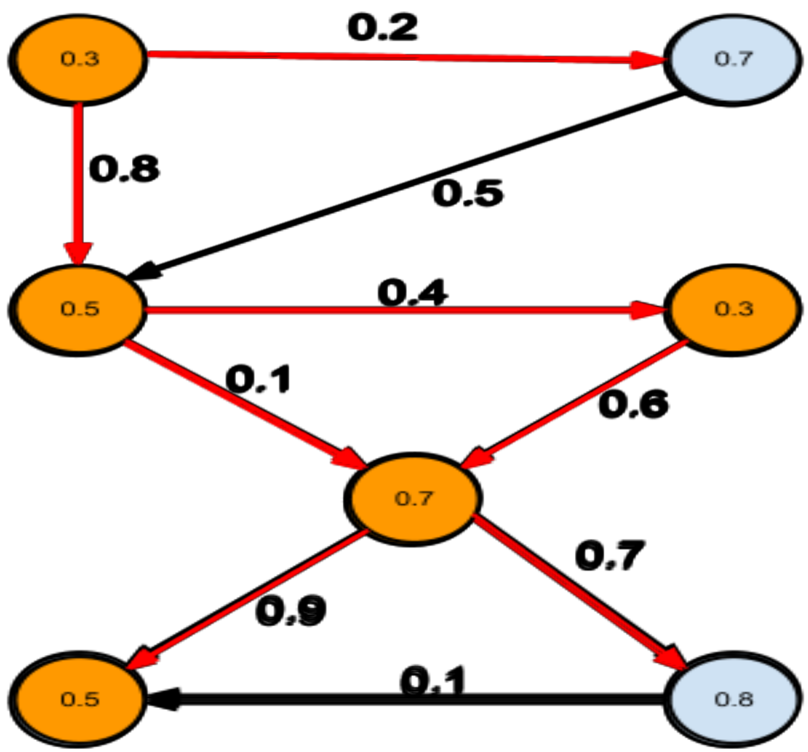
Co-evolutionary dynamics in social networks: A case study of Twitter

Demetris Antoniades
danton@cs.ucy.ac.cy
Laboratory of Internet Computing
Computer Science Department
University of Cyprus

joint: **Constantine Dovrolis** constantine@gatech.edu
College of ComputingGeorgia Institute of Technology

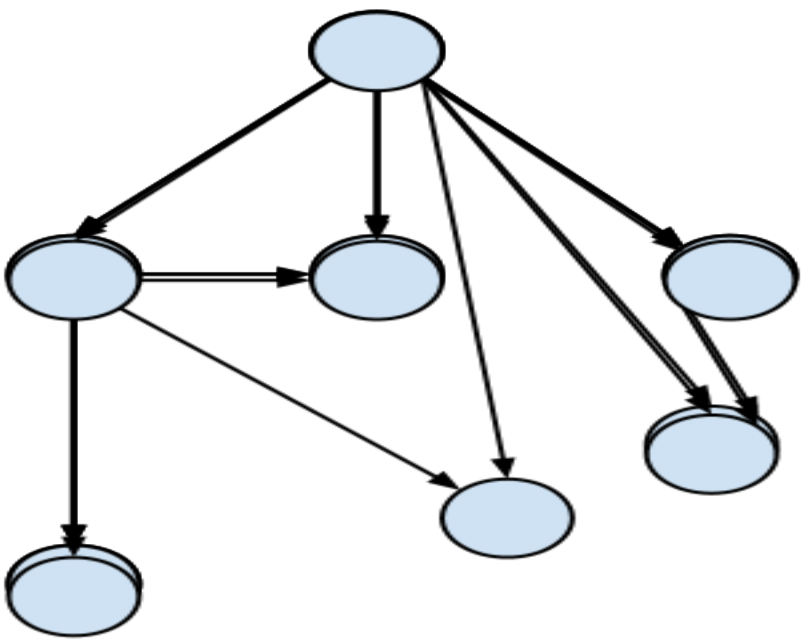
Dynamics *on* networks

- Example: information diffusion with a threshold model (similar to Granovetter's)



