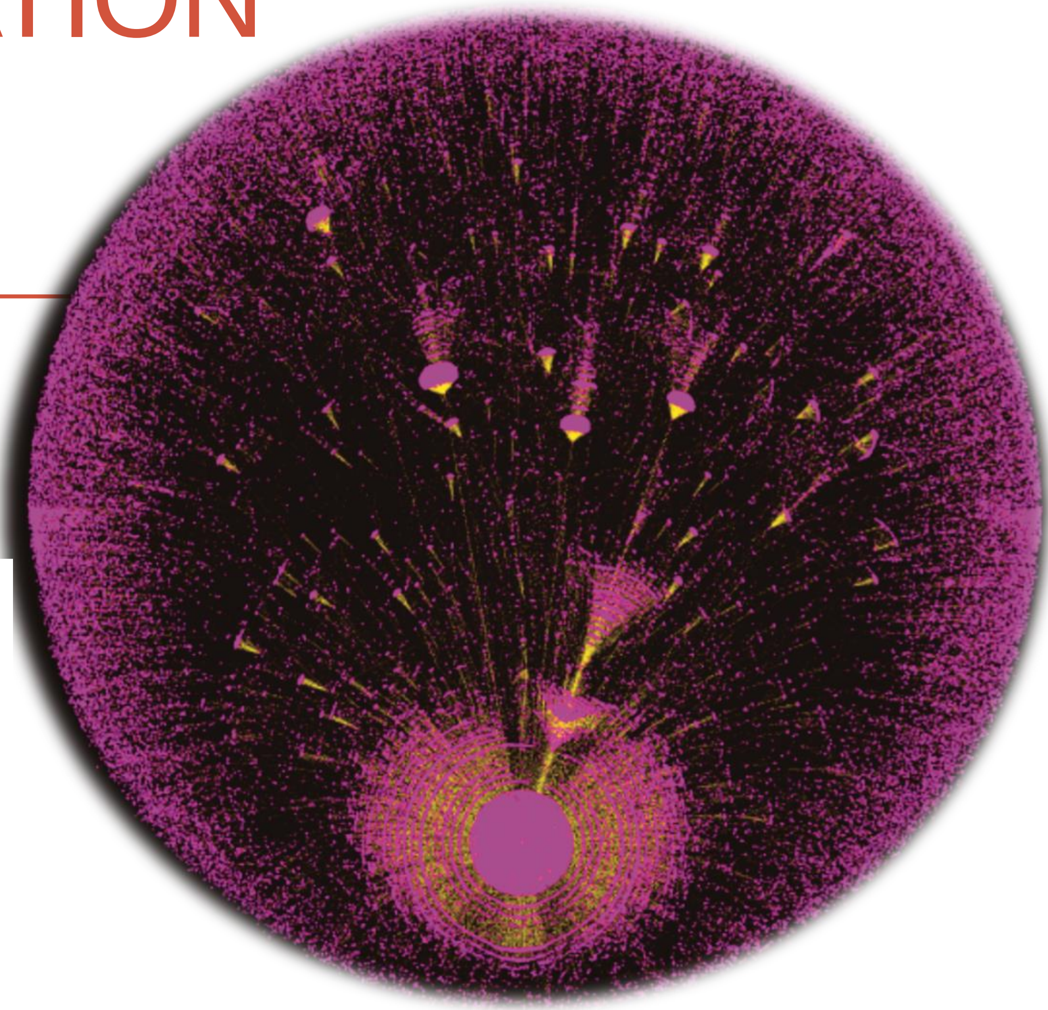
VIRAL MARKETING OF INFORMATION

Feng Ling
Complex Systems Group,
Institute of High Performance Computing
A*STAR, Singapore

Institute of High Performance Computing
A*STAR



Online Social Media (OSN)

Devastating Power of OSN

- Social Unrest

- 2011 London Chaos
  - 3443 crimes within days
  - 3100 people arrested
  - $300 million damage
  - Blackberry Messenger???

