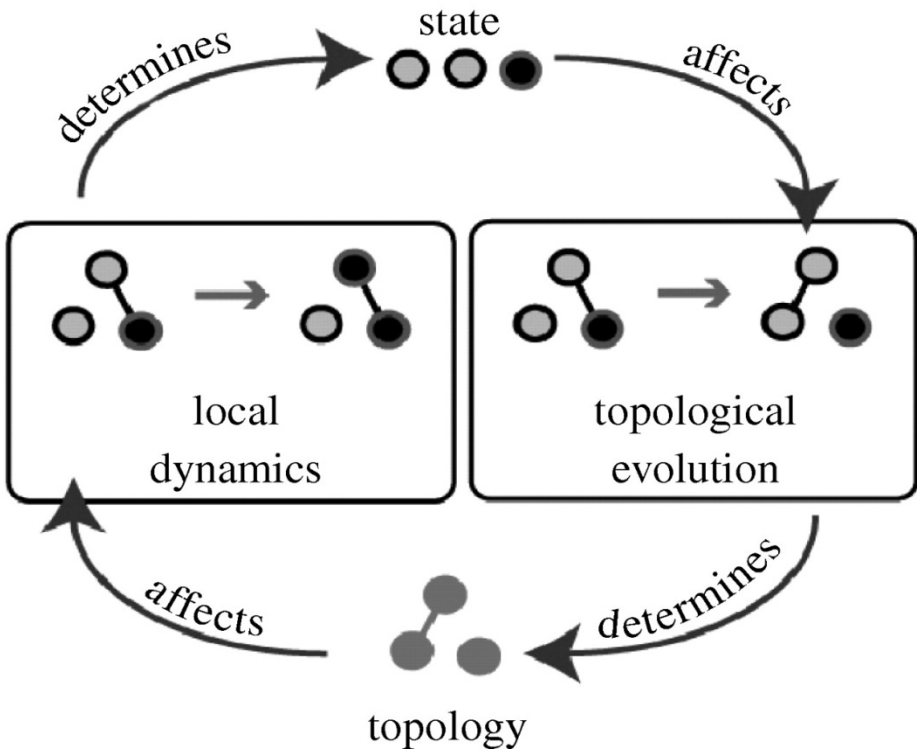
Co-evolutionary dynamics in social networks: A case study of Twitter

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

Constantine Dovrolis
constantine@gatech.edu
College of Computing
Georgia Institute of Technology