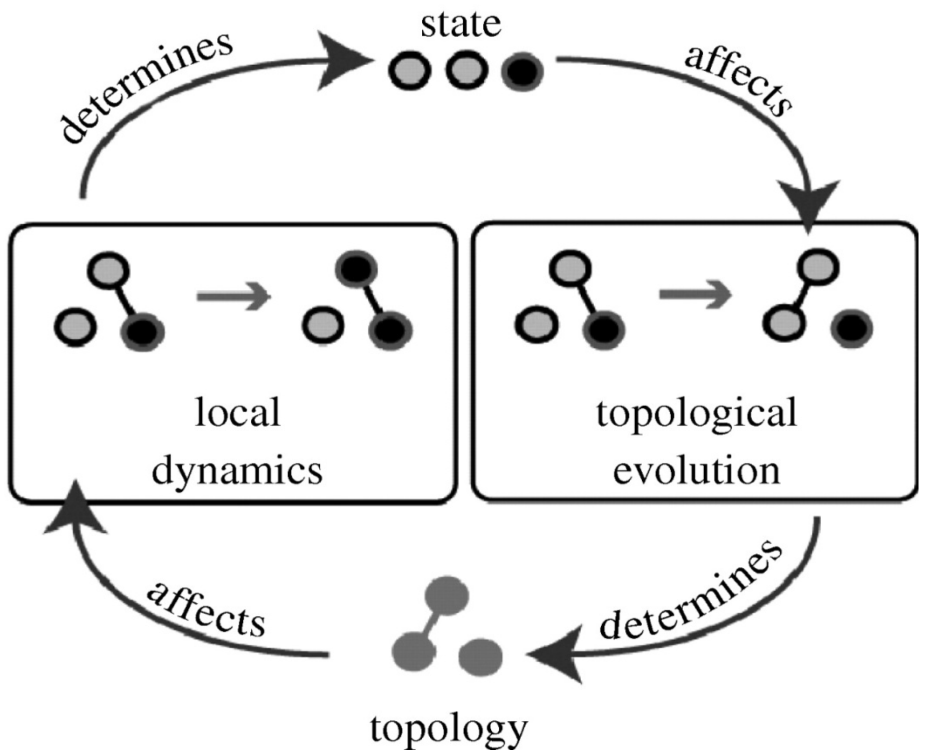
Dynamics *of* networks

- Example: Preferential-attachment model



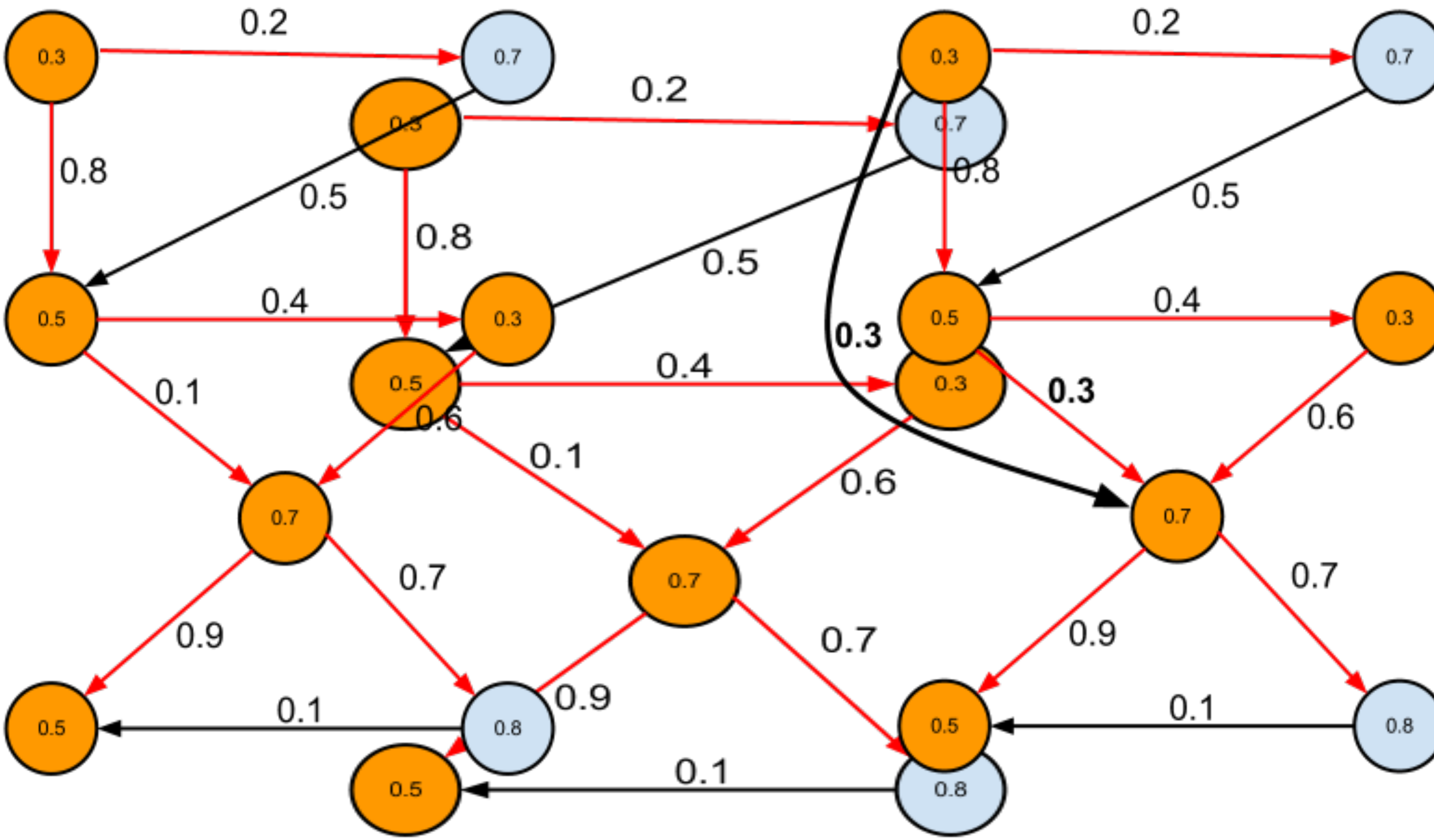
Co-evolutionary dynamics

- Coupled dynamics *ON and OF* networks



"Adaptive Coevolutionary Networks: A Review", Thilo Gross and Bernd Blasius, Journal of the Royal Society: Interface 5, 259-271, 2008

Social networks are co-evolutionary



Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network

- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - TRF events Vs. exogenous new followers
  - A data collection methodology for TRF events
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events

- Next steps

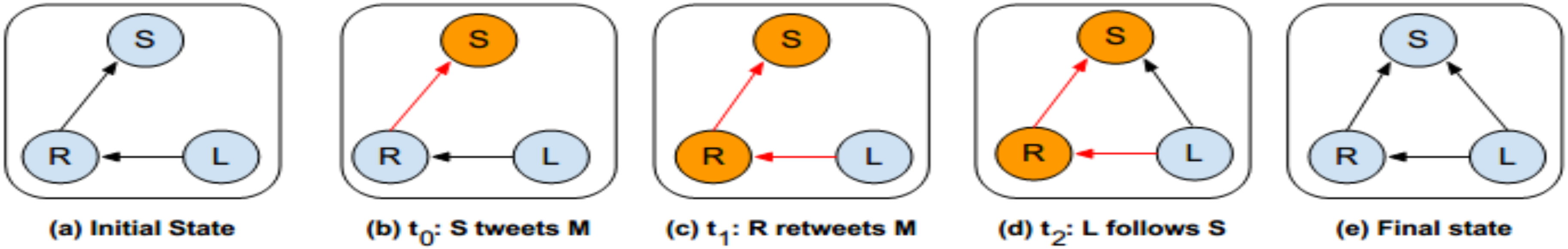
Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network

- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - TRF events Vs. exogenous new followers
  - A data collection methodology for TRF events
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events

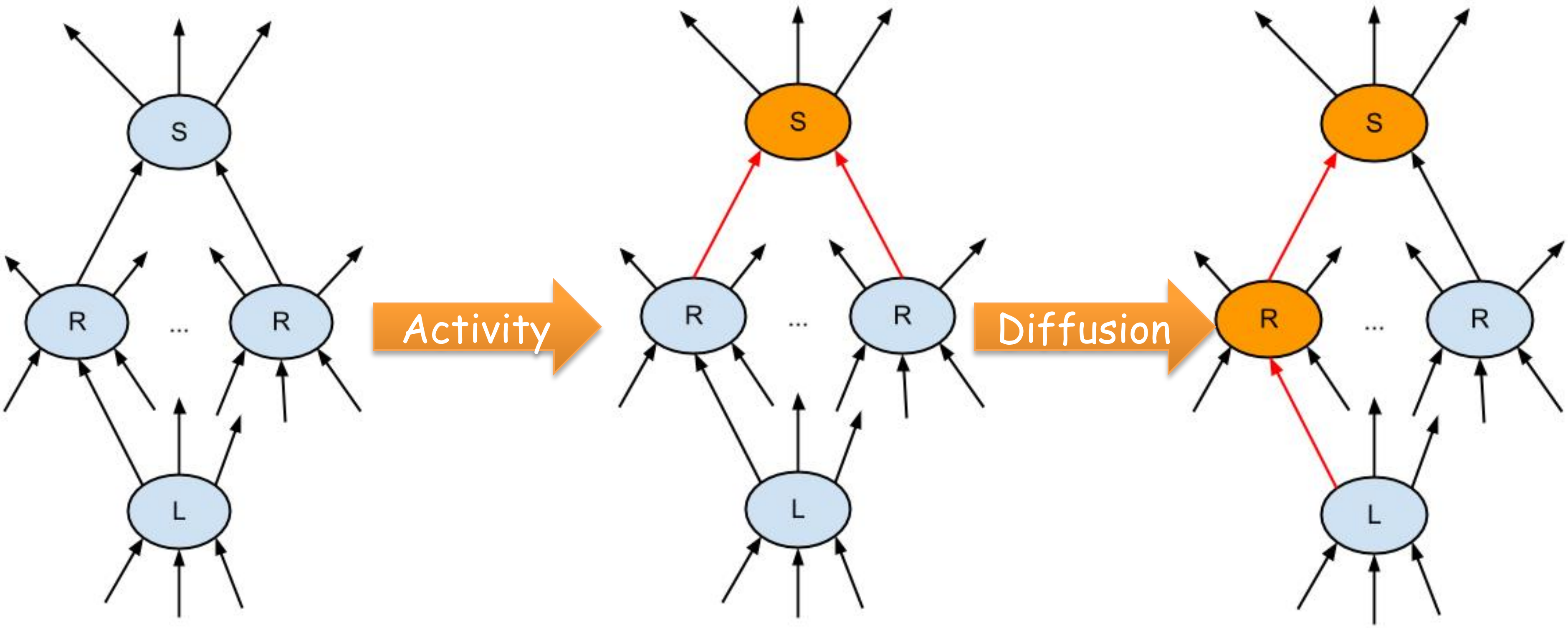
- Next steps

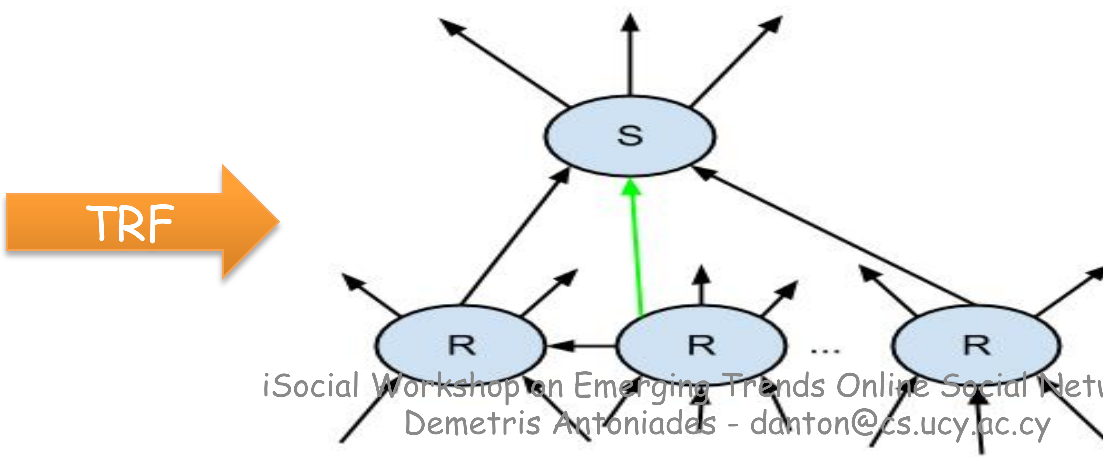
Case study: Twitter



- Tweet-Retweet-Follow (TRF) events
  - Info diffusion (retweets) leads to new followers
  - Clear case of co-evolutionary dynamics

