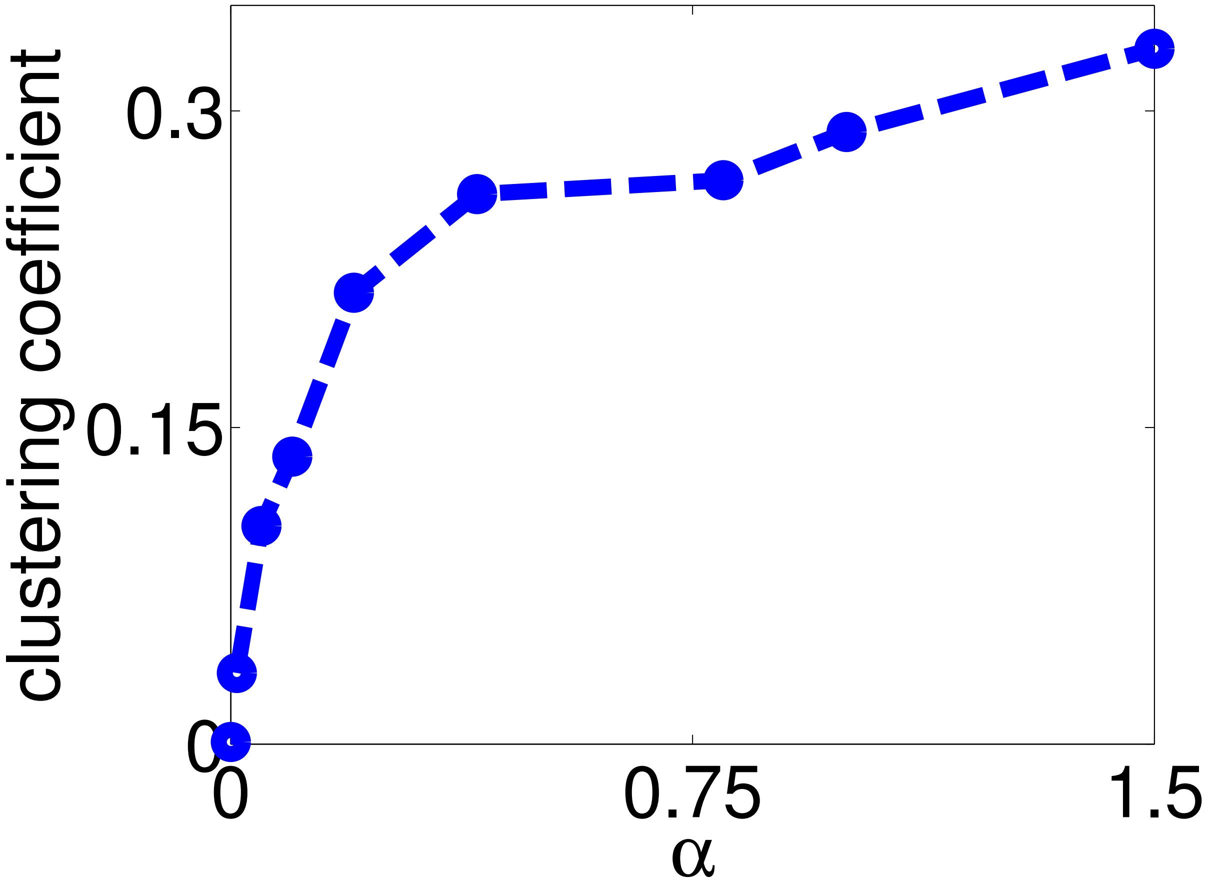
Small connected components merge



Diameter shrinks

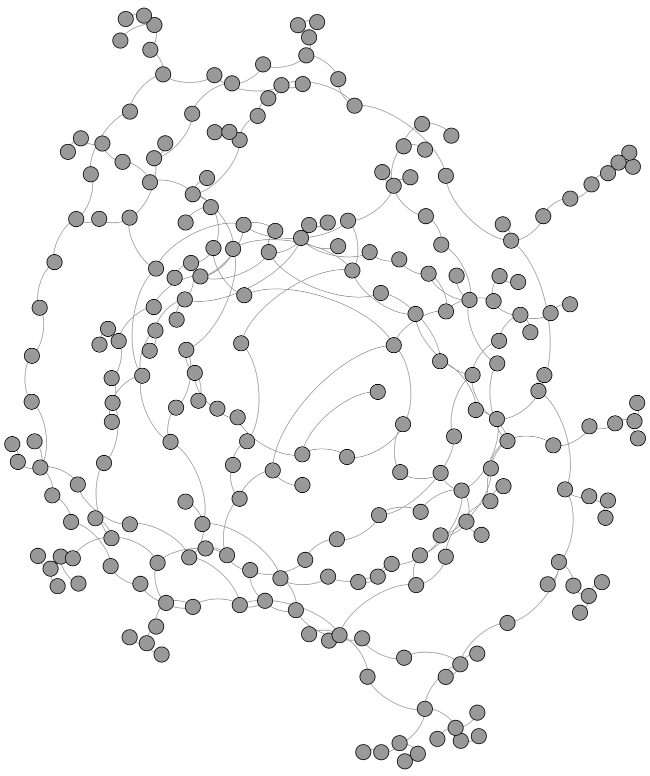
Our model generate networks with **small shrinking (or flattening) diameter