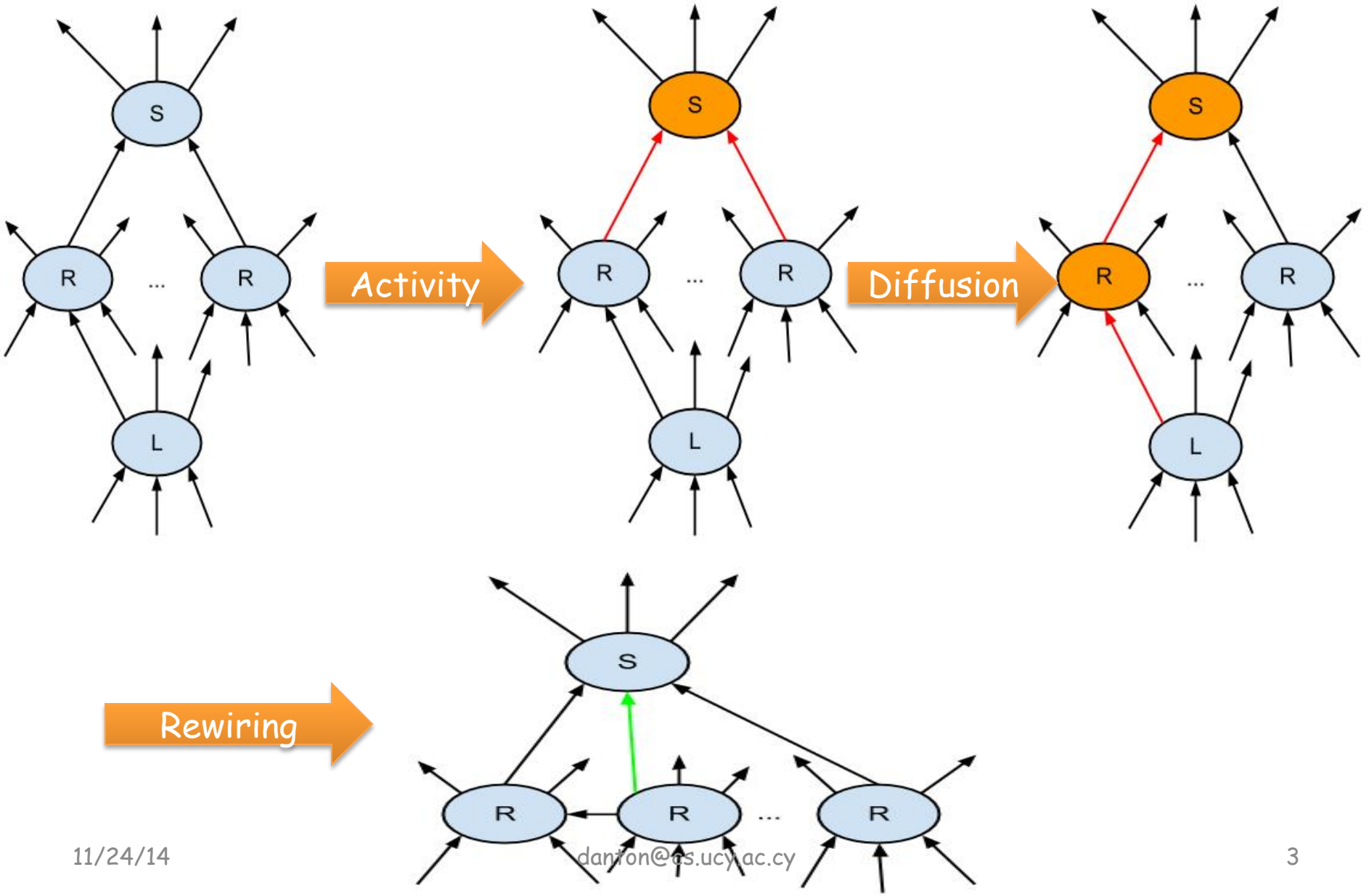
Co-evolutionary dynamics 101

- Coupled dynamics *ON and OF* networks



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

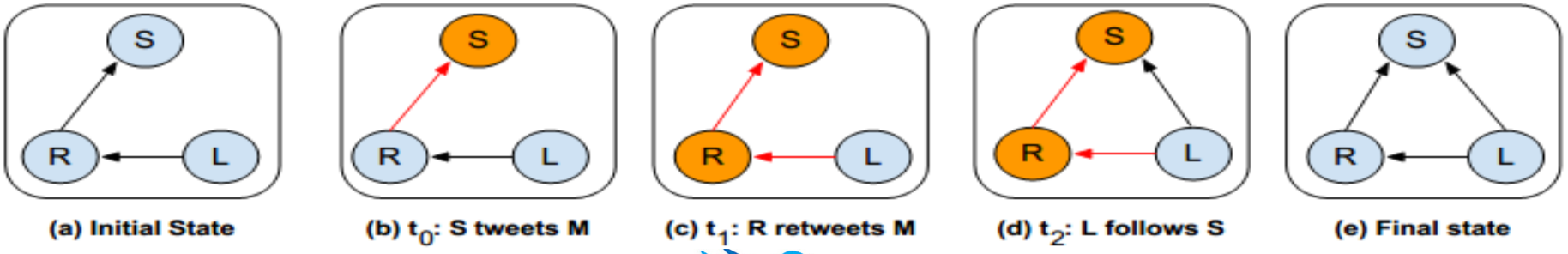
Are social nets co-evolutionary?



Main objectives of this work

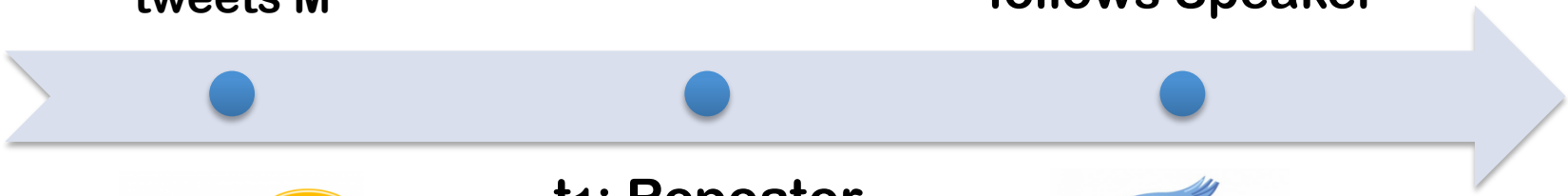
- In the context of Twitter:
  - Examine presence of co-evolutionary effects
  - Quantify their likelihood and analyze them statistically
  - Develop a simple probabilistic model, based on empirical results
  - Examine how co-evo dynamics may affect a social network in long-term