- Arab Spring
  - Organized through Social media
  - Facebook
  - Twitter

Promoting power of OSN
- Viral Video

Influence of nodes s_i

- Influence of node i
  - Expected **final** number of infected nodes s

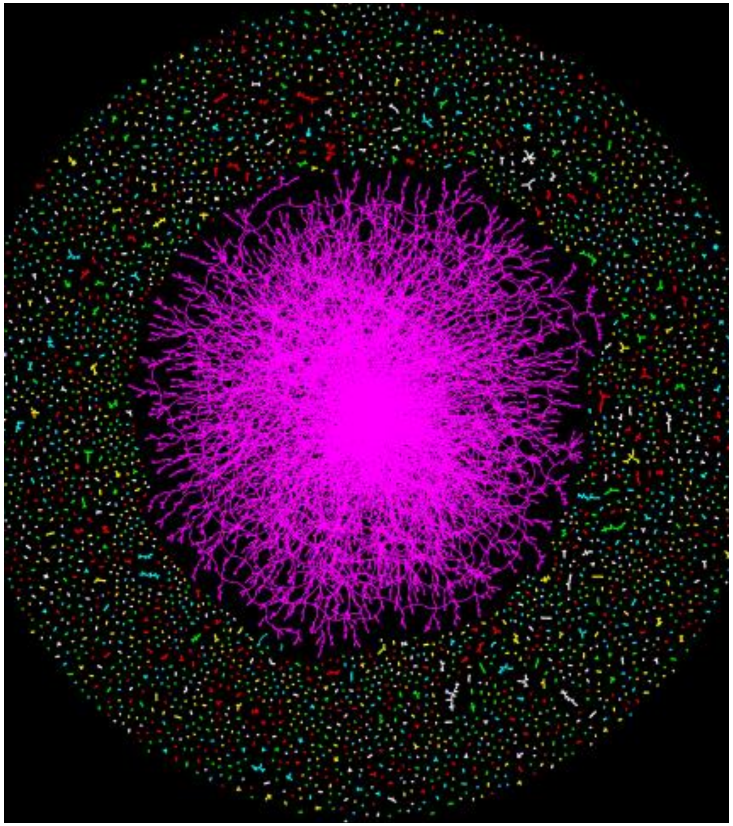
$$s_i = \sum_{s=1}^{N} s \cdot p_i(s)$$

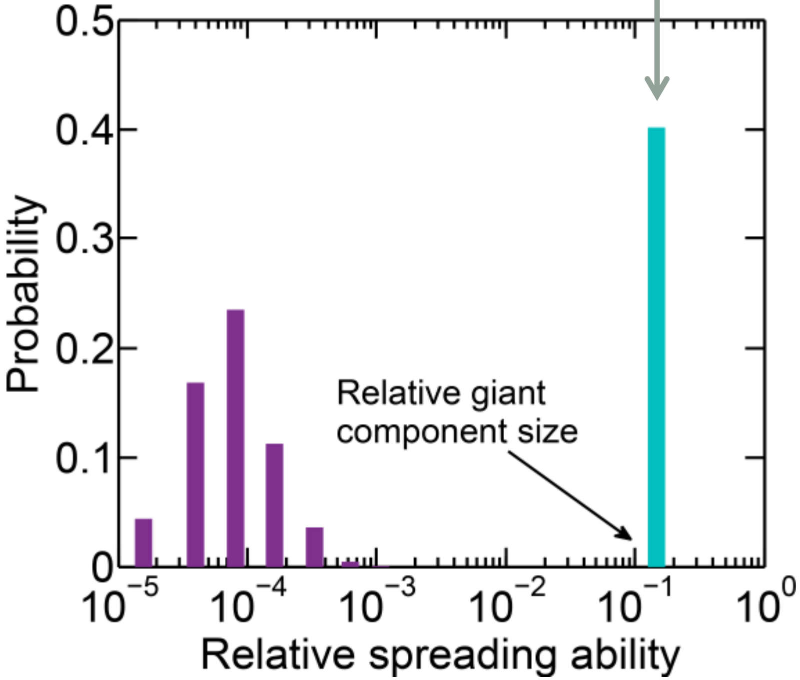
- Influence of set of nodes

$$s_V = \sum_{s=1}^{N} s \cdot p_V(s)$$

Bimodal spreading outcome

Unique
Giant Cluster size





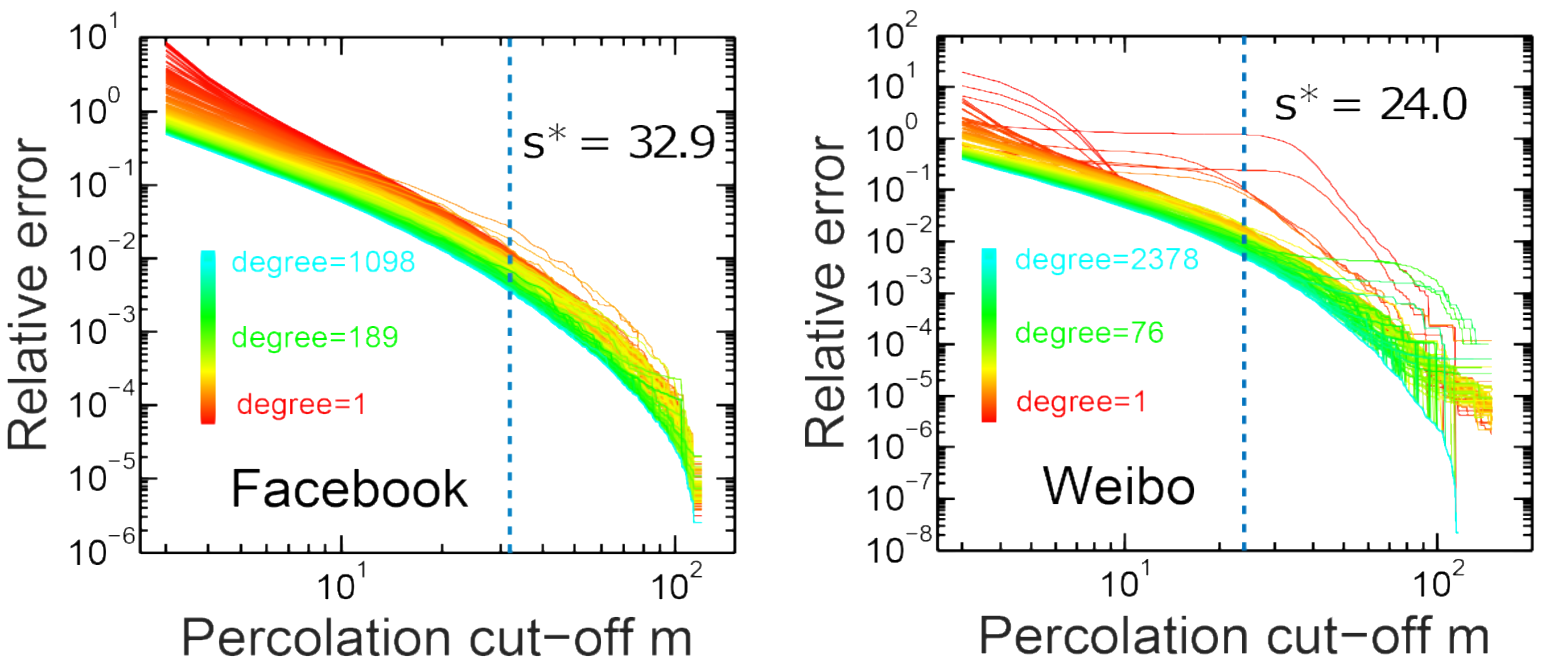
* SIR simulation on New Orleans Facebook network, http://socialnetworks.mpi-sws.org/data-wosn2009.html. (Bimal Viswanath and Alan Mislove and Meeyoung Cha and Krishna P. Gummadi, On the Evolution of User Interaction in Facebook', Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09)).

* Hu Yanqing, Shenggong Ji, Ling Feng, Yuliang Jin, Shlomo Havlin, ‘Quantify and Maximise Global Viral Influence Through Local Network Information ‘arXiv preprint arXiv:1509.03484, 2015 - arxiv.org

