The Anatomy of a Scientific Rumor

Antonio Lima
Joint work with Manlio De Domenico and Mirco Musolesi

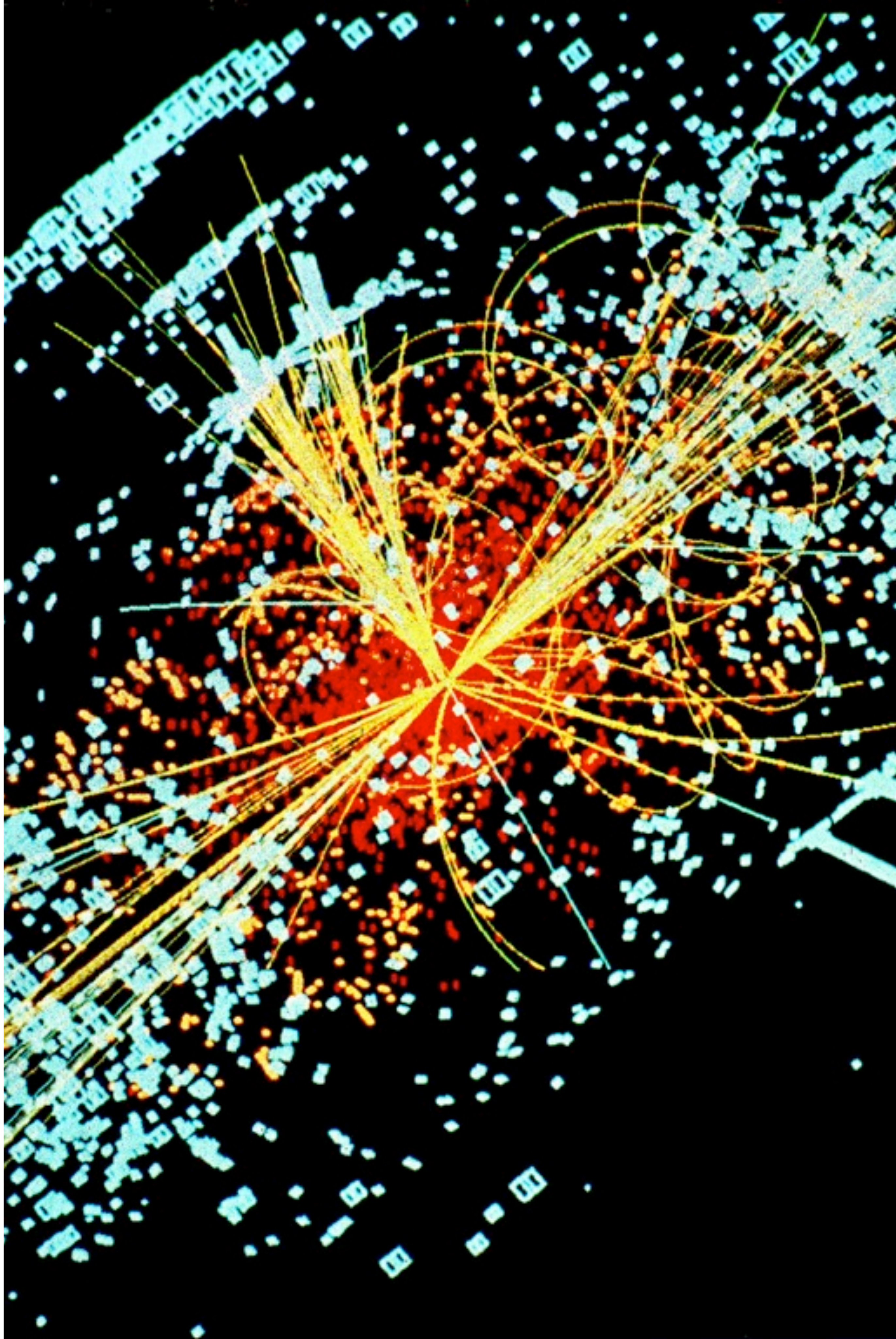
Mathematics of Networks
University of Southampton - September 16, 2013

UNIVERSITY OF BIRMINGHAM

The announcement

- On 4th July 2012 the ATLAS and CMS collaborations announced the results of the discovery of a new particle with the features of the Higgs boson.

- The world discussed the news and updates through traditional media and online social media. This represents an interesting large-scale phenomenon that can be analyzed in time and space.



Why interesting?

Why interesting?

Understanding human behavior at large-scale.

Again, why interesting?

- If you know how a piece of information spreads on a network, maybe you can study how to **control its diffusion**, range, velocity, etc.
    - Marketing. You want to maximize the diffusion of an advertisement.
    - Disaster response (disease outbreaks, earthquakes, terrorist attacks, ...).
        - In some scenarios, you might actually want to demote the info...
- On the other hand, you can **detect if something is happening** and what (trending topics), and perhaps take appropriate action.
    - Systems architecture. Systems could adapt to the context.

Twitter in a nutshell

- A user can post 140-chars messages (called *tweets*). A message can contain one or more *#hashtags* and *@mentions*, it can be a *@reply*. A user can *retweet* messages that they find interesting.

- A user can *follow* other users. Their messages show up in the *home timeline*, in descending chronological order.

The data

- We crawled from Twitter all the messages sent between July 1-7 with at least one the keywords: `lhc`, `cern`, `higgs`.
  985,590 tweets in total.

- We also crawled the social network of users 456,631 users connected by 14,855,875 edges.