over time**, as observed empirically.

Clustering coefficient

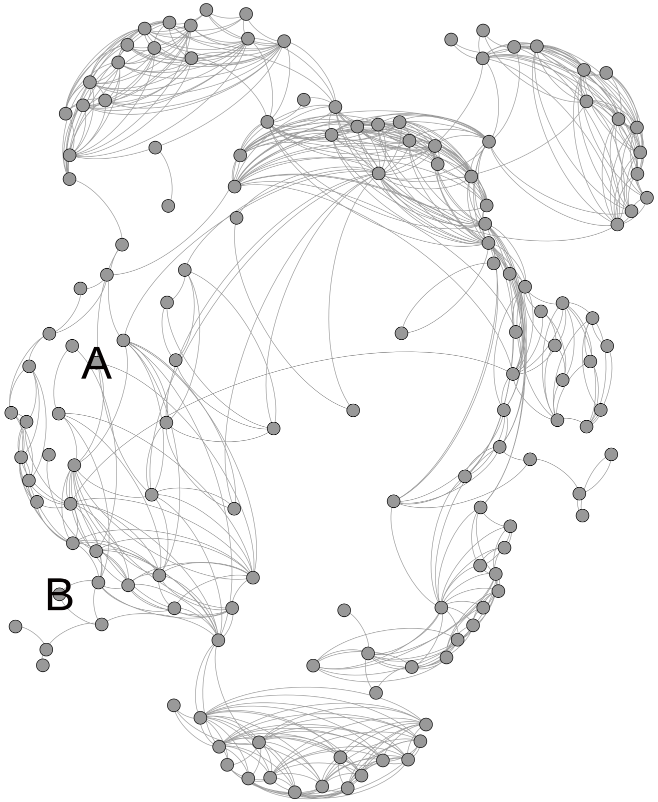


We can generate networks with **different levels of triadic closure**, as observed empirically