Tweet-Retweet-Follow (TRF) events



t0: Speaker (S) tweets M

t2: Listener (L) follows Speaker



t1: Repeater (R) retweets M

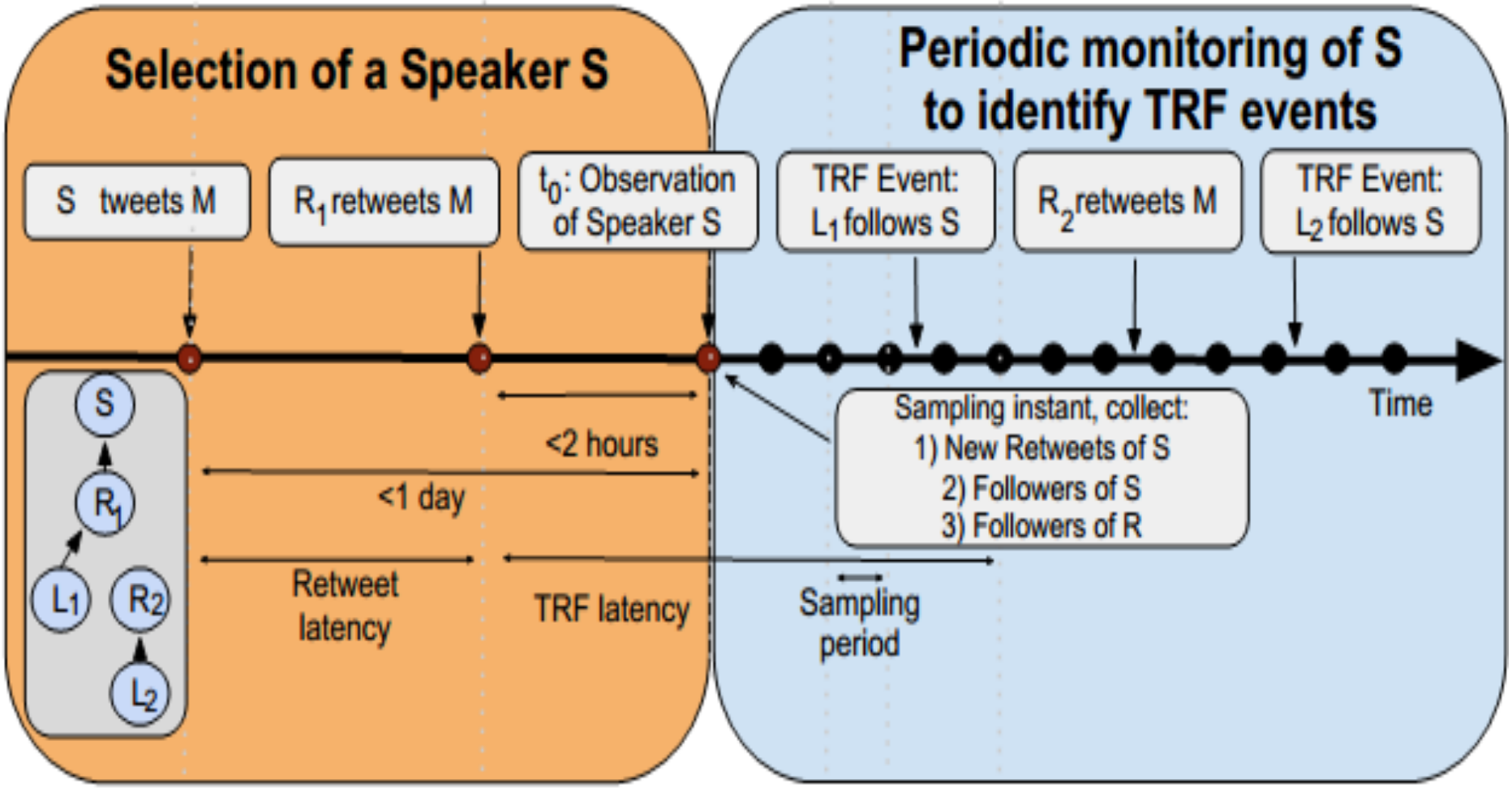
- TRF events: clear case of co-evolutionary dynamics