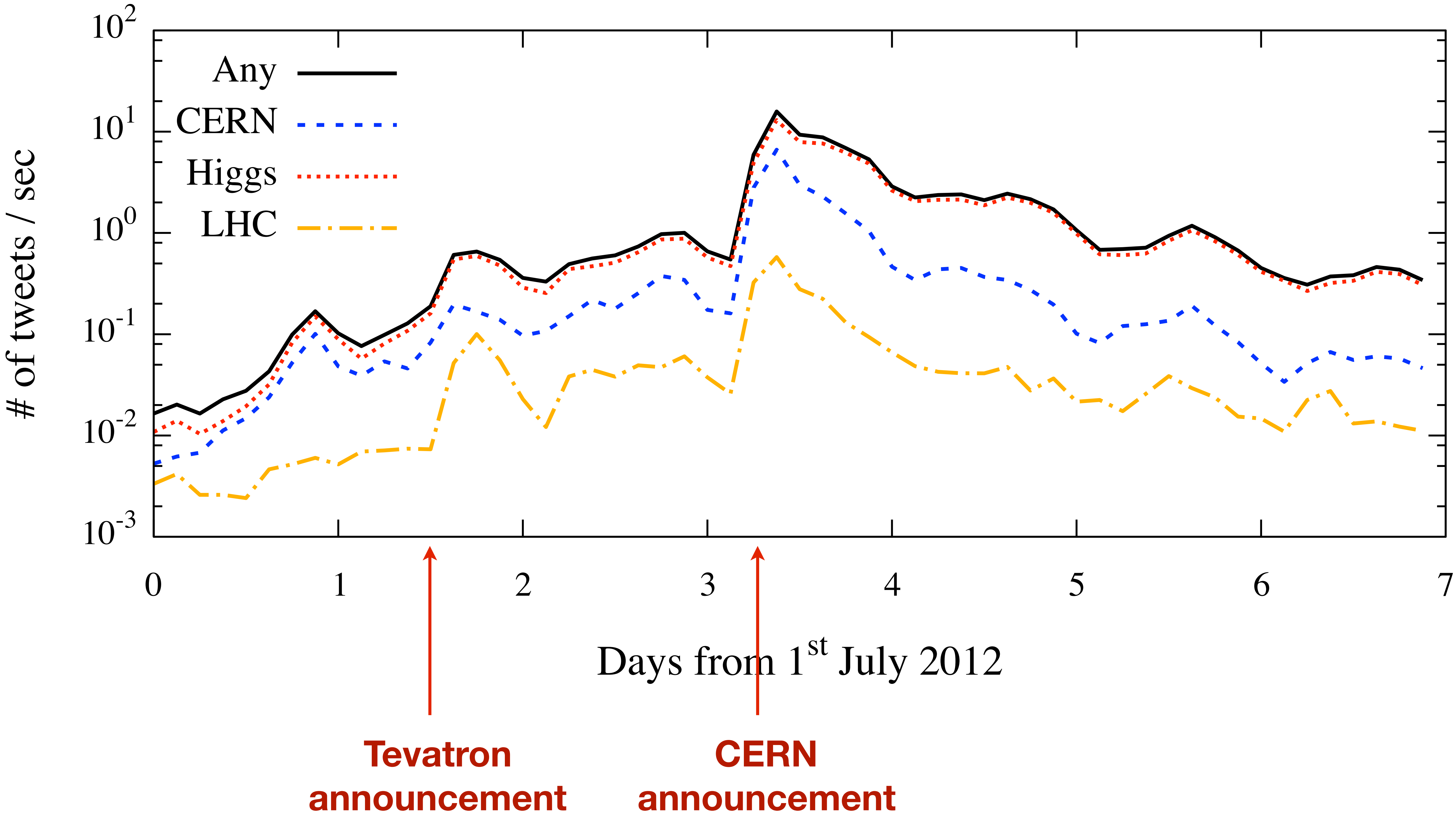
- Locations were geocoded from users' profile info.

Retweet geo-network

Each arc ranges from yellow (at the source) to red (at the destination).

UNIVERSITY OF BIRMINGHAM

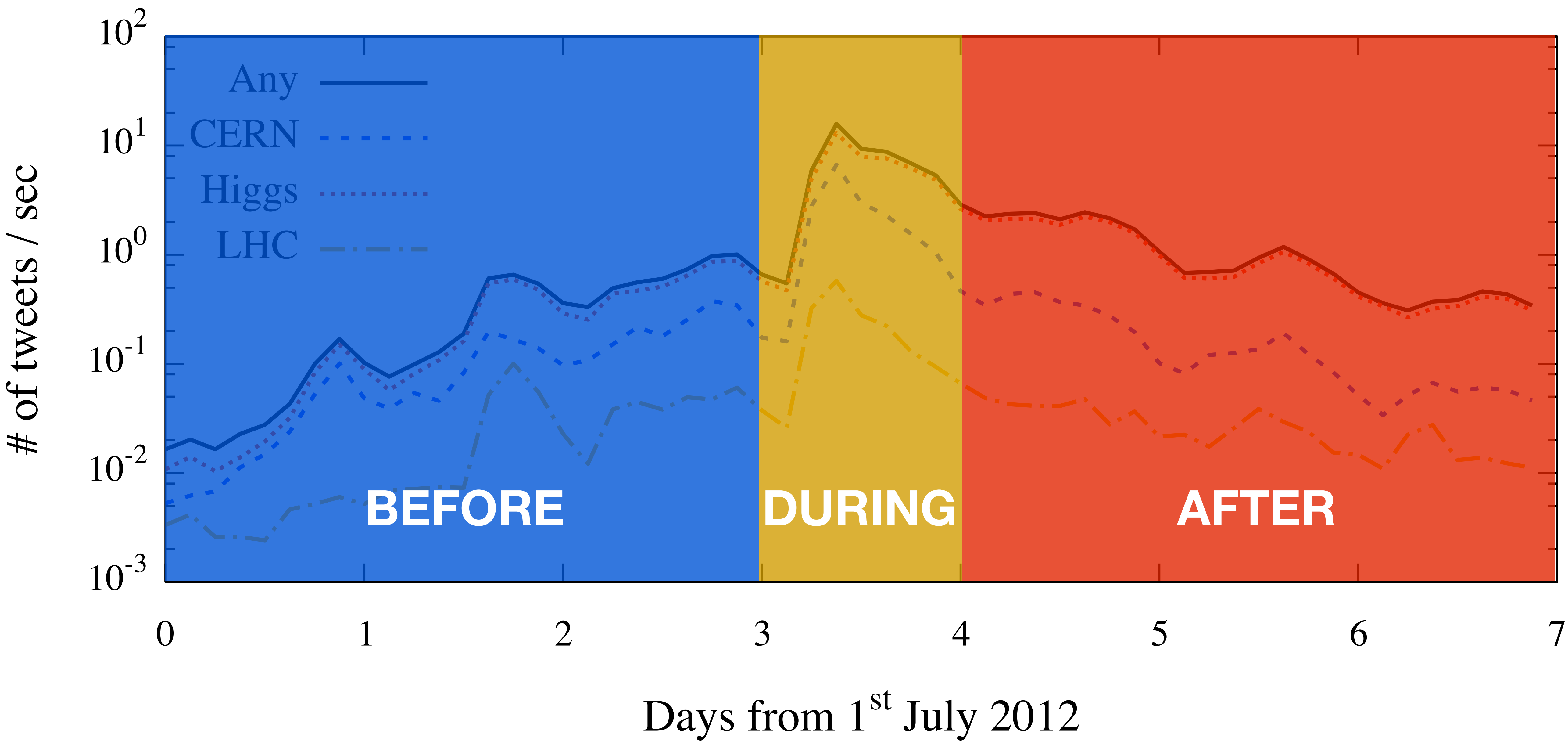
@ColinEberhardt 7141 retweets	Possibly the biggest scientific discovery of our time, the #Higgs Boson, announced in glorious MS Comic Sans Font http://t.co/jDxLkqNx
@CERN 4504 retweets	#CMS: ”we have observed a new boson with a mass of 125.3 ± 0.6 GeV at 4.9 sigma significance.” Thun- derous applause. #Higgs #ICHEP2012
@CERN 2507 retweets	CERN Press Release: CERN experiments observe particle consistent with long-sought #Higgs boson http://t.co/MBjIwytL #ICHEP2012
@timscott 2062 retweets	At the end of the #higgs announcement, one of the CERN team will pause, nonchalantly say ”oh, one more thing”, then calmly teleport away.
@ProfBrianCox 2033 retweets	Not only have taxpayers spent more on banks than we’ve spent on science since Jesus. Bob Diamond is keeping the Higgs out of the headlines!