Tweet-Retweet-Follow (TRF)





Definition of TRF event

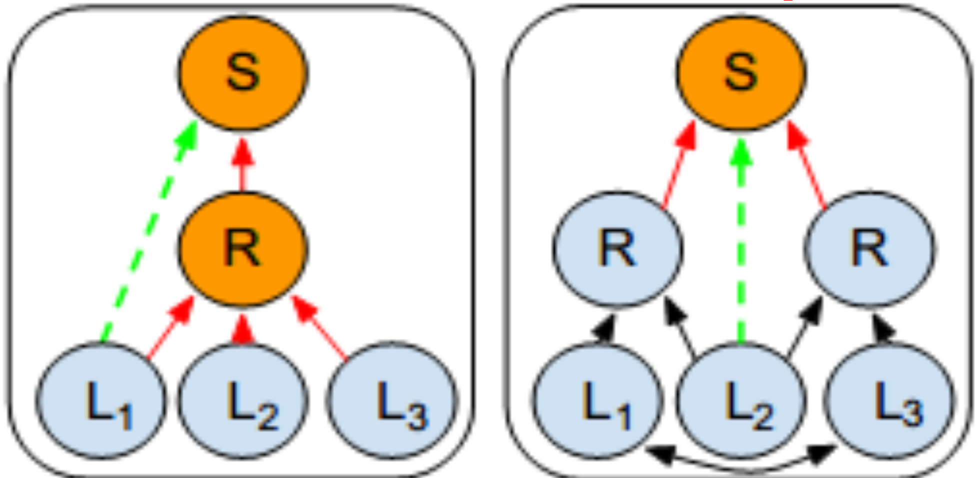
- A Tweet-Retweet-Follow event
  - Speaker S,
  - Repeater R,
  - Listener L

- Occurs when:
  a) S tweets a message M at time t_0
  b) R retweets M at some time $t_1 > t_0$
  c) A follower L of R follows S within Δ hours from t_1

Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network
- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - TRF events Vs. exogenous new followers
  - A data collection methodology for TRF events
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events
- Next steps:

Does receiving a retweet increase probability of a new follower link?
(compared to not receiving a retweet)



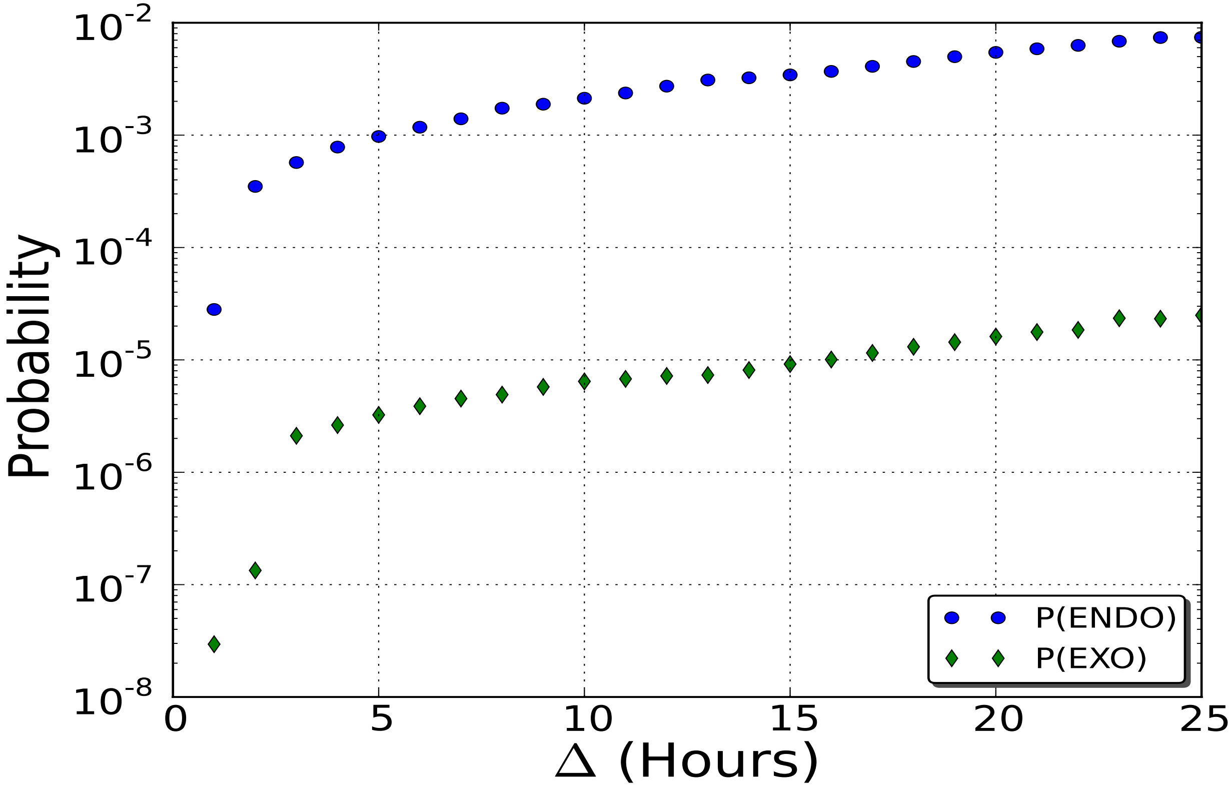
(a) TRF event: Follow after receiving a Retweet
(b) TWF event: Follow without receiving a Retweet

- Control for local structure
- Examine the probability for a new follower in a time window Δ

Does receiving a retweet increase probability of a new follower link? (compared to not receiving a retweet)

- Continuously monitored 200 users for a period of 10 days
  - Periodically collecting F(S) every 30 minutes
    - Also collecting F(F(S)) for each follower of S
  - 4,945 new follow relationships observed during this period
    - 42% of which were Endogenous followers