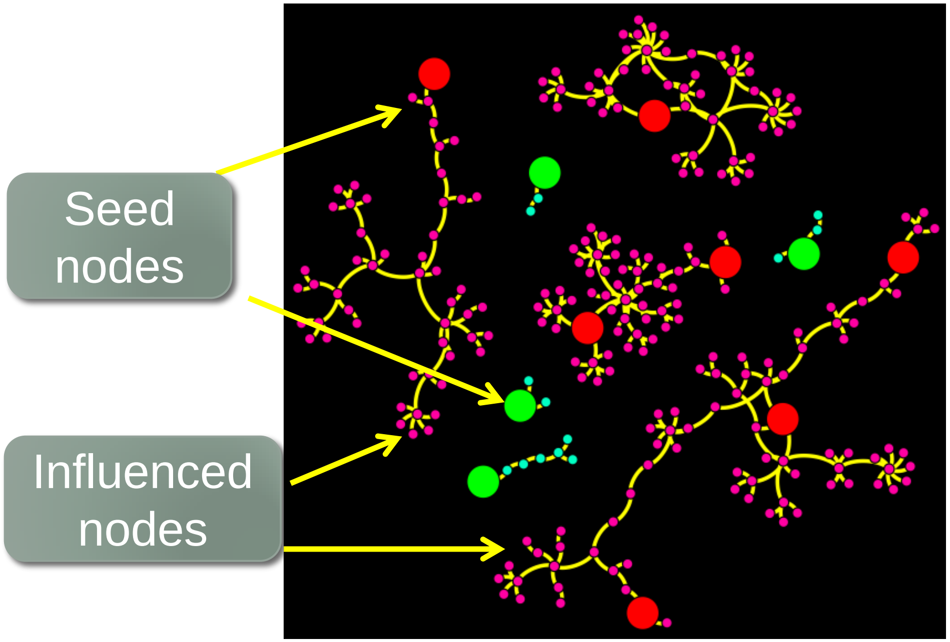
Determination of cutoff m = 2 s*

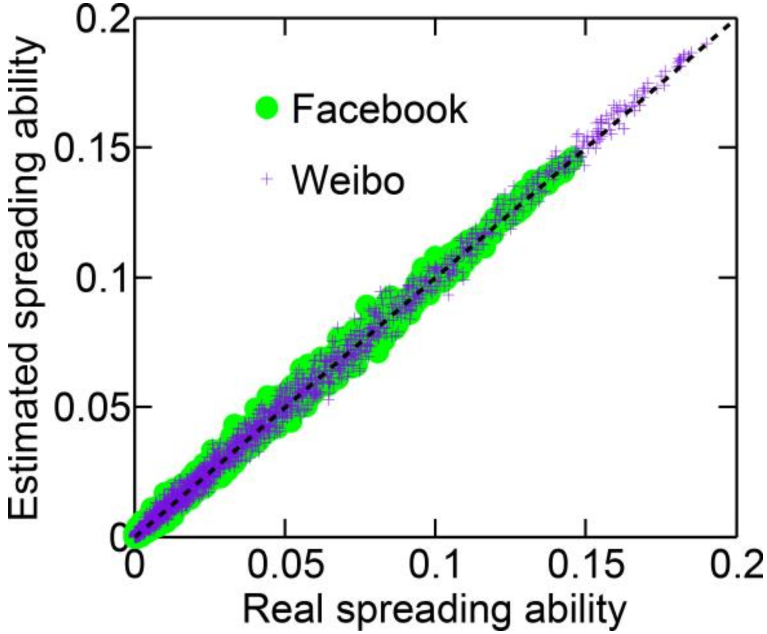
Exponential decay of cluster size with characteristic size s* （excluding GC)

$$p_{\mathrm{f}}(s) \sim s^{-\tau} e^{-s/s^*}$$



Simple Spreadability estimation - cluster size cutoff m = 50



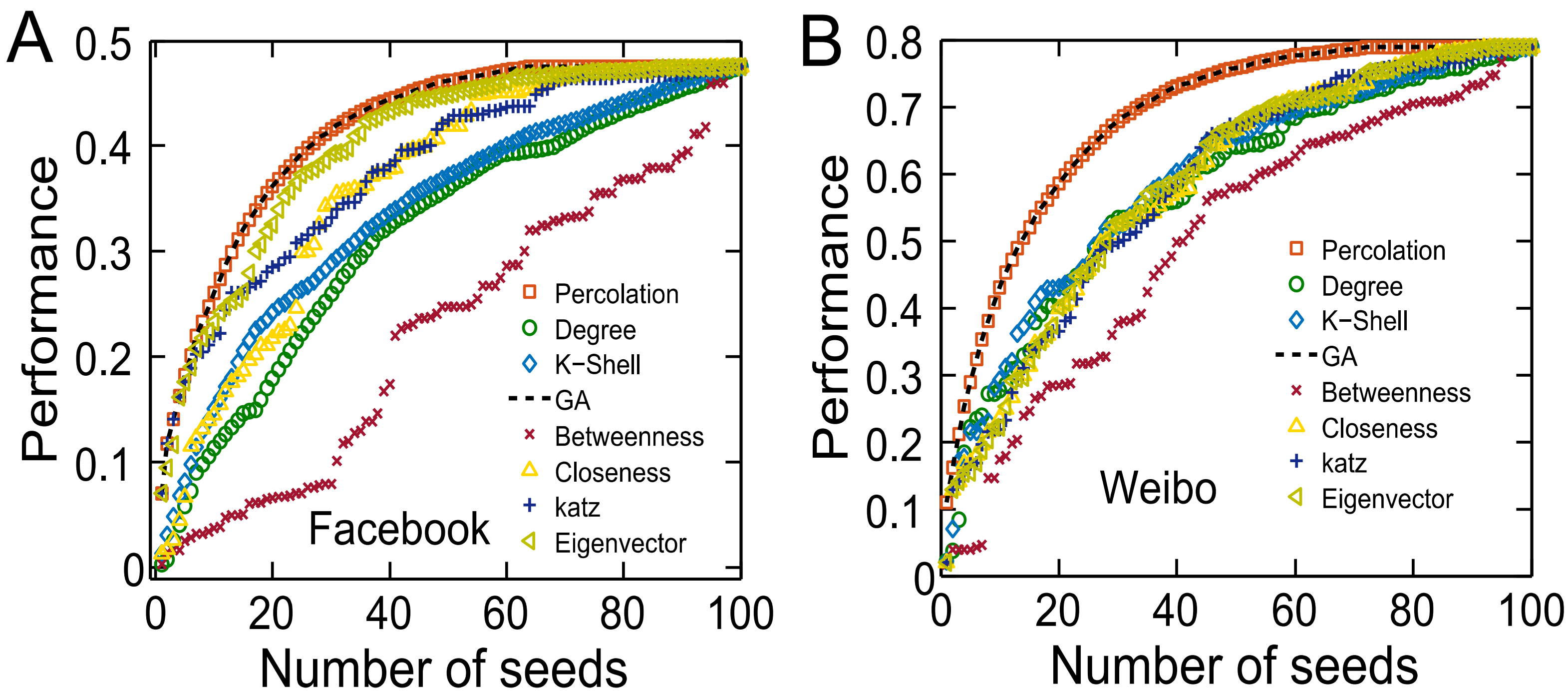


Spreading ability $s_V = s^{\infty} \cdot p_V(s^{\infty})$

GC Size $\left[1 - \sum_{s_i=1}^{m-1} p_i(s_i)\right]$

GC Probability $\left[1 - \left\langle \sum_{s_j=1}^{m-1} p_j(s_j) \right\rangle\right]$

Find the best spreader set (most influential) - Natural greedy of percolation



Algorithm	Percolation	Greedy	Genetic	K-shell	Katz	Eigen	Closeness	Between
Complexity	O(LM)	O(NLM)	O(NMPG)	O(N)	O(N^3)	O(N)	O(N^2logN)	O(N^2)