Tweets rate

Abrupt changes can be linked to specific key-events around the announcement.

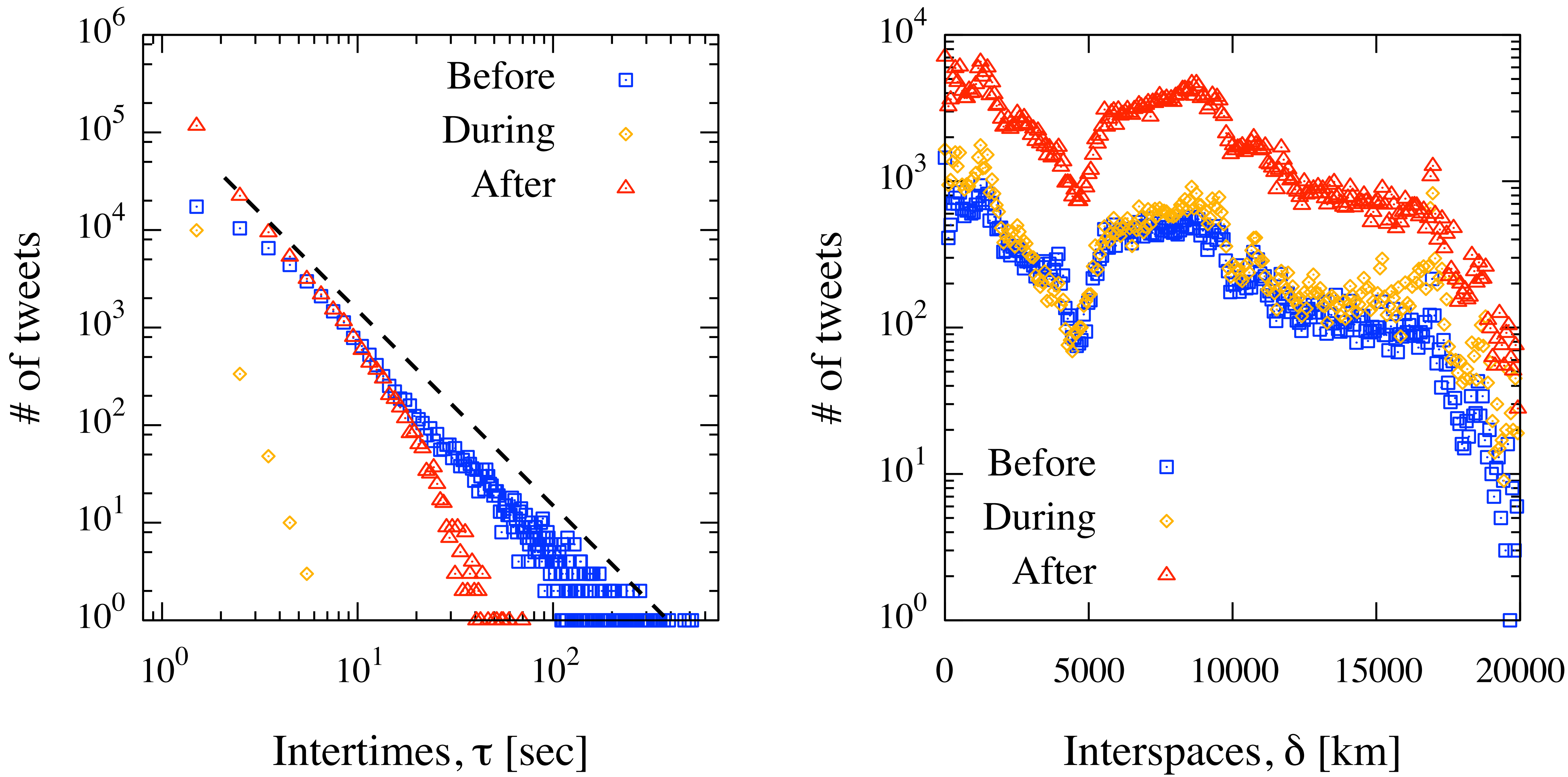
UNIVERSITY OF BIRMINGHAM



The main event: the CERN announcement

Before: July 1-3
During: July 4
After: July 5-7

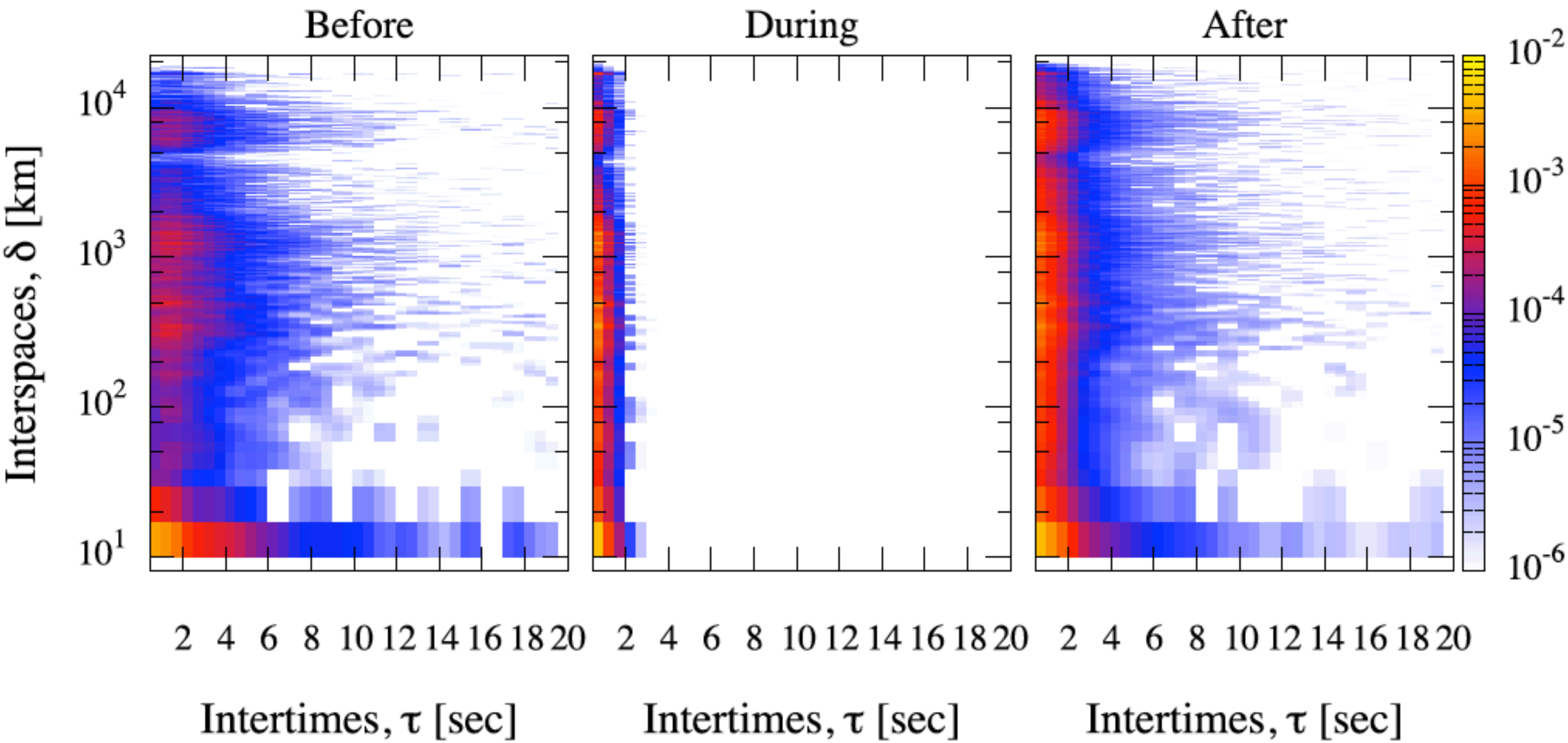
UNIVERSITY OF BIRMINGHAM



Intertime and interspace distributions

Temporal delay and spatial distance between two consecutive tweets between *any* user in the network.