Different type of networks



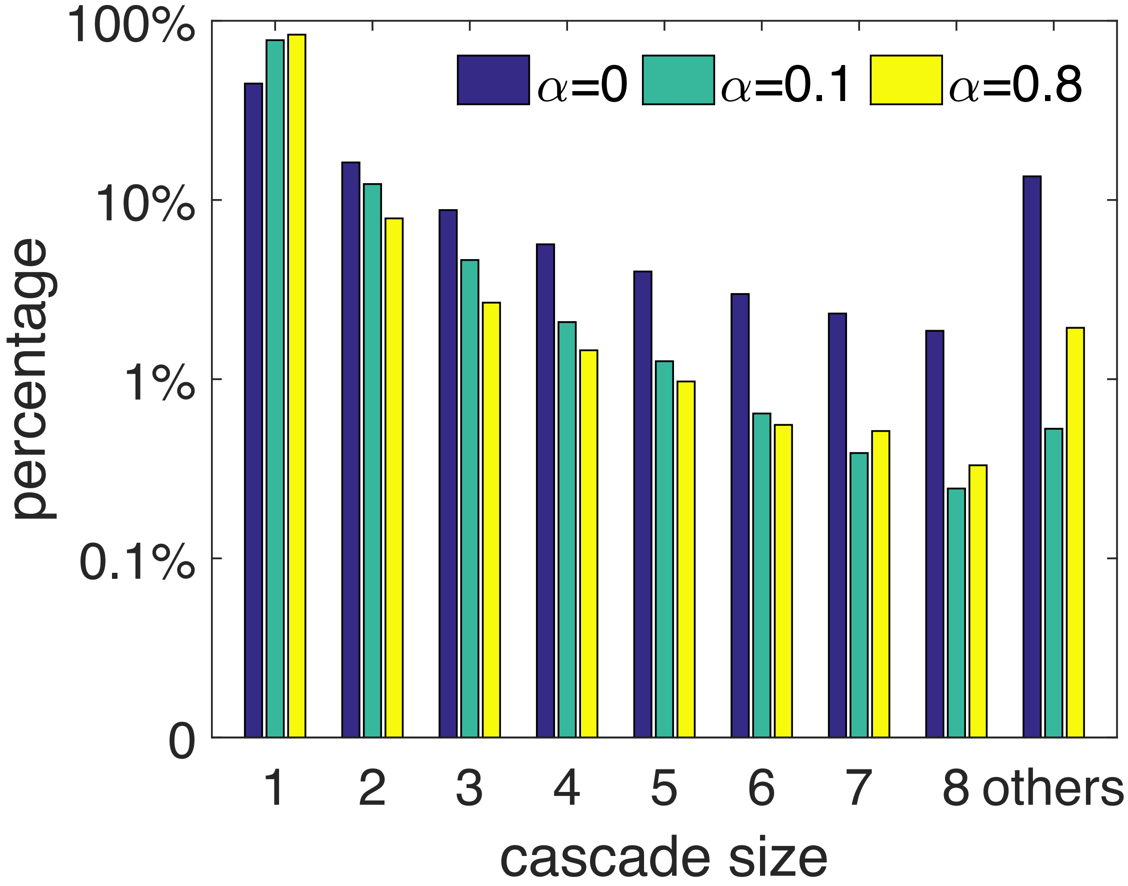
Erdos-Renyi, β = 0



Scale-free network, β = 0.8

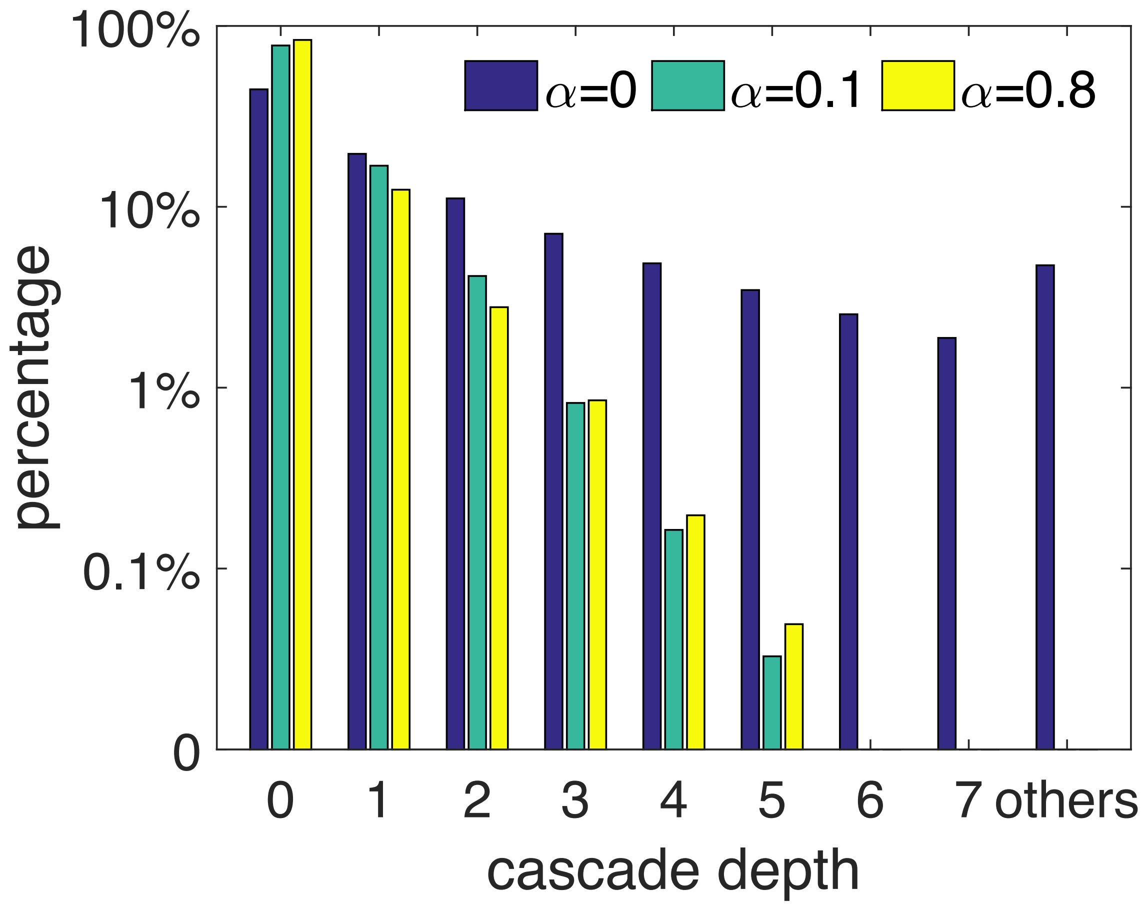
Our model allows us to generate **networks with very different structure**