Effect of receiving (or not receiving) a retweet

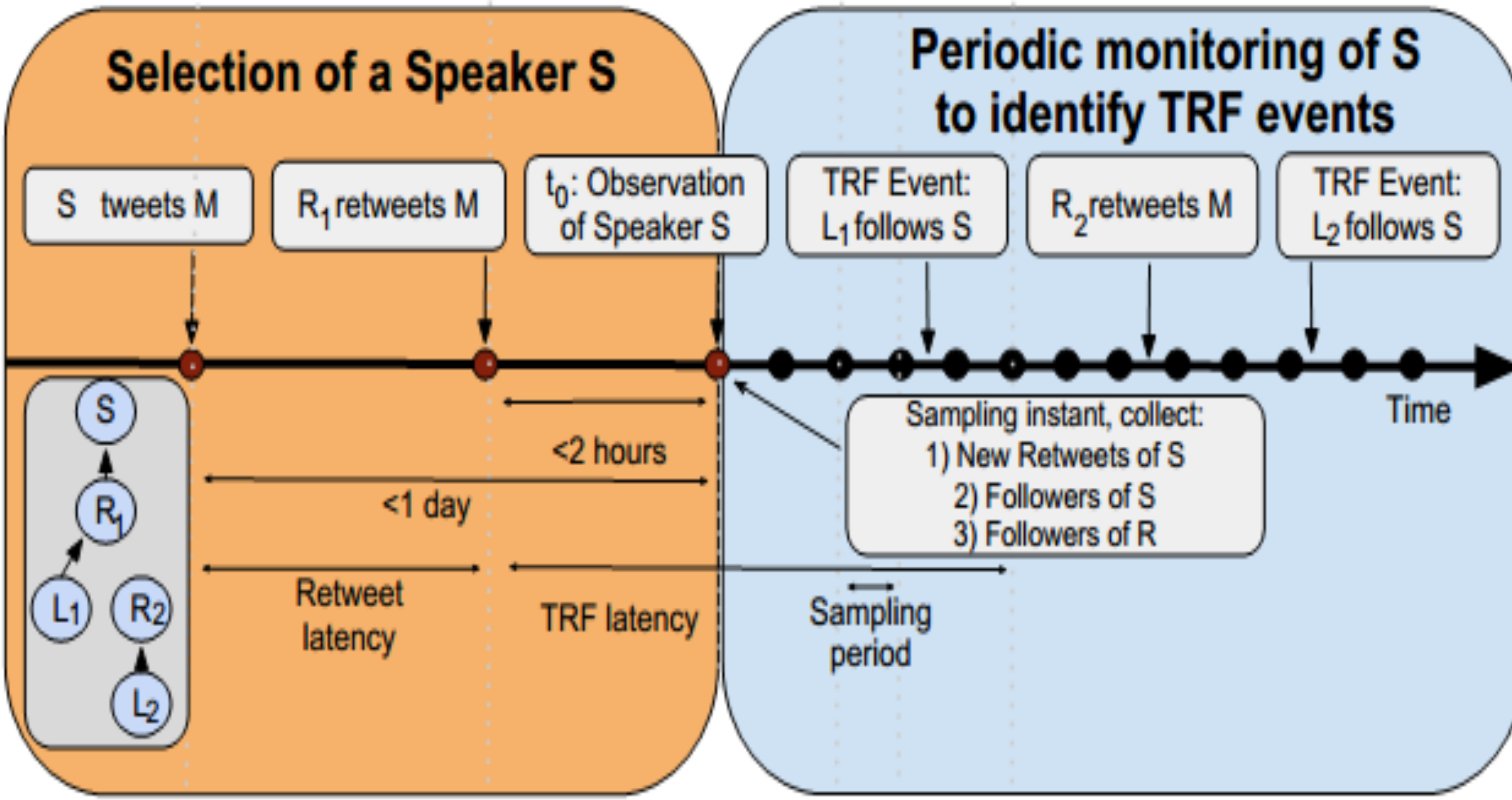


- TRF events 3 orders of magnitude more likely than TF events

Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network
- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - TRF events Vs. exogenous new followers
  - A data collection methodology for TRF events
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events
- Next steps:

Data collection methodology



Selection of Speaker S

- Obtain a number of active Twitter users
  - Get 20 most recent tweets as returned from http://www.twitter.com/search
- Mark as Speaker if:
  - Tweet during last 24 hours
  - At least one retweet during last 2 hours
- Collect
  - F(S, t), F'(S, t), F(R, t), F'(R, t)
  - Creation time, location information, number of statuses for both S and R

Monitoring of Speakers

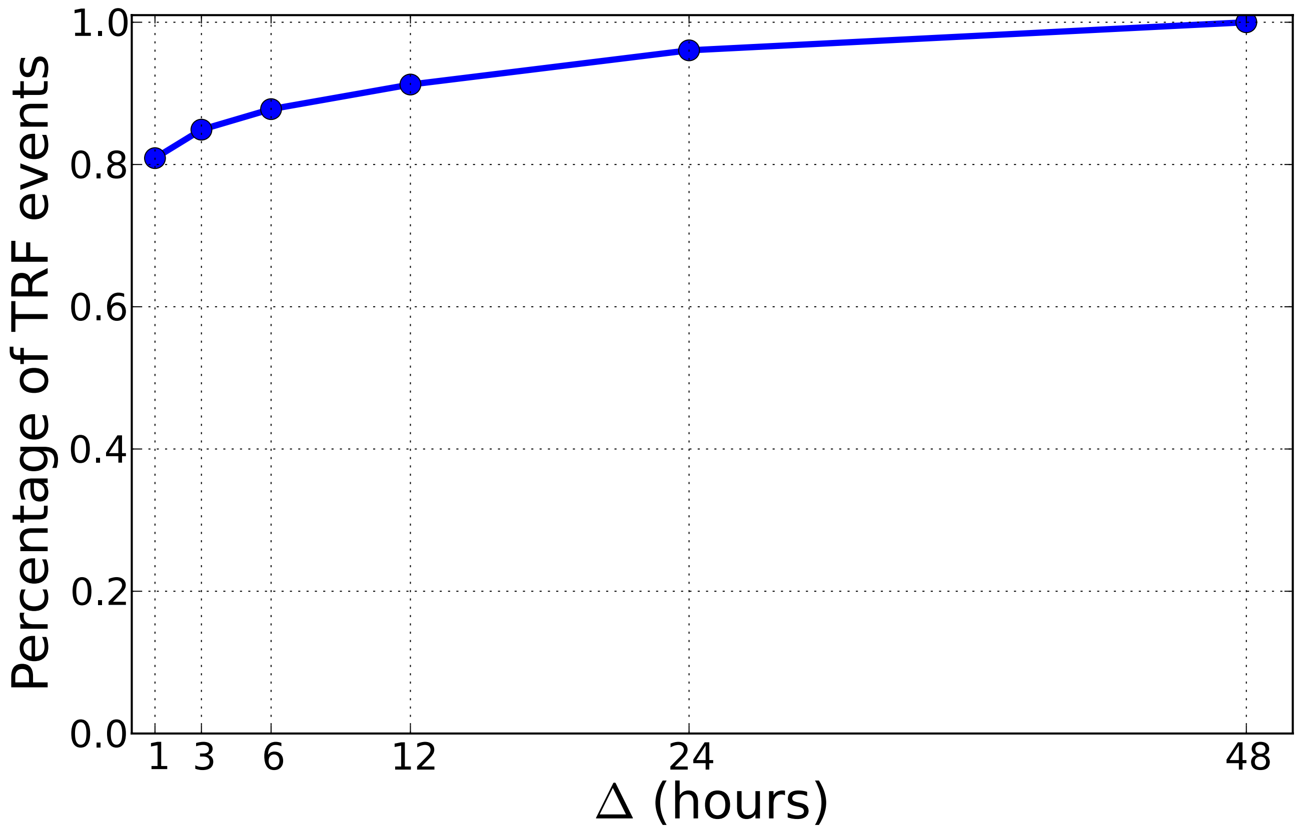
- Periodically update F(S) and F'(S)
  - Every 5 minutes
- Log a TRF event if
  - S has additional followers
  - These followers where in F(R) of at least one Repeater of S

Collected data

- September 19 to September 25 2012
  - 4746 Speakers monitored
    - Posted 386,980 tweets
  - 83860 Repeaters
    - 146,867 Retweets
    - 120 milion RT events
  - 7451 TRF events (17% of observed follow relationships)

- Bot-filtering
  - Remove all bot accounts
    - Accounts suspended by Twitter
    - 1% of collected accounts
      - Similar percentage for Speakers, Repeaters and Listeners
    - 10% of identified TRF events