UNIVERSITY OF BIRMINGHAM



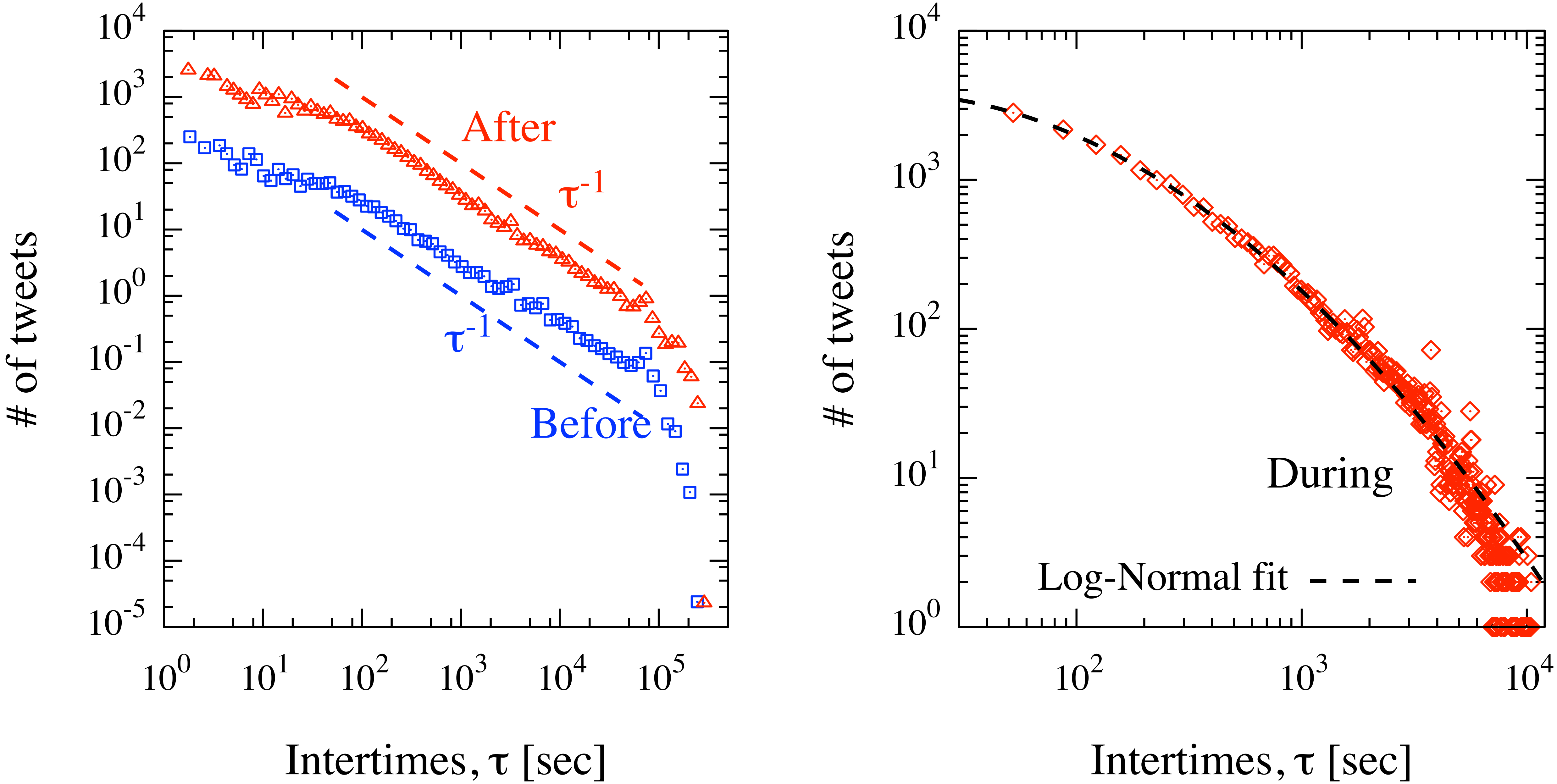
Intertimes, interspaces

Temporal delay and spatial distance between two consecutive tweets between *any* user in the network.