Independent of N – network size

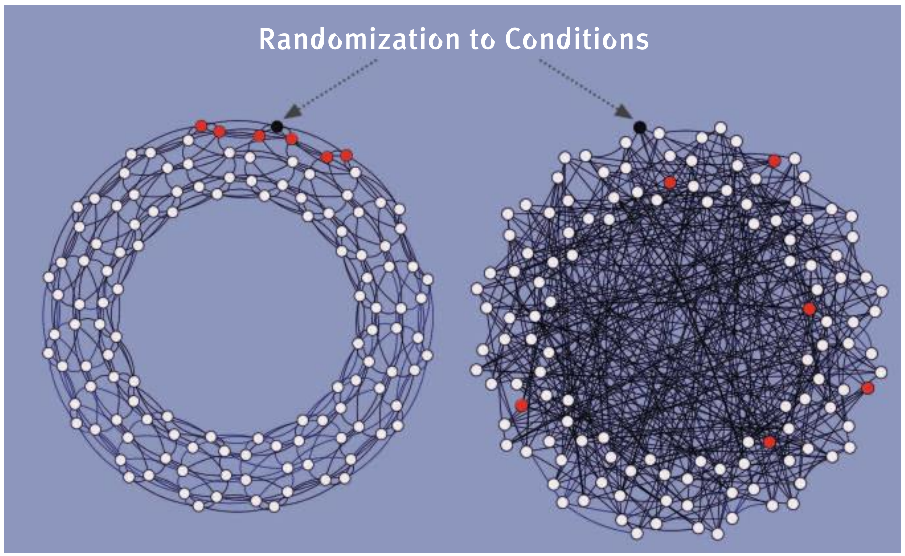
Conclusion

- Local structure hints on global influence
- Low computational complexity – independent of N
- Rich physics
  - Percolation – Spreading equivalence
  - Existence of Giant Cluster
  - Unique size of Giant Cluster
  - Exponential decay of non giant cluster sizes
  - Estimated characteristic cluster size – deciding factor of cutoff m
  - Larger N, more defined GC size, better estimation accuracy.

Hu Yanqing, Shenggong Ji, Ling Feng, Yuliang Jin, Shlomo Havlin, ‘Quantify and Maximise Global Viral Influence Through Local Network Information ‘arXiv preprint arXiv:1509.03484, 2015 - arxiv.org

Is SIR model realistic??

Experiment on OSN
- Social Reinforcement



Clustered

Links shuffled

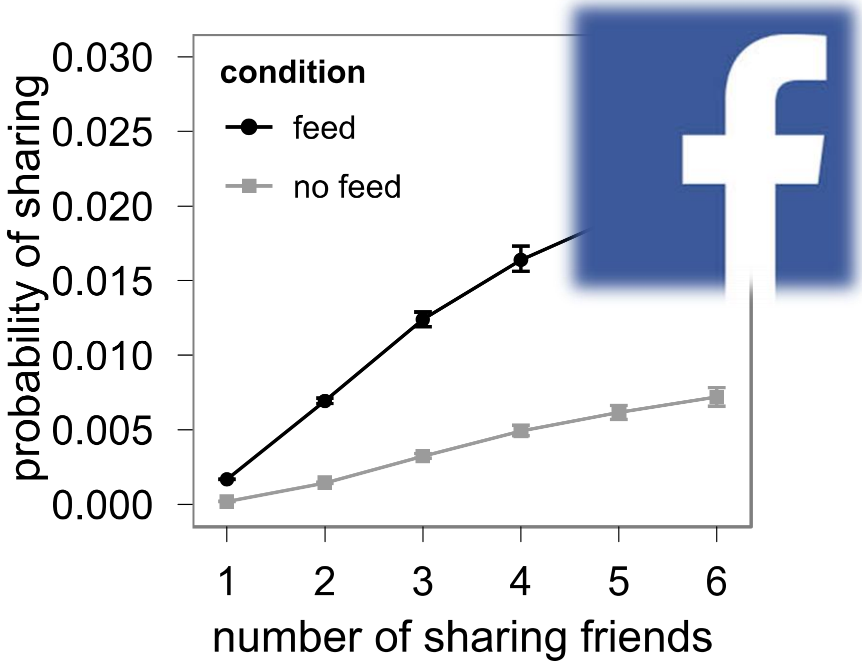
Clustered network diffusion opinions/behaviors faster

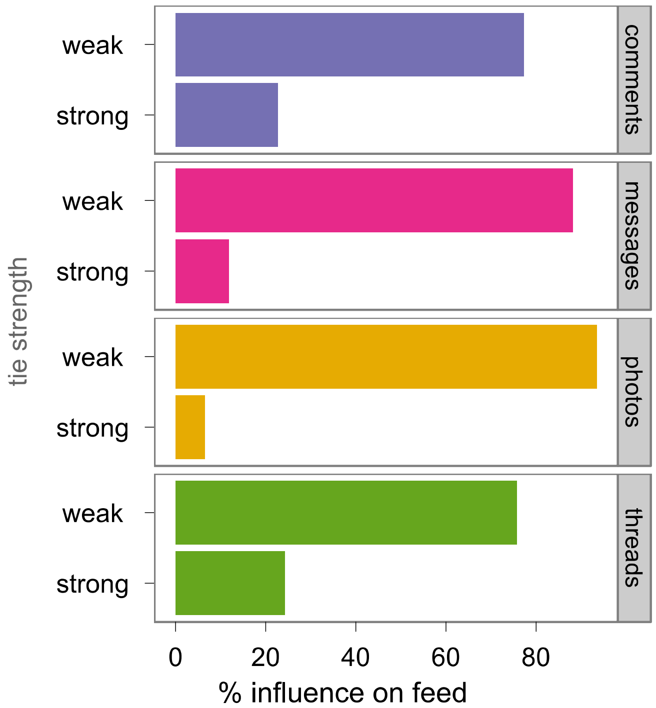
In spite c s large L

Multiple signal enhances diffusion

Daon Centola, The Spread of Behavior in an Online Social Network Experiment, Science, 2010, Vol 329

Empirical Study
- Sharing Behaviors





Multiple signal enhances Spreading/sharing

Weak links contributes dominantly to diffusion

E Bakshy, I Rosenn, C Marlow, L Adamic, The Role of Social Networks in Information Diffusion, Proceedings of the 21st International conference on WWW, 510-528

Empirical study on viral spreading
- Most popular messages on SINA Weibo

Duration	No. MSG	Popularity	No. Users	Ave. Followers	Max Popularity
Dec 2012	1000	>10,000	40M	>100	250,000