TRF events Vs. Δ



Number of TRF events increases with Δ

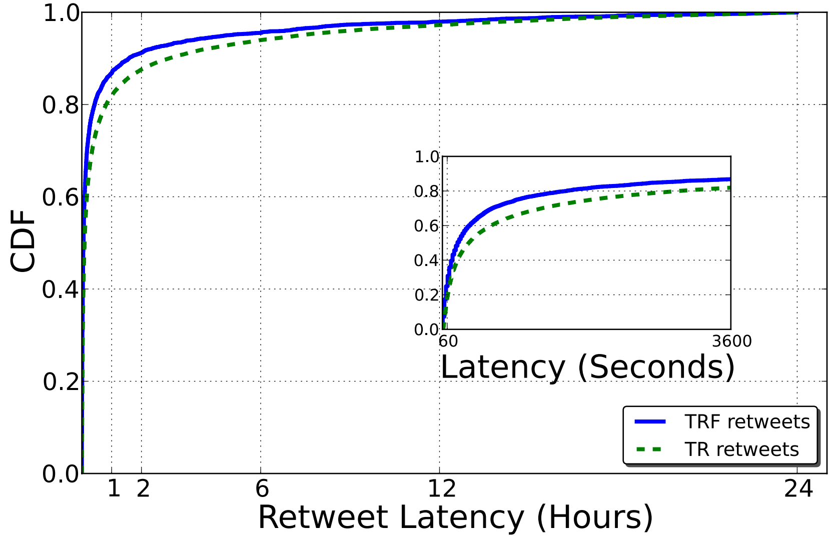
Most of the events occur within 24 hours from the retweet

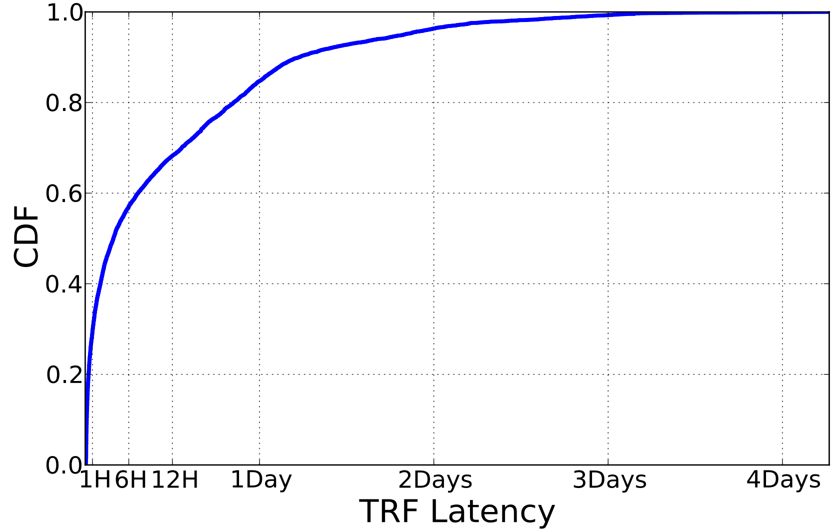
Temporal aspects of TRF events

- Retweet latency: time between the time Speaker posted the tweet and the Repeater retweeted

- TRF latency: time between the time the Repeater retweeted and the Listener followed the Speaker

Temporal aspects of TRF events (cont.)





- Most retweets in first hour from tweet
  - Users tend to act soon after information becomes available
- Users follow mostly during the same day
  - Also may follow after several days

Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network
- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - A data collection methodology for TRF events
  - TRF events Vs. exogenous new followers
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events
- Next steps:

TRF probability

- Single events: Tweet-Retweet (TR)
  - L received M from S through R at t_r
  - $TR(S, R, L, t_r, ID_t, ID_r, I_\Delta)$
  - I_Δ = 1 if L → S
  - Do not account for multiple retweets of S received by L

TRF probability (cont.)

- Retweet Groups (RG)
  - User's read their inbox periodically
  - **And** have limited attention span

$$RG(S, L, t, n, I_{\Delta})$$

  - t: time of first retweet of S in L's inbox
  - n: number of retweets of S in L's inbox during $< t, t + \Delta >$

$$P_{TRF} = \frac{RG(S, L, t, n, 1)}{RG(S, L, t, n, *)}$$

What factors affect TRF probability?

Factor	Description
	Structural Features
$\|F(S)\|$	Number of followers of S
$\|F'(S)\|$	Number of followees of S
$AGE(S)$	Number of days since S joined Twitter
$S \rightarrow L$	Reciprocity: whether the Speaker was following the Listener at the time of the TR event
	Informational Features
$\|ST(S)\|$	Total number of tweets of S
$A_{rate}(S)$	Rate of S tweets per day
$Tweets(S, L, \Delta)$	Number of distinct tweets of S received by L during period Δ
$Retweets(S, L, \Delta)$	Number of distinct retweets of S received by L during period Δ
$Repeaters(S, L, \Delta)$	Number of Repeaters R that L received tweets of S from during period Δ

Logistic Regression

Used logistic regression to examine which of the features significantly affect the TRF probability

$$ln\left(\frac{P_{TRF}}{1-P_{TRF}}\right) = \kappa_0 + \sum_{i=1}^{n} \kappa_i X_i$$

κ_i denotes the effect of feature X_i to odds of TRF events

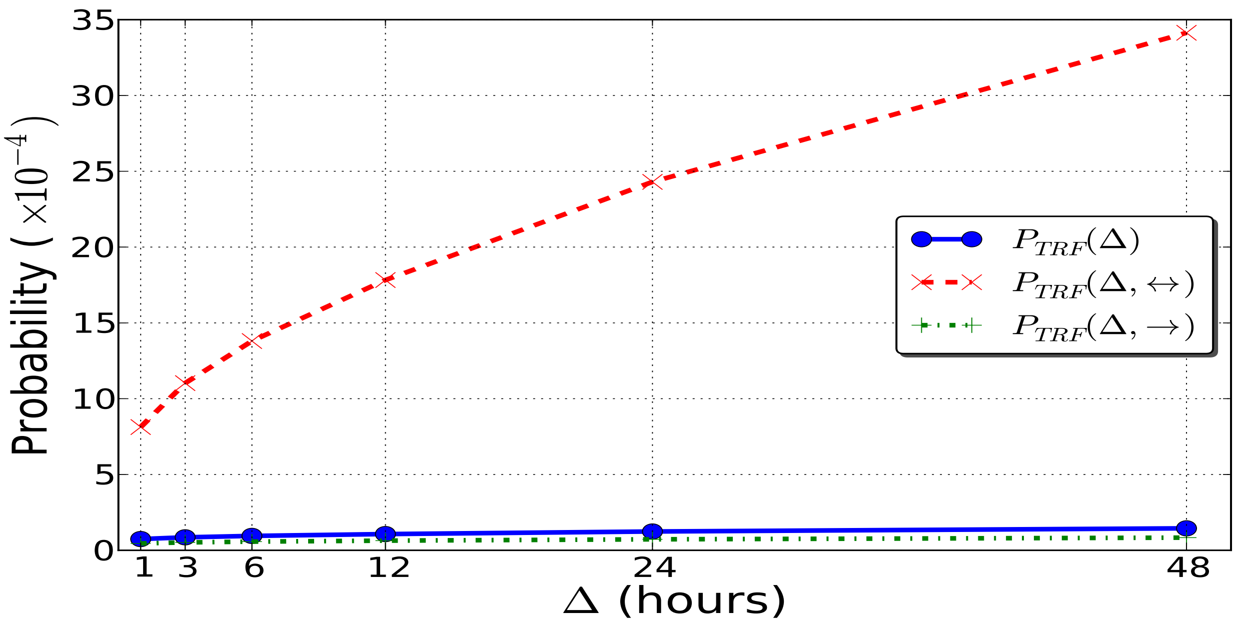
What factors affect TRF probability?