UNIVERSITY OF BIRMINGHAM

Retweets
by day

UNIVERSITY OF
BIRMINGHAM

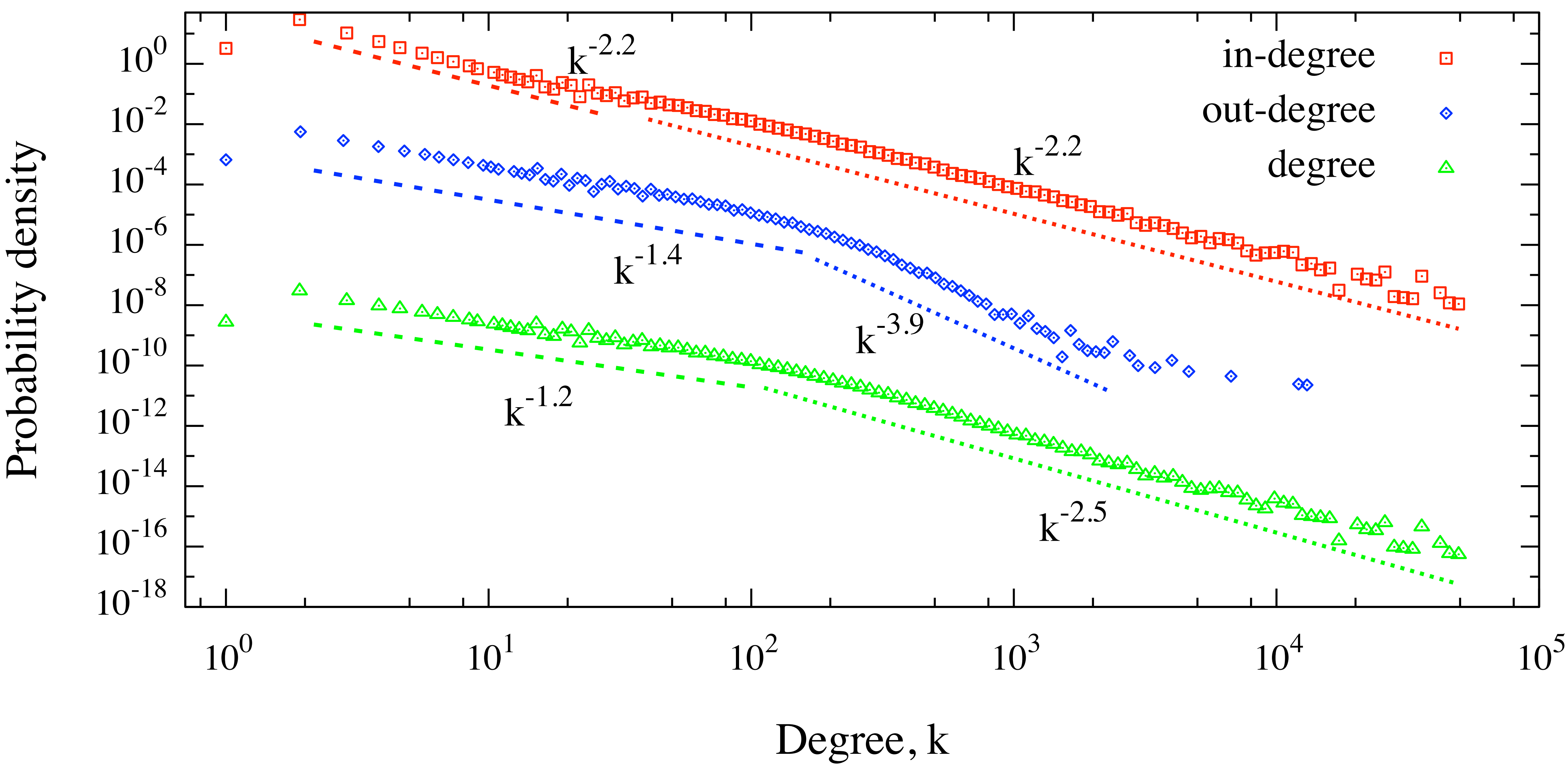


$$log(\tau) = N(\mu = 5.627 \pm 0.008, \sigma = 1.742 \pm 0.006)$$

User activity intertimes

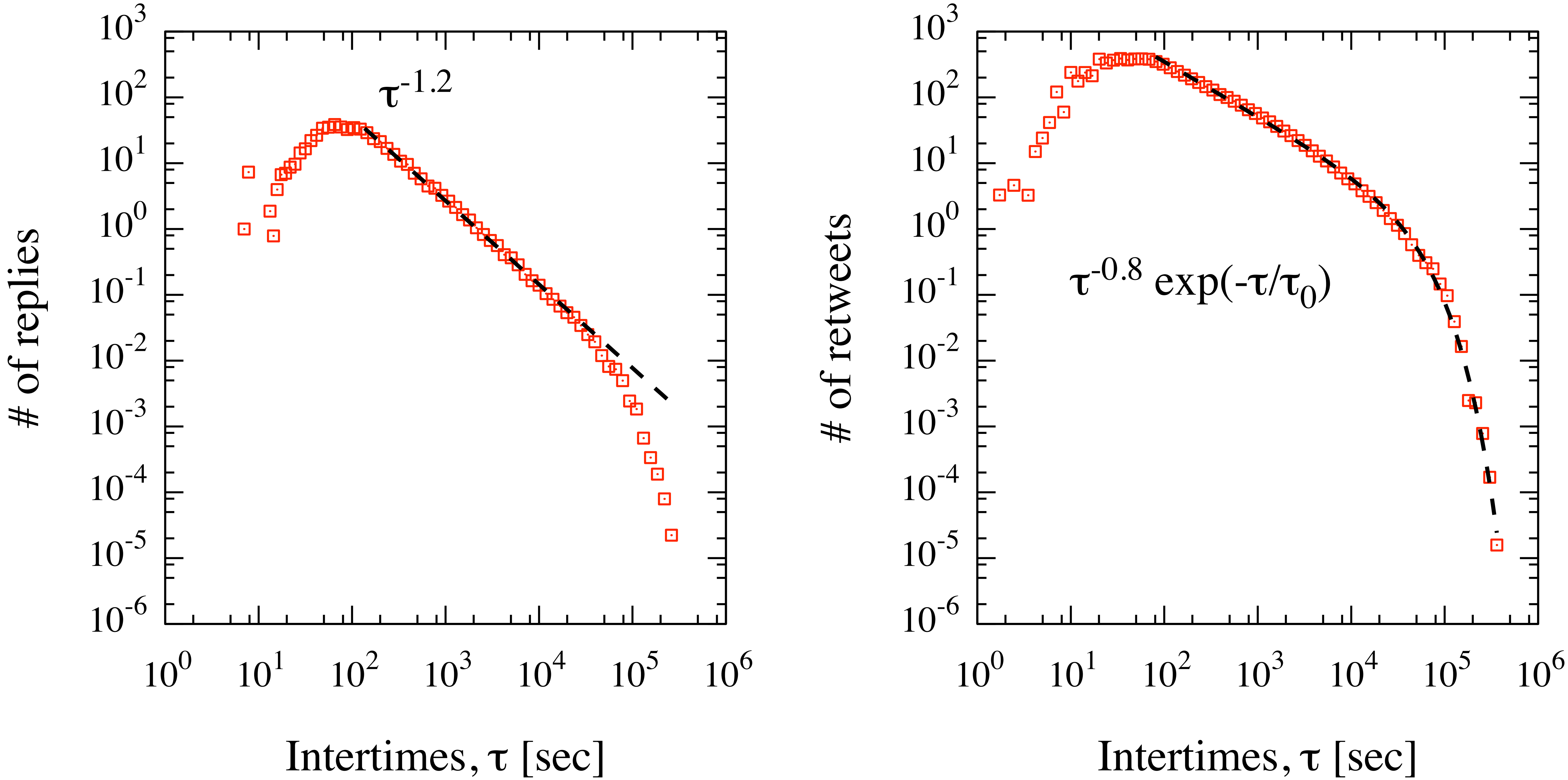
Temporal delay between two consecutive tweets between *the same* user in the network.

UNIVERSITY OF BIRMINGHAM



Degree distribution

The distributions have been shifted along the y axis to show their structure. Dashed lines are shown for guidance only.



Replies, RTs intertimes

Power-law scaling relationships with exponent < 1 occur in nature only with cut-off. Even in those cases, they are very rare.

UNIVERSITY OF BIRMINGHAM

Rumor spreading

- We investigate how the "Higgs boson rumor" spreads in the social network.