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

Data collection methodology

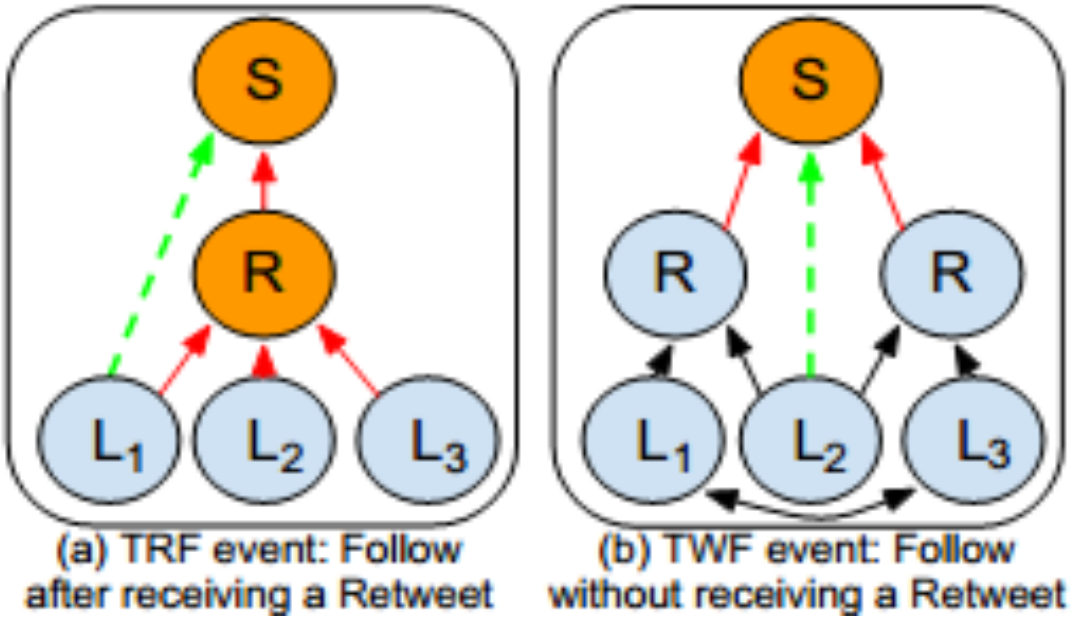


Collected data

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

- Bot-filtering
  - Remove bot accounts (accounts suspended by Twitter)
    - 1% of collected accounts
    - 10% of identified TRF events