Factor	**Description**	**Odds ratio**	**95% CI**
	Structural Features		
$\|F(S)\|$	Number of followers of S	1.000^{***}	$[1.000, 1.000]$
$\|F'(S)\|$	Number of followees of S	0.999^{***}	$[0.999, 0.999]$
$AGE(S)$	Number of days since S joined Twitter	0.998^{***}	$[0.998, 0.998]$
$S \rightarrow L$	Reciprocity: whether the Speaker was following the Listener at the time of the TR event	27.344^{***}	$[25.663, 29.136]$
	Informational Features		
$\|ST(S)\|$	Total number of tweets of S	1.000^{***}	$[1.000, 1.000]$
$A_{rate}(S)$	Rate of S tweets per day	0.989^{***}	$[0.988, 0.991]$
$Tweets(S, L, \Delta)$	Number of distinct tweets of S received by L during period Δ	2.010^{***}	$[1.781, 2.270]$
$Retweets(S, L, \Delta)$	Number of distinct retweets of S received by L during period Δ	1.603^{***}	$[1.371, 1.873]$
$Repeaters(S, L, \Delta)$	Number of Repeaters R that L received tweets of S from during period Δ	2.076^{***}	$[1.889, 2.282]$

1. Reciprocity: Speaker already follows Listener (about half of TRF events)
2. Number of retweets of S received by L: how many times does S appear in L's timeline?

Reciprocity

- In 44% of the observed TRF events the Speaker was already following the Listener.



Received information

- The number of times L sees S in her inbox affects the probability for L to follow S
  - This number aggregates
    - Number of unique tweets of S seen by L
    - Number of unique retweets of S seen by L
    - Number of unique Repeaters forwarding tweets of S to L
    - All features seen to affect P_{TRF}

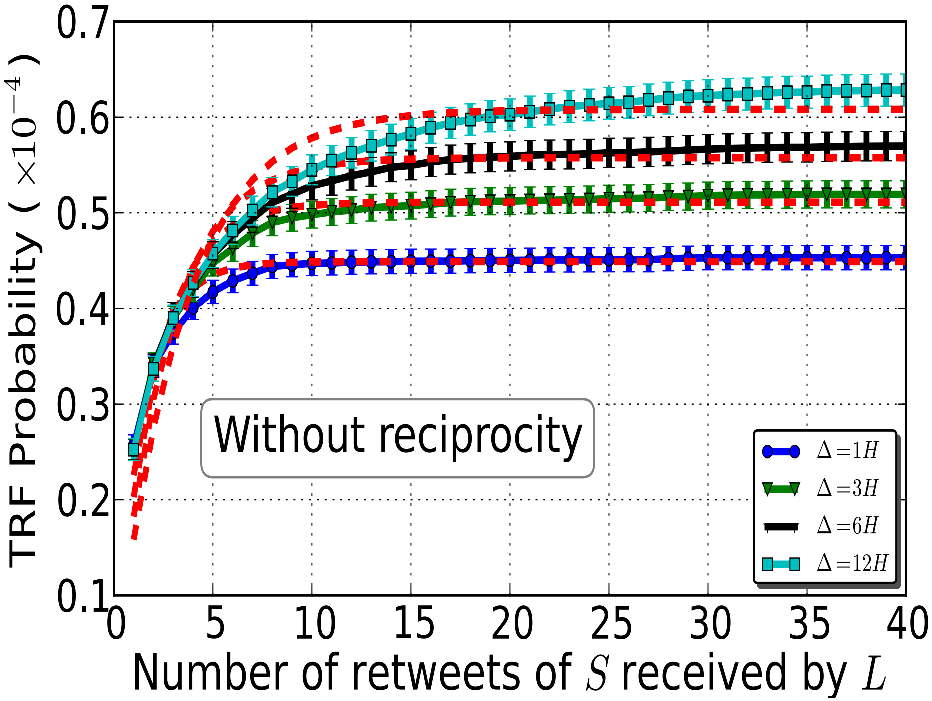
A simple model of TRF events

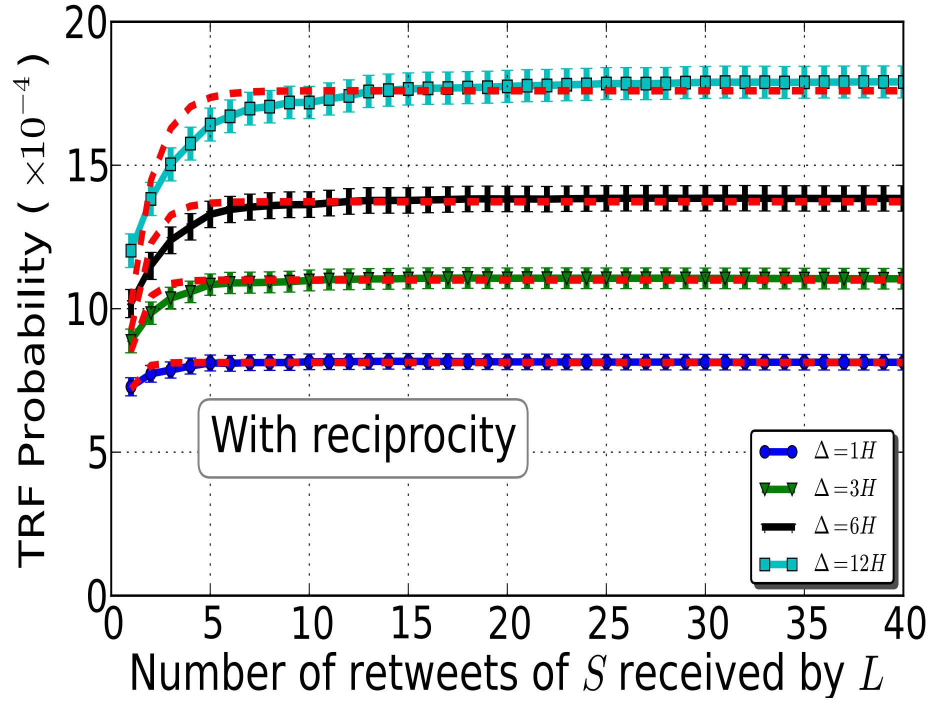
- Suppose each retweet leads to TRF event independently with probability *q*
- After receiving *n* retweets, probability of TRF =
  $1 - (1 - q)^n$
- But, Listener does not read all tweets/retweets
  - "Observation" probability *p*

$$P_{TRF}(n) = p \times (1 - (1 - q)^n)$$

- Reciprocity increases product p×q by a factor of 100
- Time window Δ affects mostly probability p
  - With reciprocity, $p \approx 25 \times 10^{-4}$ and $p \times q \approx 10^{-3}$ (Δ=24hours)

TRF model evaluation





Outline

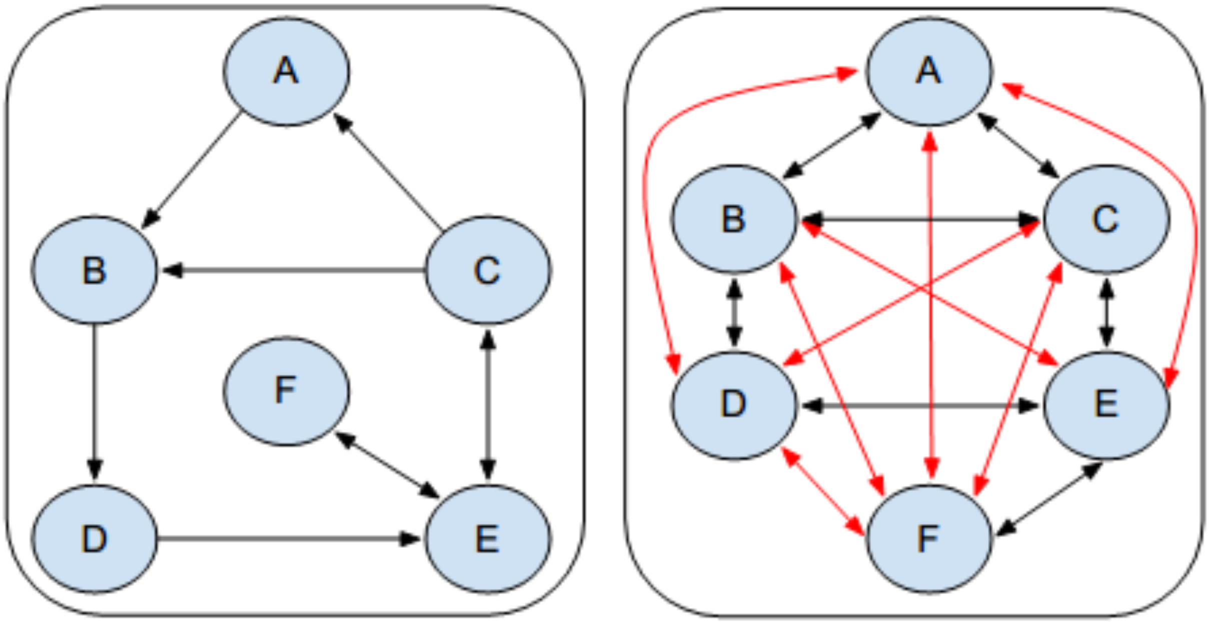
- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network
- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - A data collection methodology for TRF events
  - TRF events Vs. exogenous new followers
  - A probabilistic model for TRF events
  - What are the implications of TRF events in the structure and function of social networks?
- Unfollow events
- Next steps:

What is the effect of TRF events on the structure of the network

- Consider two scenarios:
  - A network with a pre-existing cycle
  - A hierarchical network

- How may TRF events change the network in the long-term?