- Different from disease epidemics, we cannot use the unstructured SI model.
- We distinguish between two different states for users in the network. A user is either *active* or *non-active*.
- $A(t), D(t)$ represent the number of active and non-active users, respectively.
- $a(t), d(t)$ represent the fraction of active and non-active users, respectively.

$$A(t) + D(t) = N$$

Model without social ties, without de-activation

- As a first step, we **ignore social ties**.
- Once a user has tweeted, he/she is considered **active**. Hence, the number of active users is monotonically non-decreasing.
- We model the number of new activations as proportional to the number of users who haven't been activated yet. More formally:

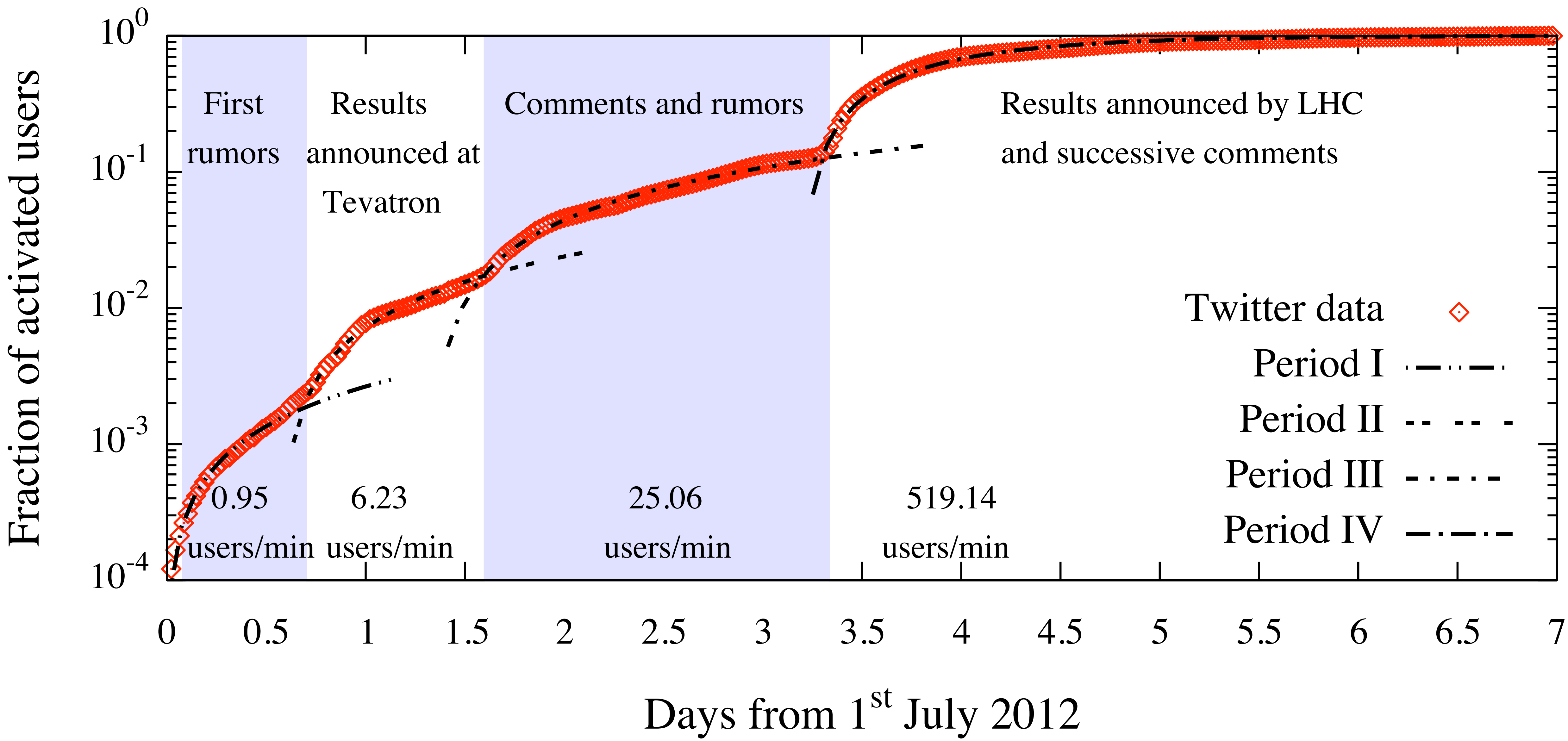
$$A(t + \Delta t) = A(t) + \lambda[N - A(t)]\Delta t$$