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

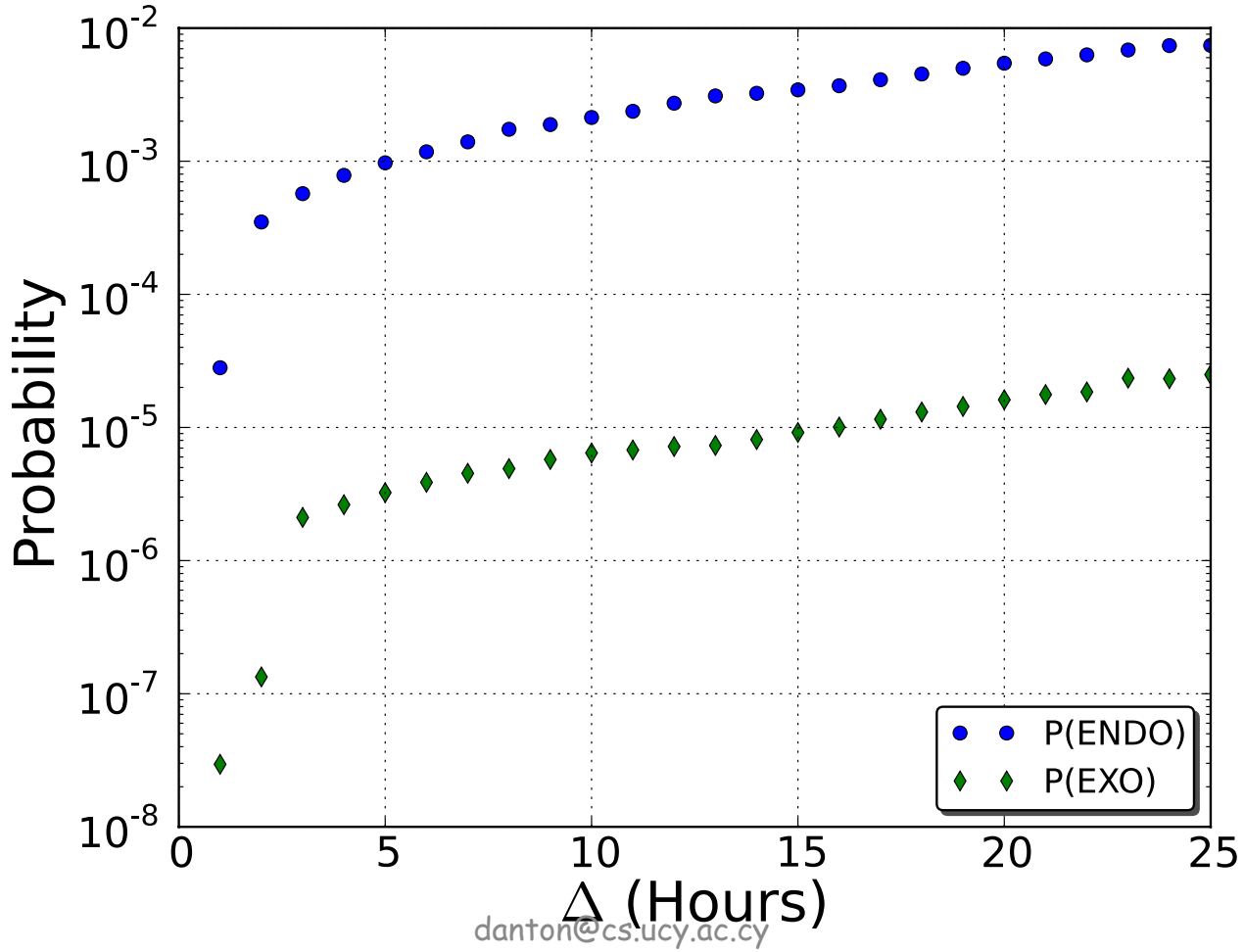


(a) TRF event: Follow after receiving a Retweet

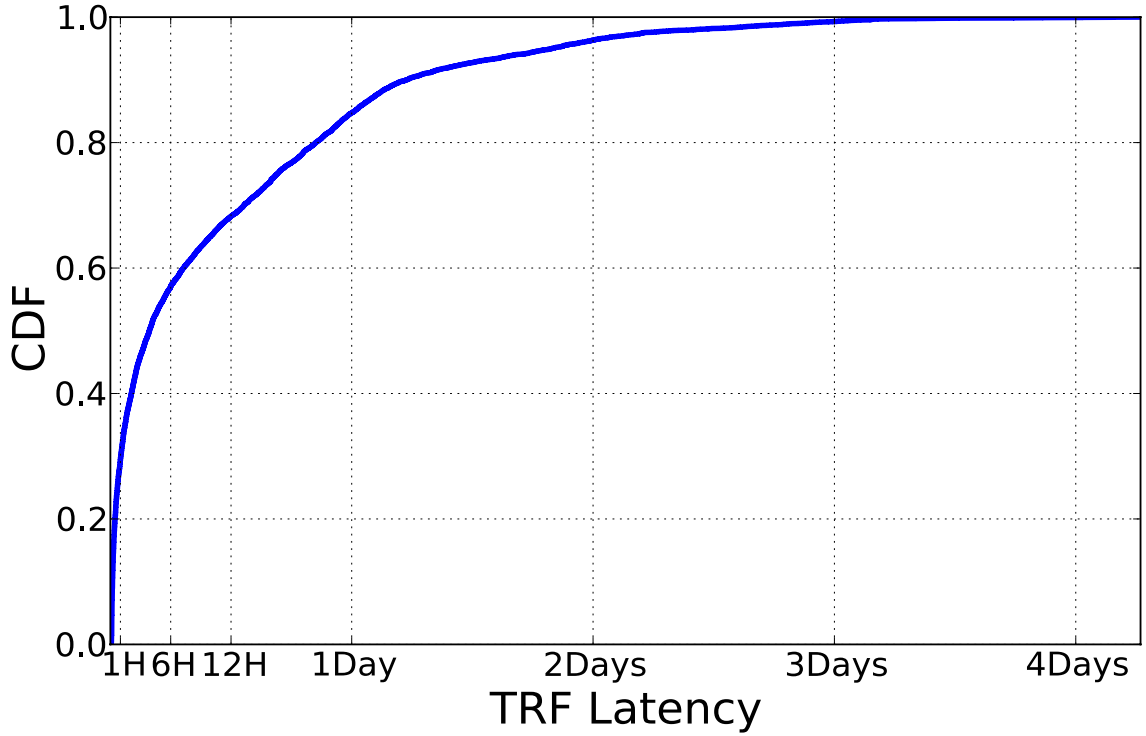
(b) TWF event: Follow without receiving a Retweet

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

Effect of receiving (or not receiving) a retweet



TRF Latency



- A new Listener follows Speaker typically within first 24-48 hours from last received retweet

Which 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

Examine which of the previous features affect TRF probability significantly:

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

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

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?