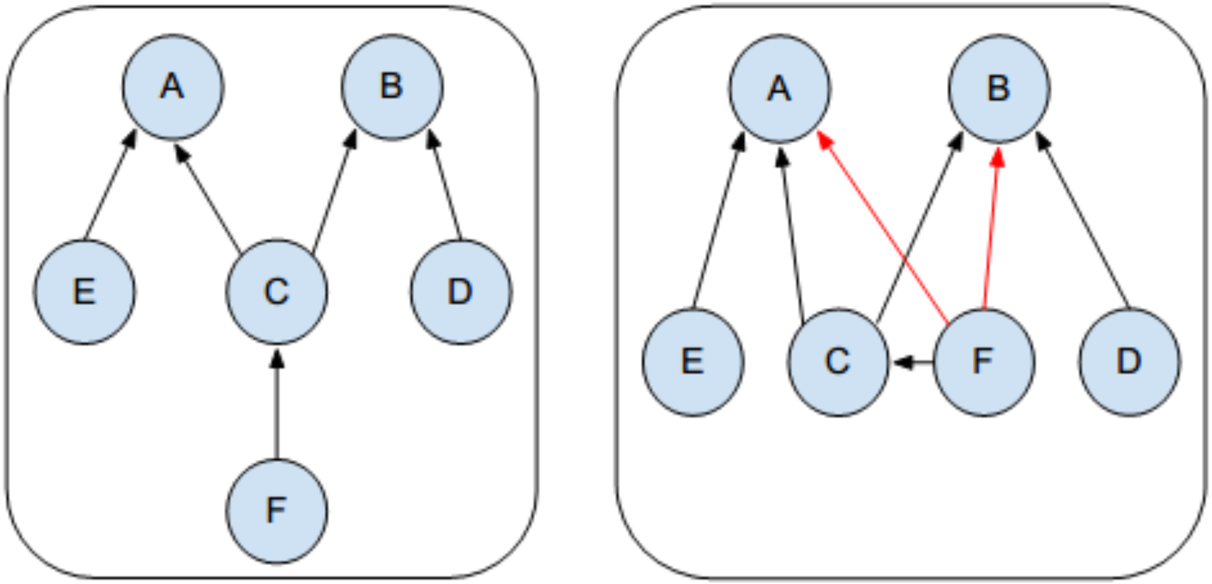
Does sub-network form a Strongly-Connected Component?



(a) Initial network topology (b) Final network topology

- It will evolve to fully connected network
- TRF events create cliques (strong communities)

Does sub-network have hierarchical structure (no directed cycles)?



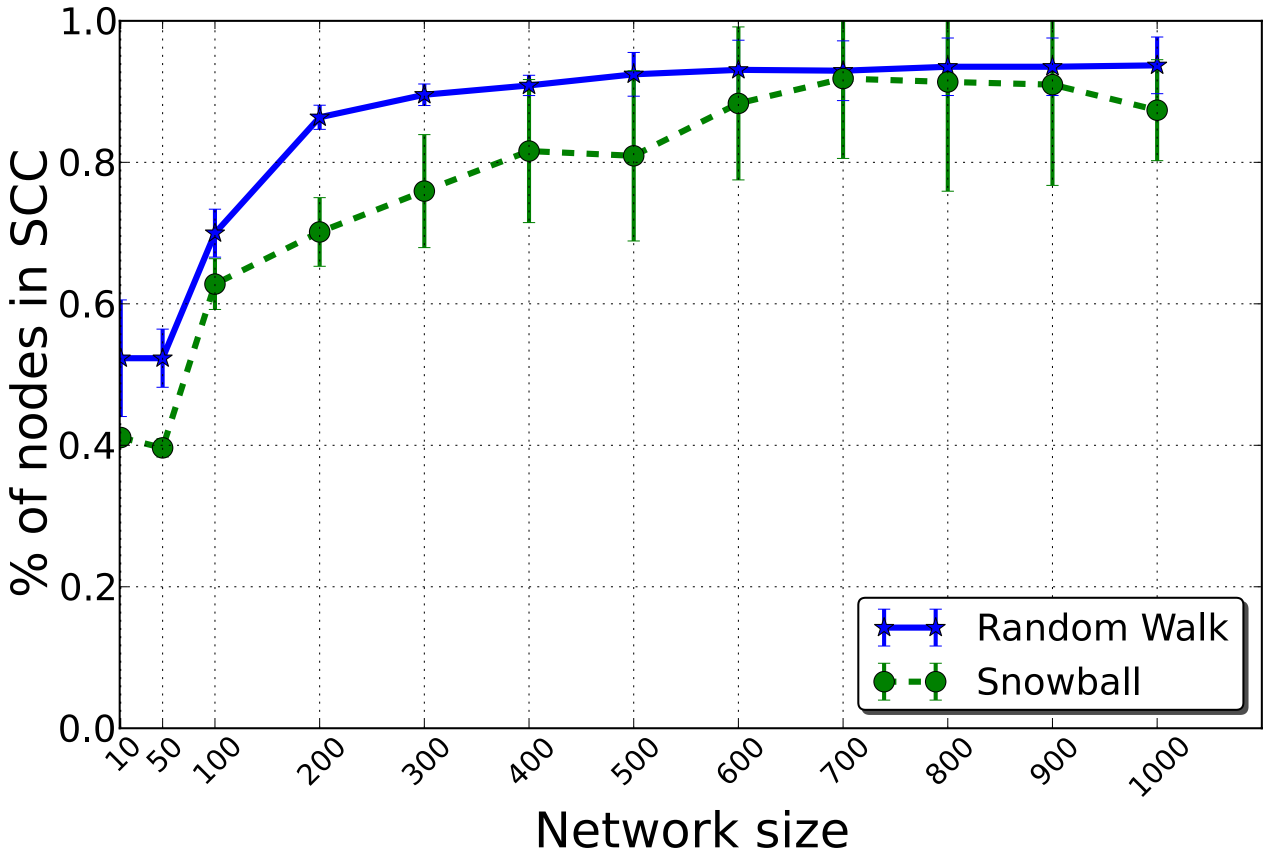
(a) Initial network topology (b) Final network topology

- Network evolves to a two-level hierarchy
- In each “sphere of influence”, an influencer is directly connected to her followers

How common are directed cycles in connected sub-graphs of theTwitter topology?