$$\frac{da(t)}{dt} = \lambda[1 - a(t)]$$

constant activation rate

$$a(t) = 1 - [1 - a(t_k)]e^{-\lambda(t - t_k)}$$

starting time of period *k*

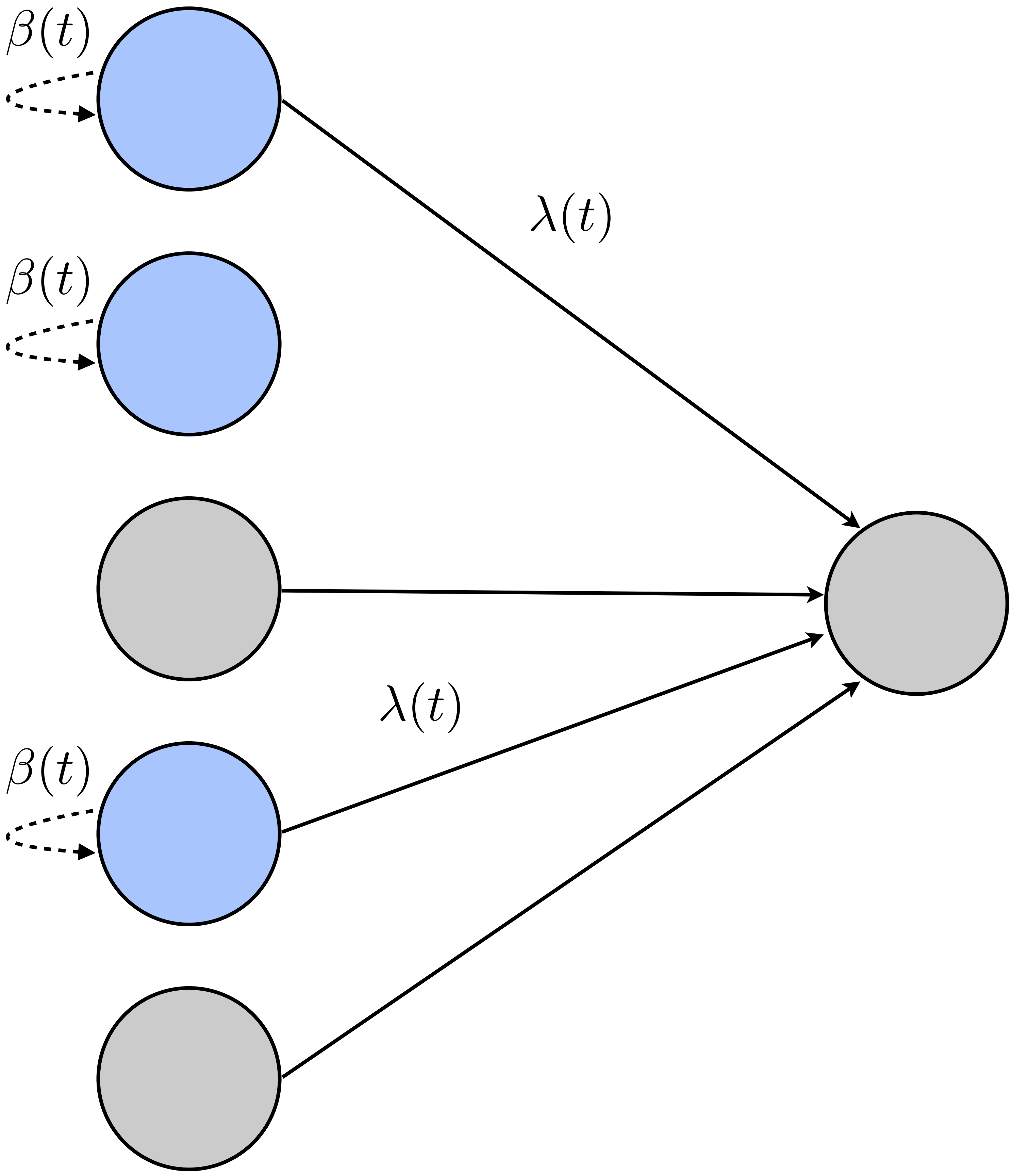


Model w/out social ties, w/out de-activation

Red dots represent real data, black lines represent model data.

Model with social ties, with de-activation

- We now build a model that takes into consideration **social cascading**.
- In this refined model, a user that tweets about the Higgs boson is **active**, then can become **non-active** after a certain amount of time.
- **Activation**: an active user can make each of their followers active with probability rate $\lambda(t)$.
- **Deactivation**: an active user can become non-active with probability rate $\beta(t)$. This models the time-limited visibility of tweets (newer tweets replace old ones, users read only most recent tweets). Non-active users cannot activate their followers.