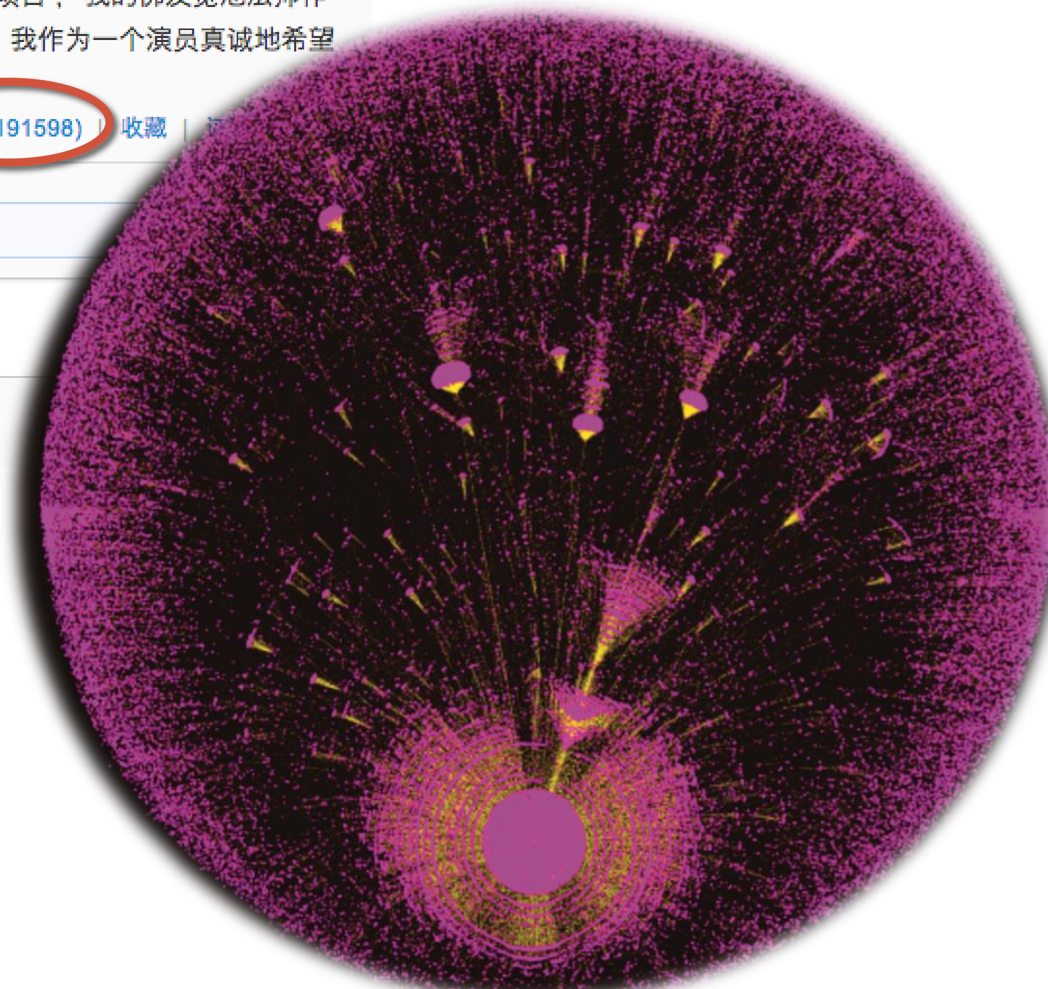
Feng Ling*, Hu Yanqing, Li Baowen, Stanley H Eugene, Havlin Shlomo, Braunstein Lidia A (2015) Competing for Attention in Social Media under Information Overload Conditions.
PLoS ONE 10(7): e0126090. doi:10.1371/journal.pone.0126090

User data

- detailed spreading path for each message

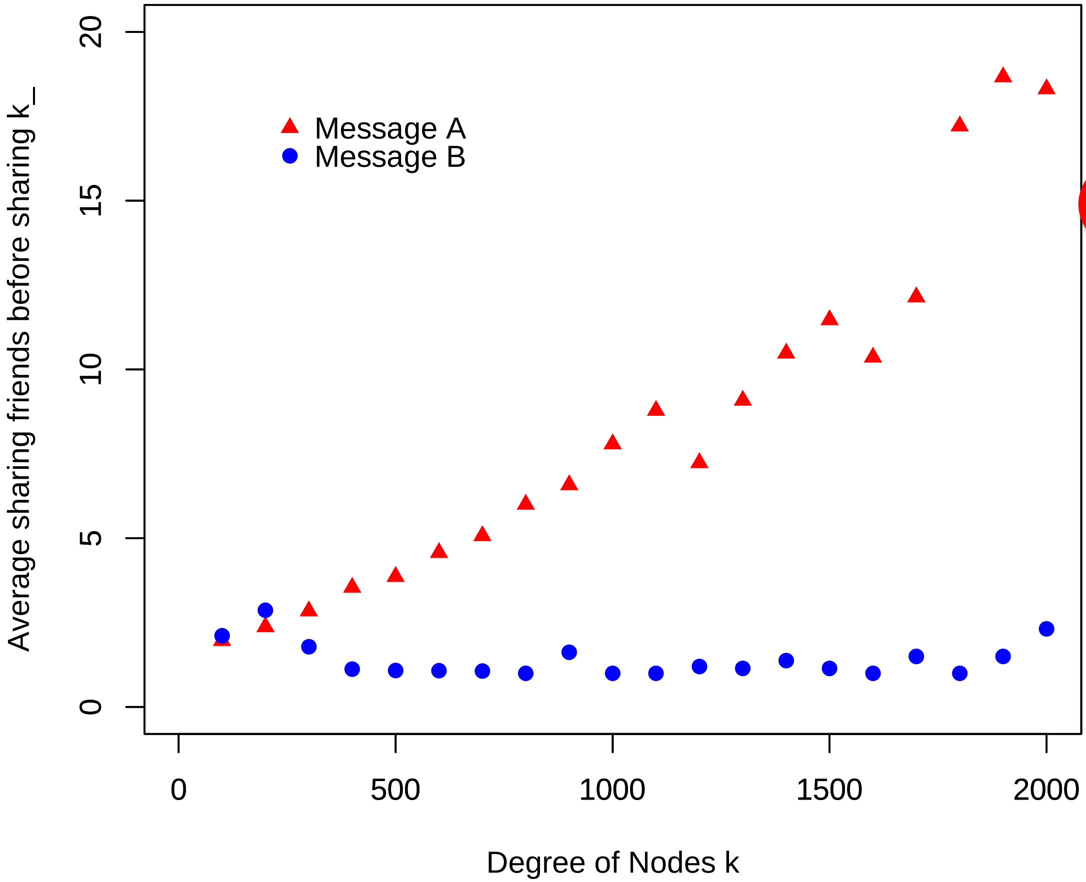
One Message	One URL
Seed node	n_0
1^{ST} Infected	n_1
No. Nei	$k_1 = 4$
Prior Infect nei	$k_1^- = 1$
Ave. k_	$< k_i^- >$
Total Users	N = 16
Spams/fake	removed
Final Popularity	R = Ni / N = 8 / 16