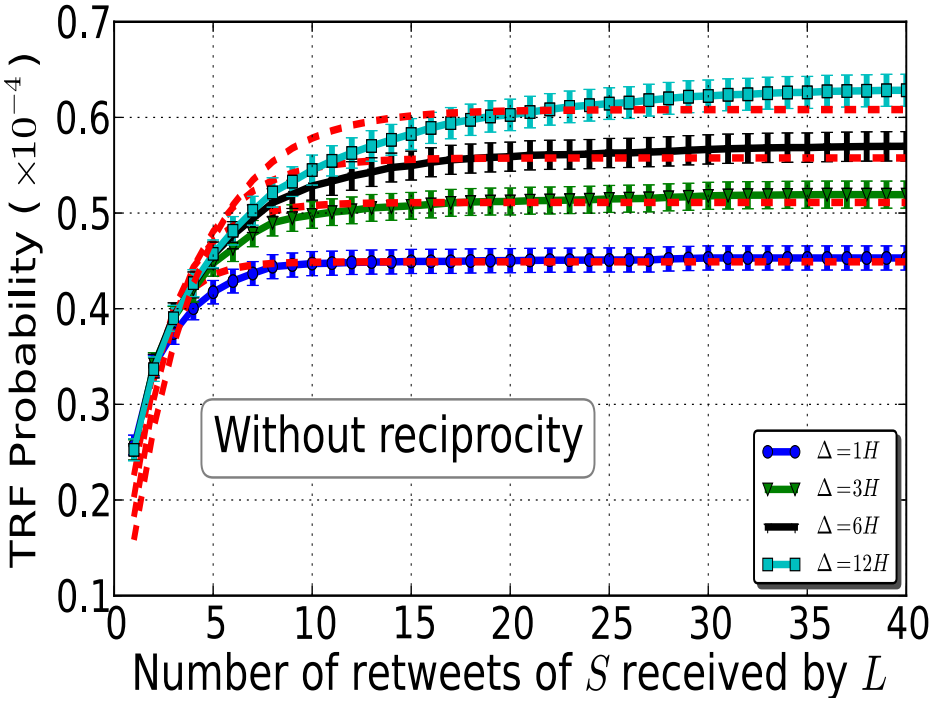
A simple model of TRF events

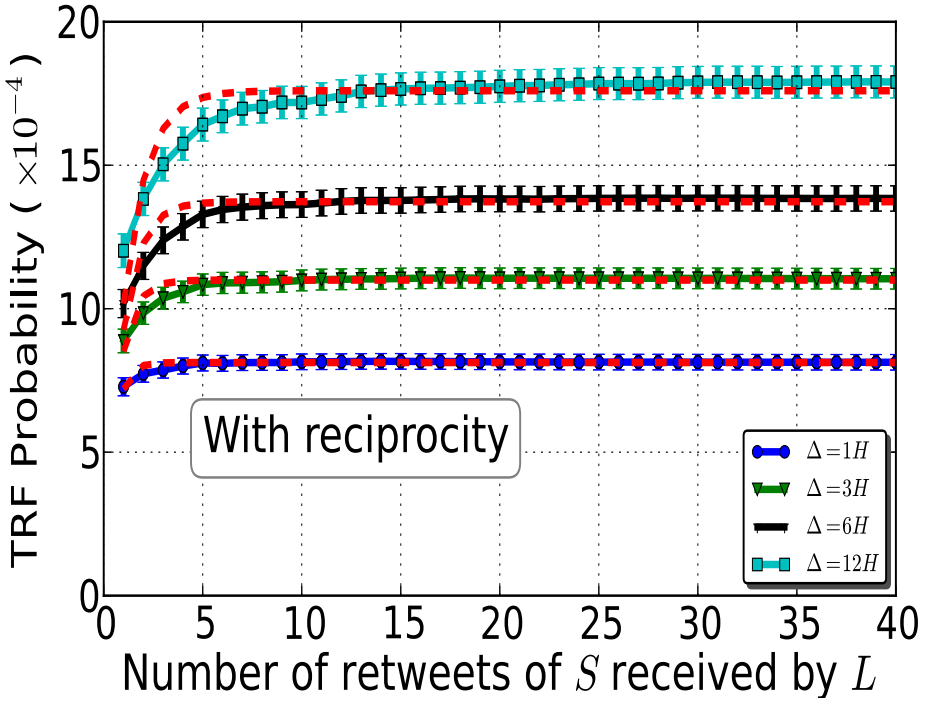
- Suppose each retweet leads to TRF event independently with probability *q*
- After receiving *n* retweets, probability of TRF = $1 - (1 - q)^n$
- But, Listener does not read all tweets/retweets
  - "Observation" probability *p*

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

- 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)