Cascade patterns: size



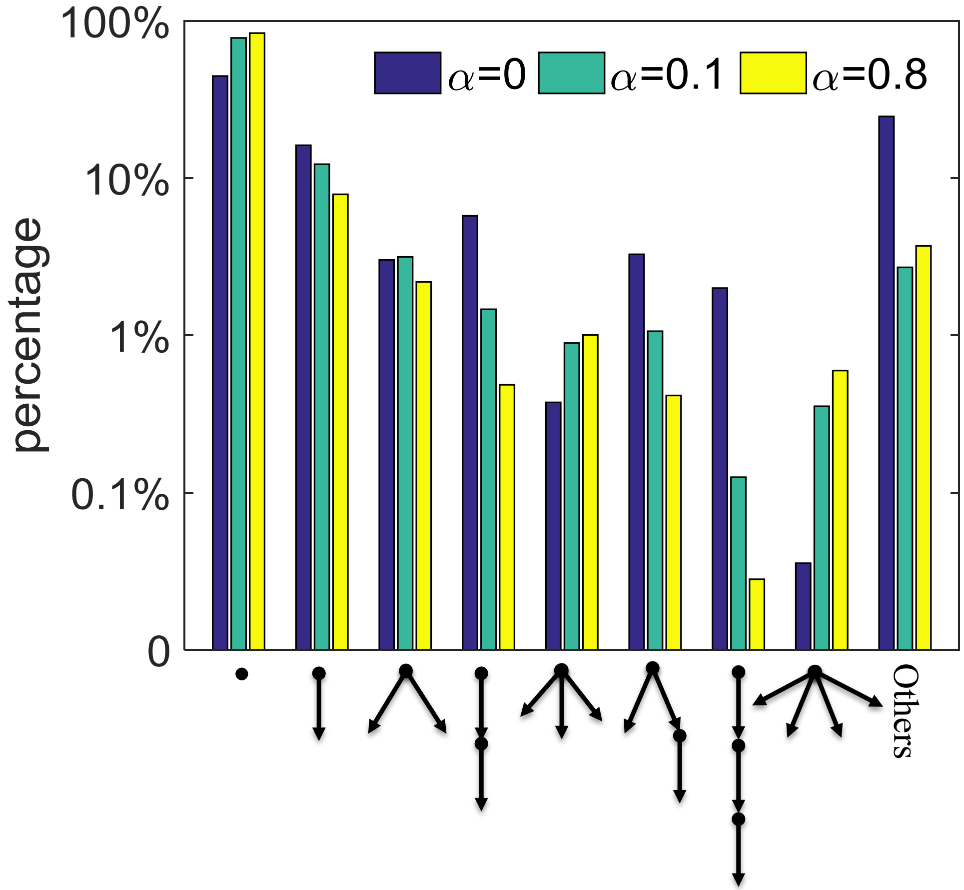
As α (or β) increases, **longer cascades become more seldom**.

Cascade patterns: depth



As α (or β) increases, **deeper cascades are more seldom**, as observed in real cascade data.

Cascade patterns: structure



The **structure** of the generated cascades becomes ***more realistic*** **as α (or β) increases**.