Example of a popular Message



- Popularity = 191598
- Purple dot : user
- Yellow line: sharing path

Empirical result I
- two types of mechanisms

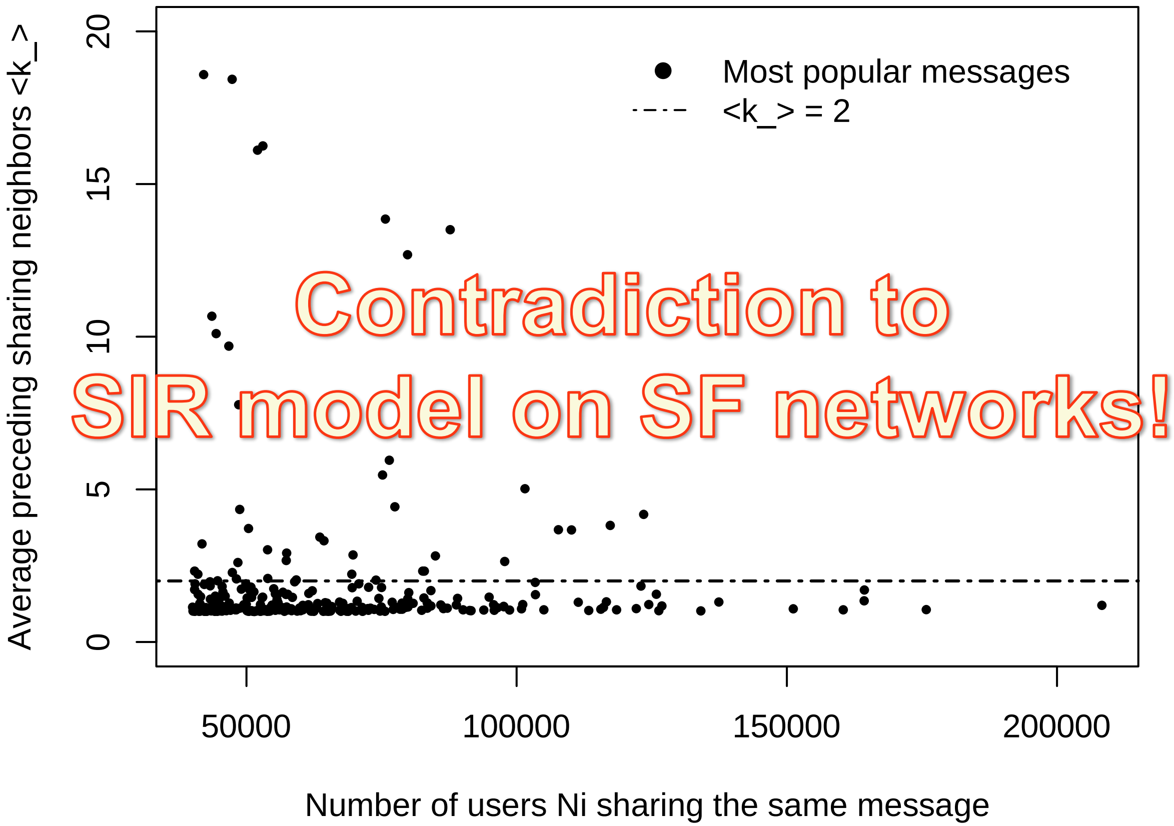


Type A:
Social reinforcement

Type B:
Immediate decision

Simple percolation

Empirical result II
- <k_> smaller than 2



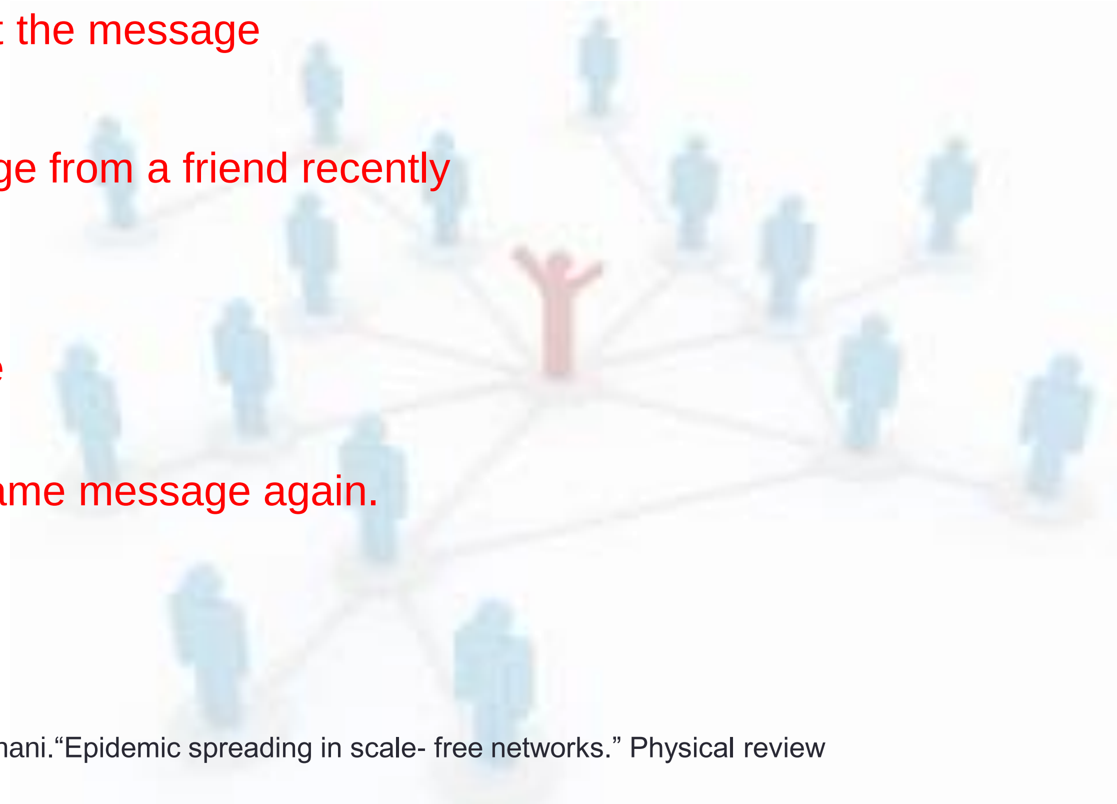
Contradiction to SIR model on SF networks!

1) Pastor-Satorras, Romualdo, and Alessandro Vespignani.“Epidemic spreading in scale- free networks.” Physical review letters 86.14 (2001): 3200.

2) Meloni, Sandro, Alex Arenas, Sergio Gmez, Javier Borge-Holthoefer and Yamir Moreno. “Modeling epidemic spreading in complex networks: concurrency and traffic.” Handbook of Optimization in Complex Networks. Springer US, 2012.

SIR model on Scale-Free networks

- Three states for nodes in epidemic (information) spreading:
  - **S** - Susceptible
    - User who have not retweet the message
  - **I** - Infected
    - Having retweet the message from a friend recently
    - Remain visible to followers
  - **R** - Recovered
    - Have retweet the message
    - No longer visible to friends
    - Not going to retweet the same message again.