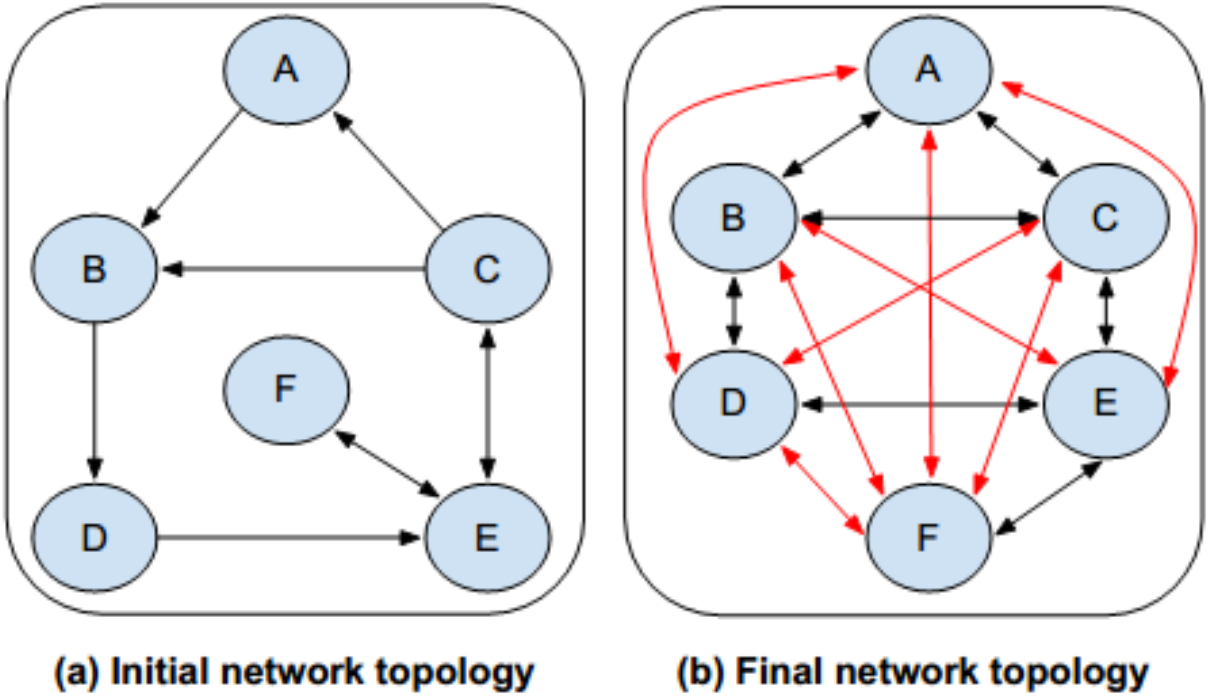
TRF model evaluation





What is the effect of TRF events in the long-term evolution of a social (sub-)network?

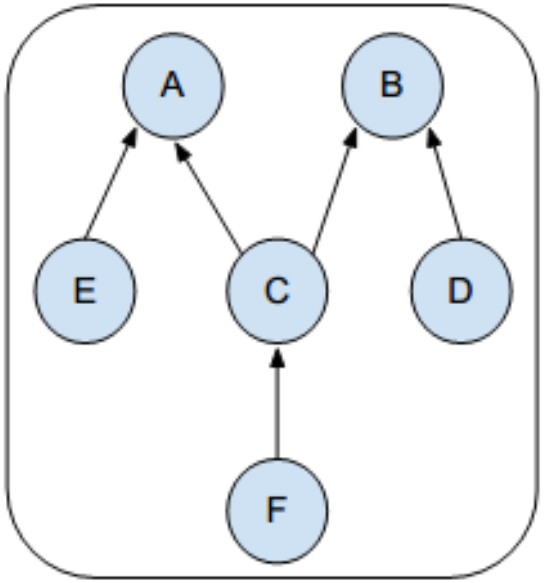
Does sub-network form a Strongly-Connected Component?



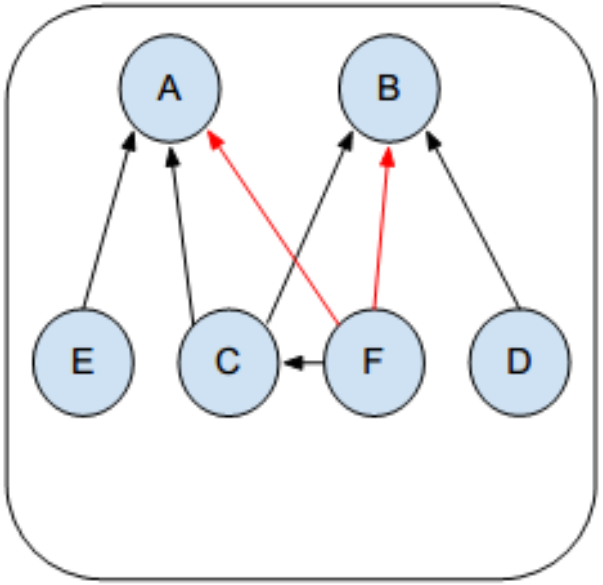
(a) Initial network topology (b) Final network topology