- Analyzed an older measured Twitter topology (41.7M nodes)
  - Sampling using "forest-fire" and "snowball" methods
  - Each sampled sub-network is weakly connected
  - Samples of different sizes
- Use Tarjan's algorithm to identify longest cycle (largest SCC) in sampled sub-network

For sub-graphs with more than 500 nodes, about 90% of nodes belong in SCC component



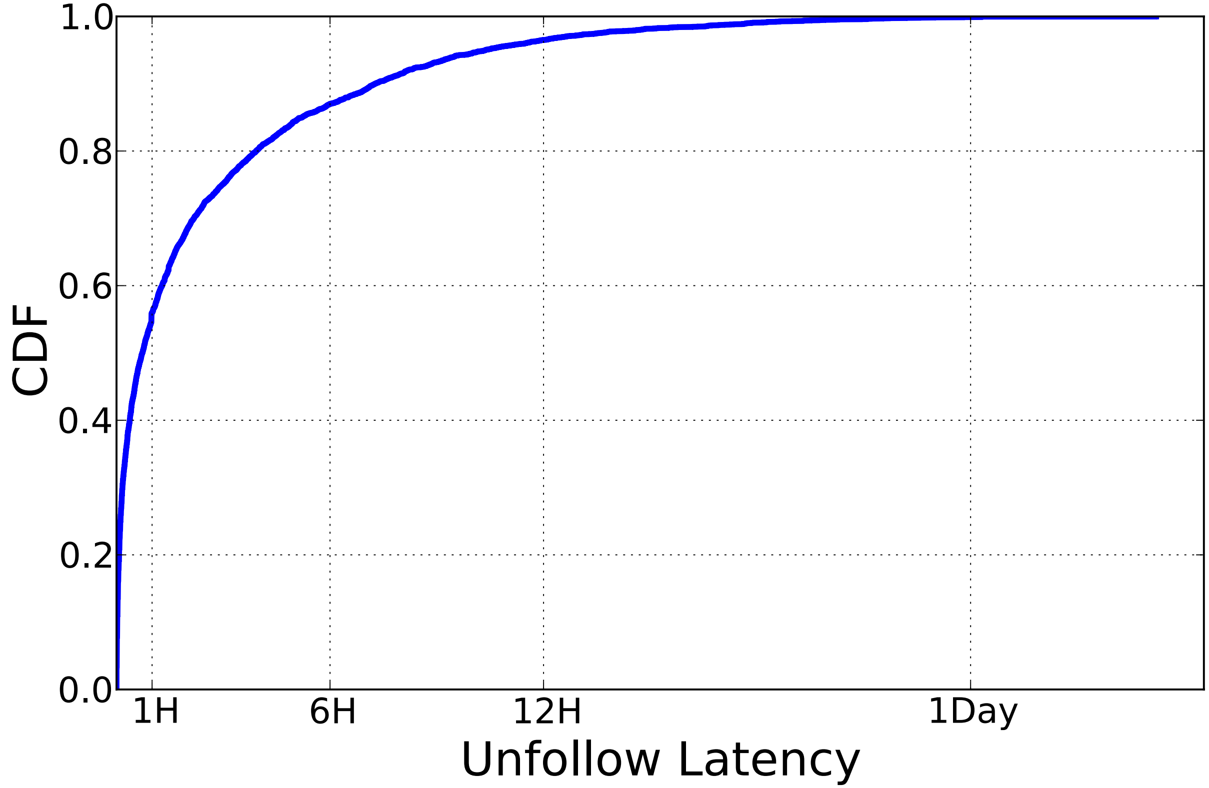
Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network
- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - A data collection methodology for TRF events
  - TRF events Vs. exogenous new followers
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events
- Next steps

Unfollow Events

- A sequence of one or more tweets of Speaker S received by a follower L may cause L to remove the link to S
- Unfollow dataset: Monitored follower lists and activity for 3,648 Speakers for 1 week
  - 4,055,327 total followers
  - 5,325 unfollow events for 983 Speakers

Unfollow Latency



- 60% of unfollow events in the first hour after S posted some content

Unfollow Probability

- How likely is for a Listener L to unfollow Speaker S during a time period Δ after receiving a tweet from S
- Activity Groups

$$AG(S, L, t_a, n, n_t, n_r, I_\Delta)$$

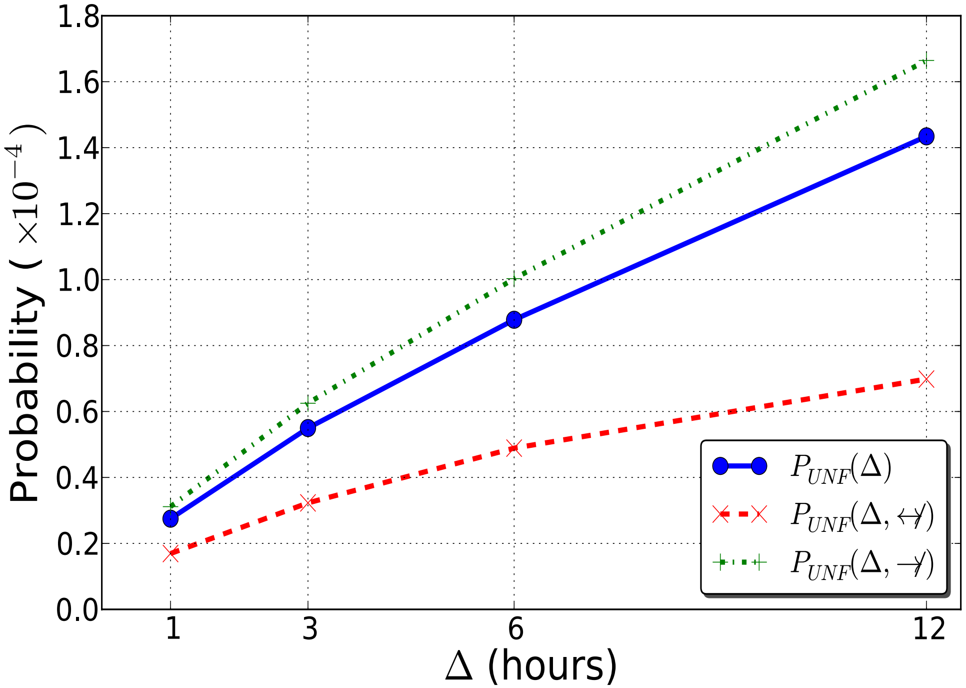
  - Similar to RG: capturing activity of S seen by L

$$P_{UNF} = \frac{AG(S, L, t, n, n_t, n_r, 1)}{AG(S, L, t, n, n_t, n_r, *)}$$

Unfollow Probability

	Odds ratio	95% CI
Structural Features		
$\lvert F(S) \rvert$	0.999^{***}	$[0.999, 0.999]$
$\lvert F'(S) \rvert$	1.000^{***}	$[1.000, 1.000]$
$AGE(S)$	0.998^{***}	$[0.998, 0.998]$
$S \rightarrow L$	0.302^{***}	$[0.261, 0.348]$
Informational Features		
$ST(S)$	1.000^{***}	$[1.000, 1.000]$
$A_{rate}(S)$	0.972^{***}	$[0.967, 0.978]$
$Tweets(S, L, \Delta)$	1.041^{***}	$[1.025, 1.057]$
$Retweets(S, L, \Delta)$	1.026	$[0.992, 1.006]$

*p<0.1; **p<0.05; ***p<0.01



- 60% less likely to unfollow when relationship is reciprocal

Outline

- Dynamics on/off networks
- Joint dynamics on and of networks:
  - Co-evolutionary network
- Case-study: co-evolutionary nature of Twitter
  - Tweet-Retweet-Follow (TRF) events
  - A data collection methodology for TRF events
  - TRF events Vs. exogenous new followers
  - A probabilistic model for TRF events
  - What are the long term implications of TRF events in the structure and function of social networks?
- Unfollow events
- Next steps

Conclusions

- Examined co-evolutionary dynamics on Twitter
  - Tweet-Retweet-Follow events
- TRF events are responsible for 20% of the new edges on Twitter
  - 80% occur in 1 day after the retweet
- Proposed a probabilistic model for TRF events
- TRF events tend to transform cycles topologies to cliques
  - 80-90% of the nodes in weakly-connected groups sampled from Twitter showed to belong to a directed cycle
- Unfollow events are also co-evolutionary
  - 60% in the first hour after Speaker's activity

Future work

- Simulate a Twitter like network considering TRF and Unfollow events
  - How does the network change?
  - How is information diffusion affected?

- Further examination of co-evolutionary dynamics on social networks
  - What other types of such dynamics are present?

Thank you!