$\beta(t)$
$\lambda(t)$
$\beta(t)$
$\lambda(t)$
$\beta(t)$

Model with social ties, with de-activation

- At time t , the probability that a non-active user, connected to j_a active users, becomes active (i.e. is activated by at least one of them) is:

$$p_\lambda(t; j_A) = 1 - [1 - \lambda(t)]^{j_A}$$

- In general, the probability that a non-active user is connected to a certain number of active users depends on degree correlations, i.e., the probability of observing a vertex with out-going degree *k_out* connected to a vertex with in-going degree *k_in*.

- However, it has been shown [Boguná et al., PRL 2003] that for *pure* scale-free networks with exponent between 2 and 3 degree correlations do not affect the spreading dynamics. We use this simplifying assumption for this network, exhibiting scale-free degree of 2.5 for k>200.

Model with social ties, with de-activation

- At time t, the probability that a non-active user *with a certain in-degree* is connected to j_A active users (with *any* out-going degree) is:

$$\tilde{p}(t; j_A, k^{in}) = \frac{\binom{A(t)}{j_A}\binom{N-A(t)-1}{k^{in}-j_A}}{\binom{N-1}{k^{in}}},$$

which accounts for all the ways you can arrange j_A users in k^{in} places.

- Hence, the probability of activation of a user with a certain in-degree is:

$$P_{\lambda,k^{in}}(D \to A) = \sum_{j_A=1}^{k^{in}} \tilde{p}(t; j_A, k^{in}) p_\lambda(t; j_A).$$

Model with social ties, with de-activation

- By summing this probability over the in-degree distribution, we obtain the probability of activation at time t.

$$\Theta_\lambda(t) = \sum_{k^{in}} \mathcal{P}(k^{in}) P_{\lambda,k^{in}}(D \to A)$$

- Finally, we use this to derive a discrete activation model:

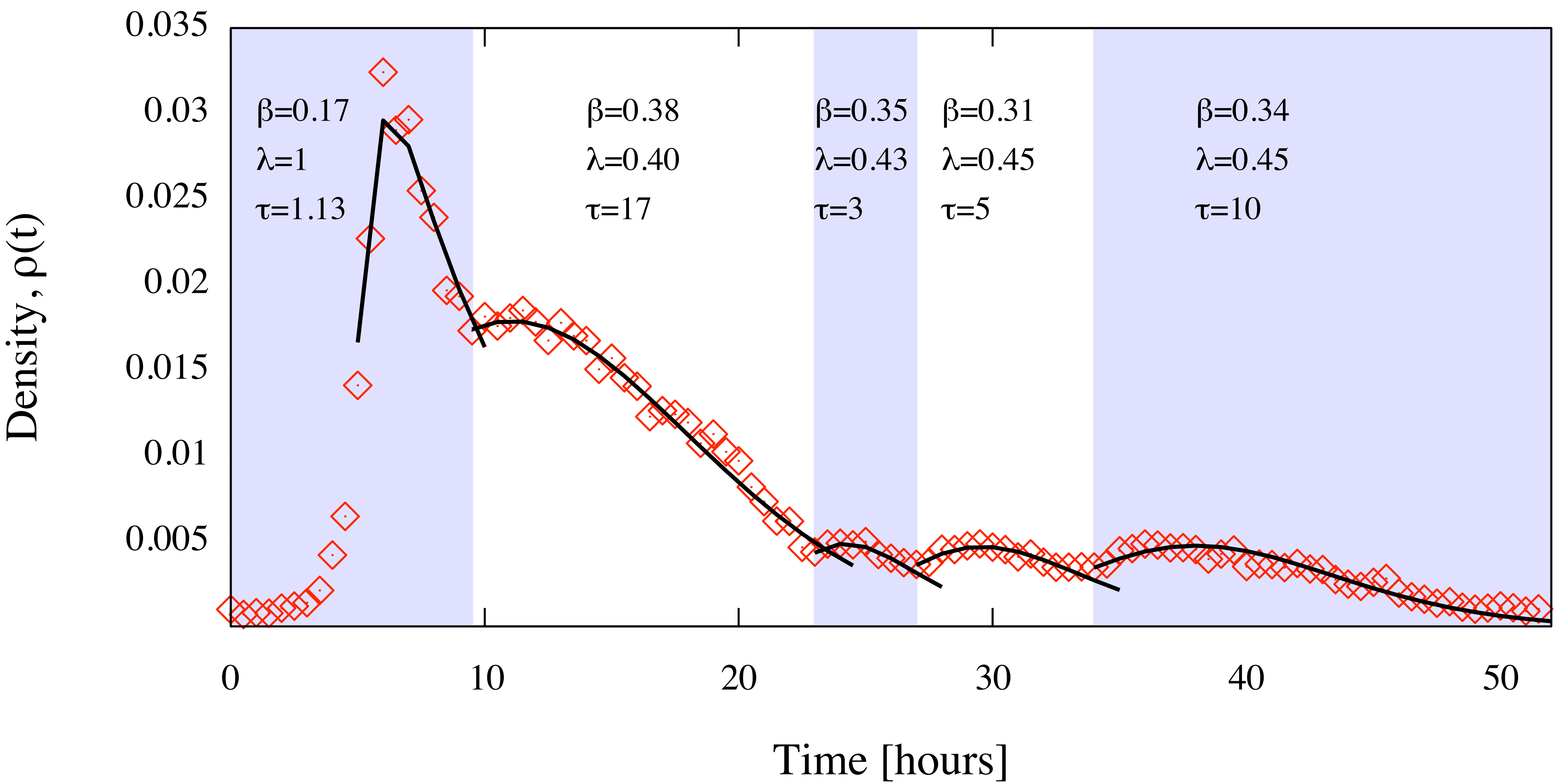
$$A(t+1) = (1-\beta(t))A(t) + (N - A(t))\Theta_{\lambda(t)}(t),$$

Model with social ties, with de-activation

- In order to account for the **decreasing interest** on a topic over time, we model the activation rate as time-varying, with an exponential decay with time scale $\tau = 1/\xi$

$$\lambda(t+1) = (1-\xi)\lambda(t)$$

- Finally, we perform large-scale Monte-Carlo simulations over the parameters, to fit our model with the observed data.



Real data and model

Red dots represent real data, solid lines represent model data. Parameters were chosen so that chi-squared is minimized.

UNIVERSITY OF BIRMINGHAM

Thanks!

Questions?

Antonio Lima

a.lima@cs.bham.ac.uk
http://cs.bham.ac.uk/axl162
@themiurgo

UNIVERSITY OF BIRMINGHAM