- 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



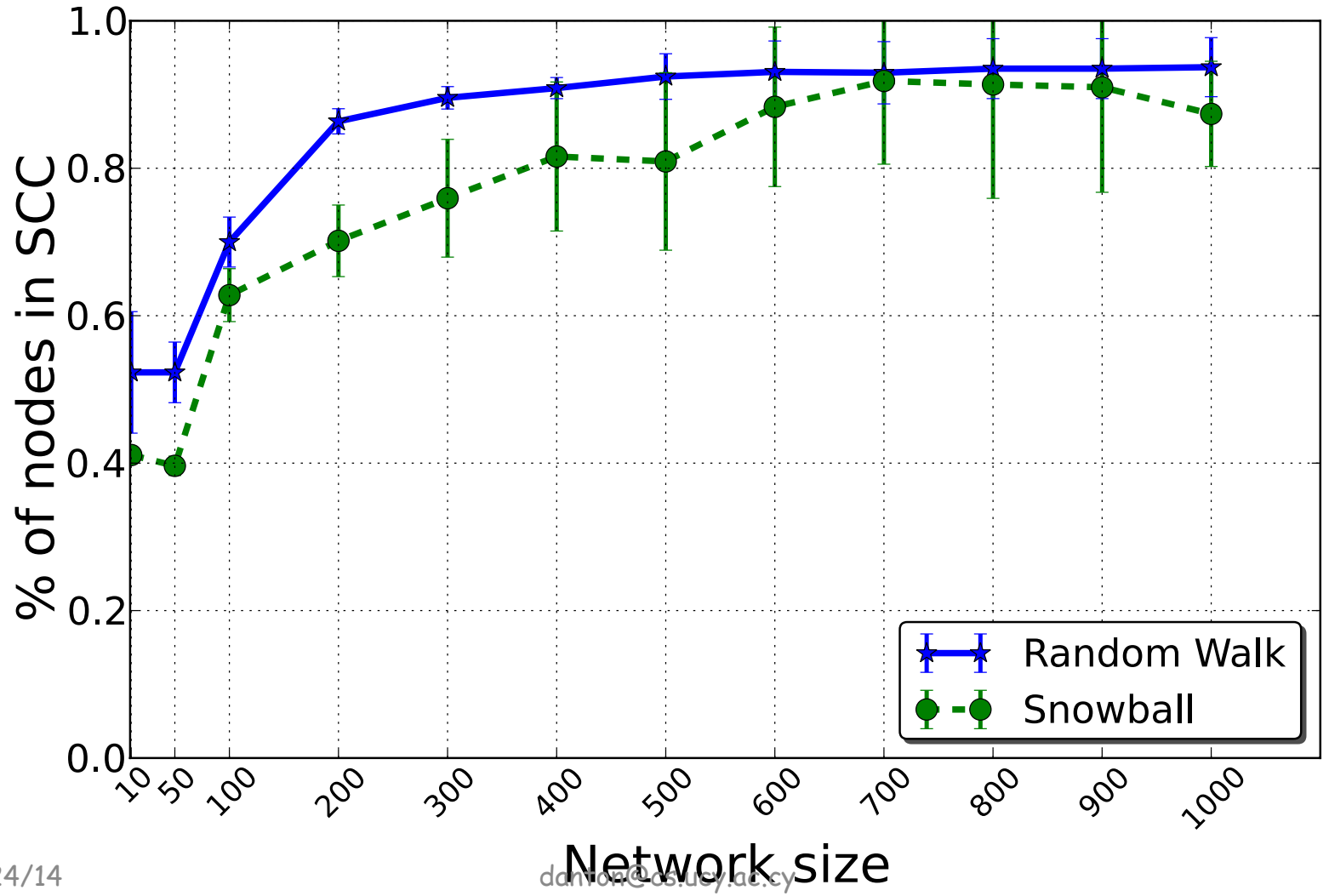
(b) Final network topology

- 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



Conclusions

- Observed co-evolutionary dynamics on Twitter
  - Tweet-Retweet-Follow events
- TRF events are responsible for 20% of the new edges in Twitter's network
  - 80% occur within 1 day after the retweet
- Proposed a probabilistic model for TRF events (simple enough for analytical studies?)
- TRF events tend to transform Twitter sub-nets to cliques

Q&A