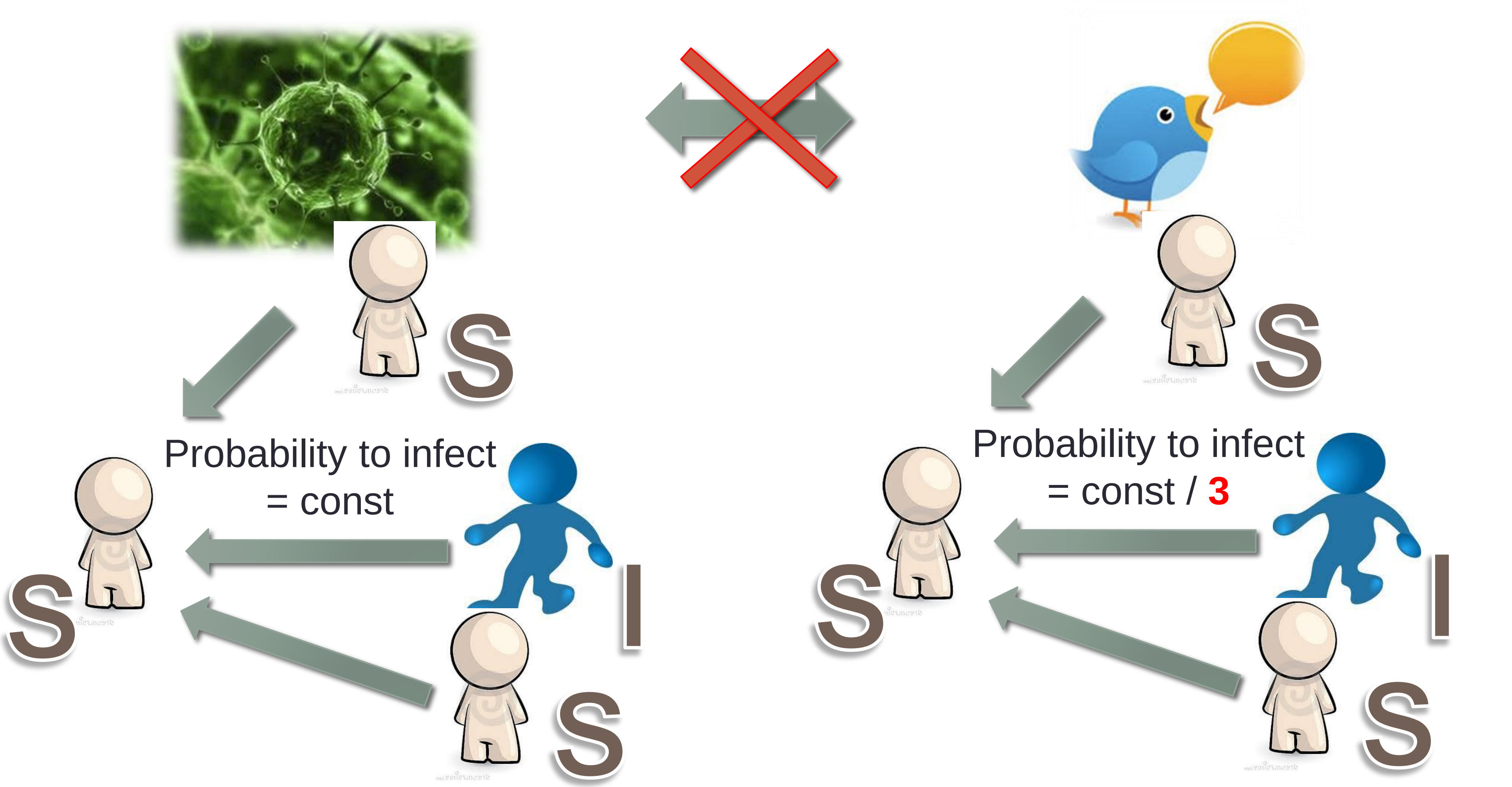


1) Pastor-Satorras, Romualdo, and Alessandro Vespignani."Epidemic spreading in scale- free networks." Physical review letters 86.14 (2001): 3200.

2) Meloni, Sandro, Alex Arenas, Sergio Gmez, Javier Borge-Holthoefer and Yamir Moreno. "Modeling epidemic spreading in complex networks: concurrency and traffic." Handbook of Optimization in Complex Networks. Springer US, 2012.

Difference between disease and tweet

<k_> larger than 2 <k_> smaller than 2



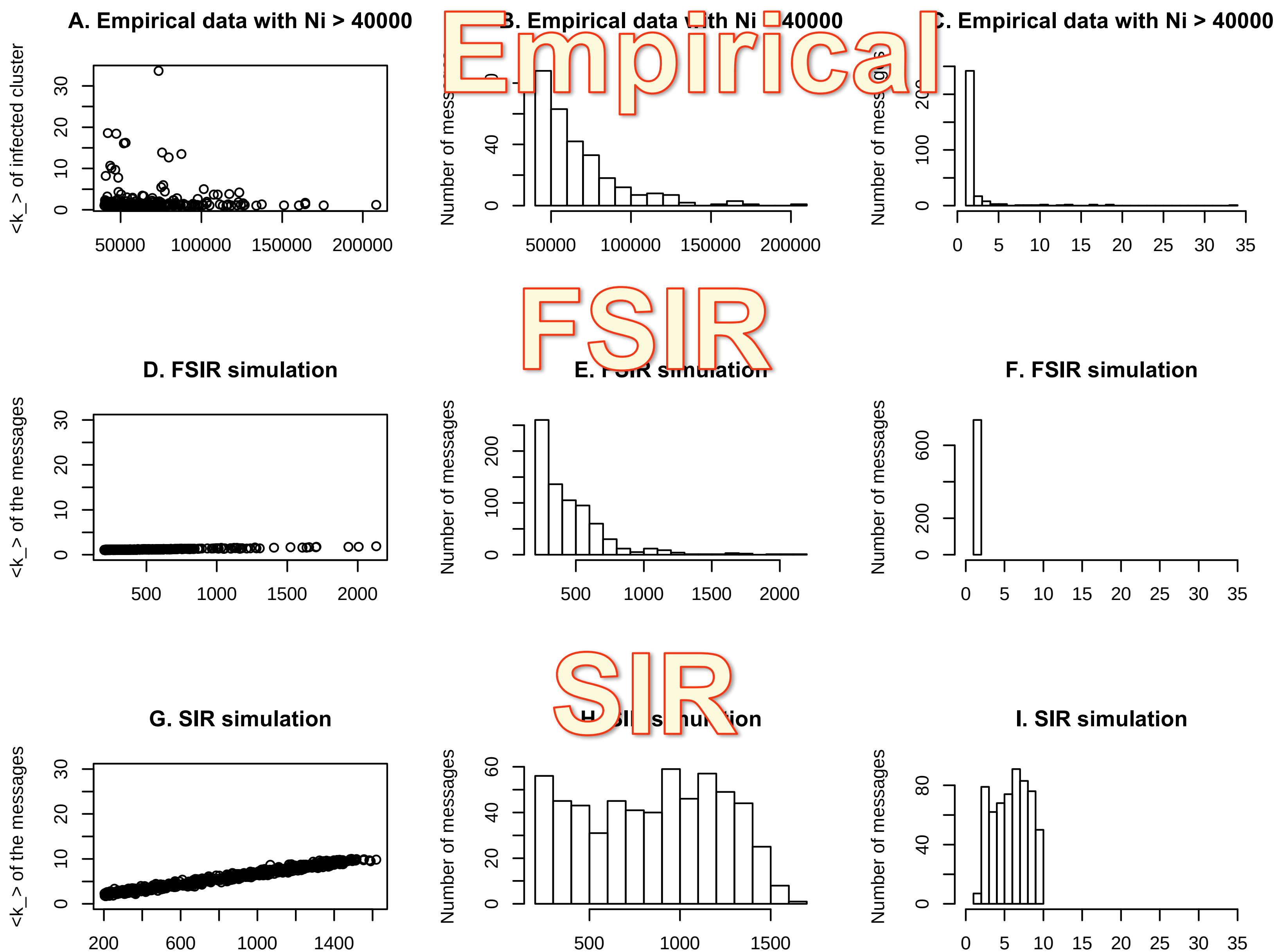
FSIR Model

- Network structure:
  - Total No. Nodes **N = 100,000**.
  - Scale-free Network (numerically resembles OSN) with
    - **$P(k) \sim k^{-2.5}$**
  - Average degree **$\langle k \rangle = 50$** (not sparse network)

- Simulation model:
  - Seed node 0 initiate a message at time 0
  - At time t, a S node i retweets the message from each of its already retweeting followees with **probability γ/k_i**.
  - After τ time steps, node i is no longer visible, and turns into R. Thus the total probability of sharing the tweet from a infected neighbor is
    - $T_i = 1 – (1- \Upsilon/k_i)^{\tau}$.
  - Spreading stops when there is no existing infected node.

The only parameter Υ is chosen to give matching infection fraction with empirical data

Empirical v.s. Model



Theoretical understanding of phase transition - Generating function

- Degree distribution p_k

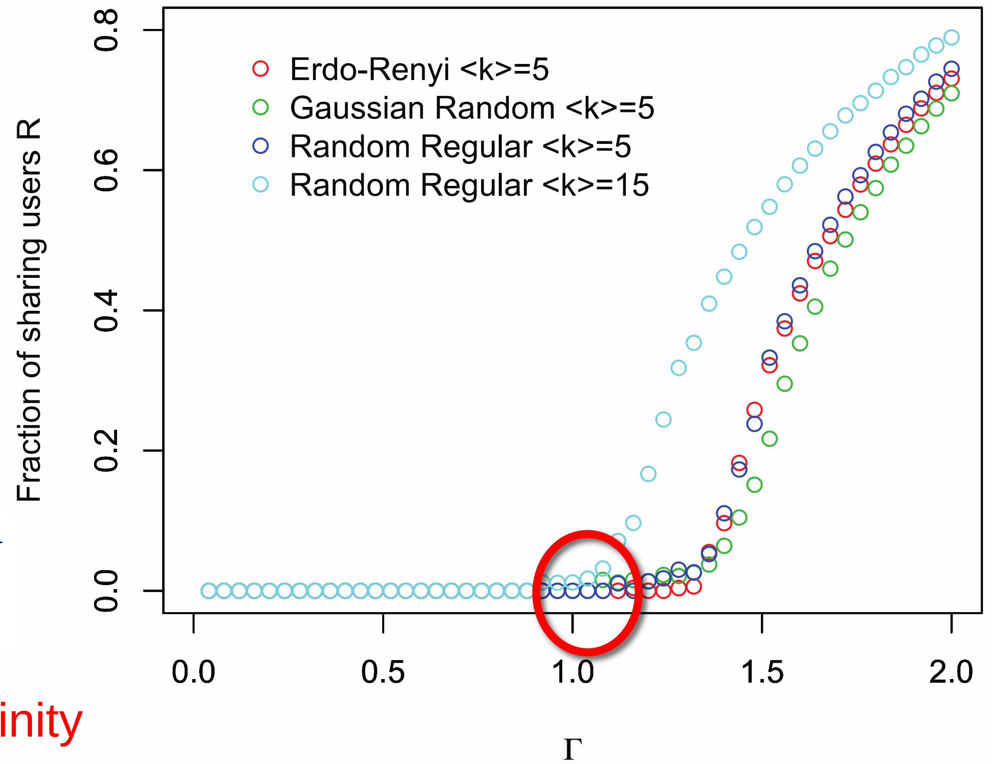
$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k$$

- Expected infected population is

$$< s > = 1 + \frac{G_0'(\Gamma, x = 1)}{1 - G_1'(\Gamma, x = 1)}$$

$$G_1(\Gamma, x) = \sum \frac{kP_k}{< k >}\left(1 + \frac{\Gamma}{k} - \frac{\Gamma}{k}x\right)^{k-1}$$

- Epidemics happens when <s> is infinity



$$\Gamma_c \approx \frac{< k >}{< k > -1}$$

Γ_c > 1 for any network structure P_k

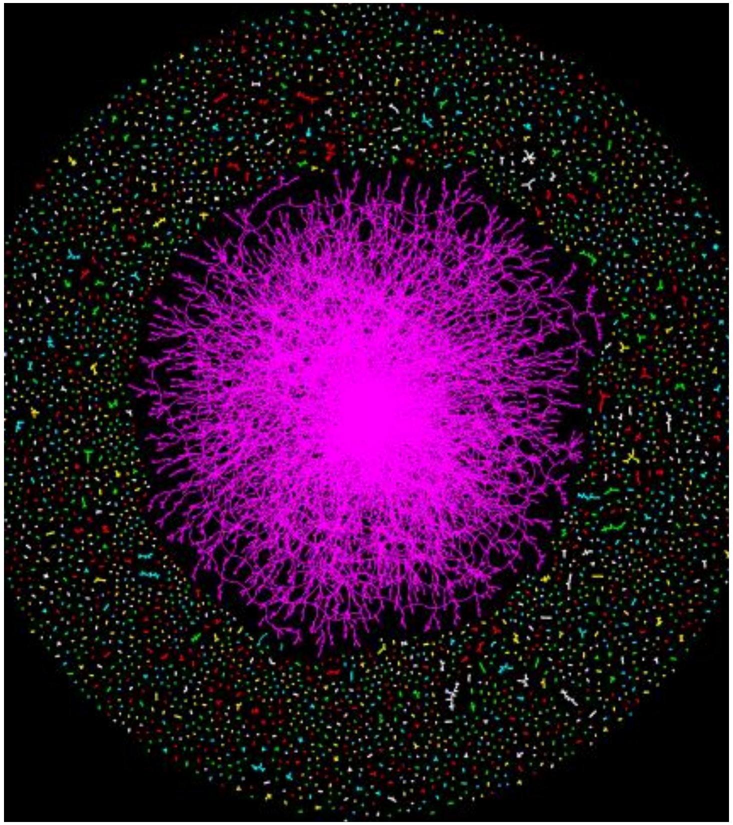
Percolation of FSIR and spreading quantification

Unique Giant Component ✔

Definite GC ✔

Small size scale of non-GC ✔

Larger N, better result ✔



Thank You

Feng Ling, fengl@ihpc.a-star.edu.sg
Complex Systems, Institute of High Performance Computing, A*STAR