Inspire…Educate…Transform.

10. Influence Propagation in Social Networks

Dr. Manish Gupta

Sr. Mentor – Academics, INSOFE

Adapted from
http://keg.cs.tsinghua.edu.cn/jietang/publications/WSDM13-tutorial-social-influence-analysis.pptx

Course Content

- Collection of three main topics of high recent interest.
  - Search engines (Crawling, Indexing, Ranking)
    - Language Modeling
    - Text Indexing and Crawling
    - Relevance Ranking
    - Link Analysis Algorithms
  - Text Processing (NLP, NER, Sentiments)
    - Natural Language Processing
    - Named Entity Recognition
    - Sentiment Analysis
    - Summarization
  - Social networks (Properties, Influence Propagation)
    - Social Network Analysis
   - **Influence Propagation in Social Networks**

Scale of Social Networks Data

- Twitter
  - A billion registered of which 316M are active
  - 6000 tweets per second
- Facebook
  - 1.49 billion monthly active users
    - 8 new users join every second
    - 0.5M likes every minute
    - 1.3M pieces of content shared every minute
    - 350M photo uploads per day
    - 31.25M messages per minute
- Instagram
  - 300M+ users
  - 48611 photos posted per minute

- Whatsapp
  - 700M users
  - 30B messages per day
- LinkedIn
  - 17361 profile views per minute
- Tinder
  - 694444 swipes per minute
- Youtube
  - 300 hours of video uploaded per minute
- Vine
  - 1041666 video loops per minute

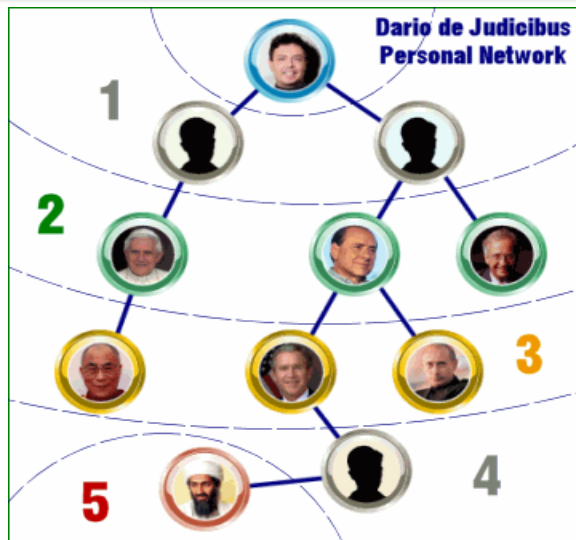# Today's Agenda

- **Does social influence really matter?**

"Love Obama" or "Hate Obama"?
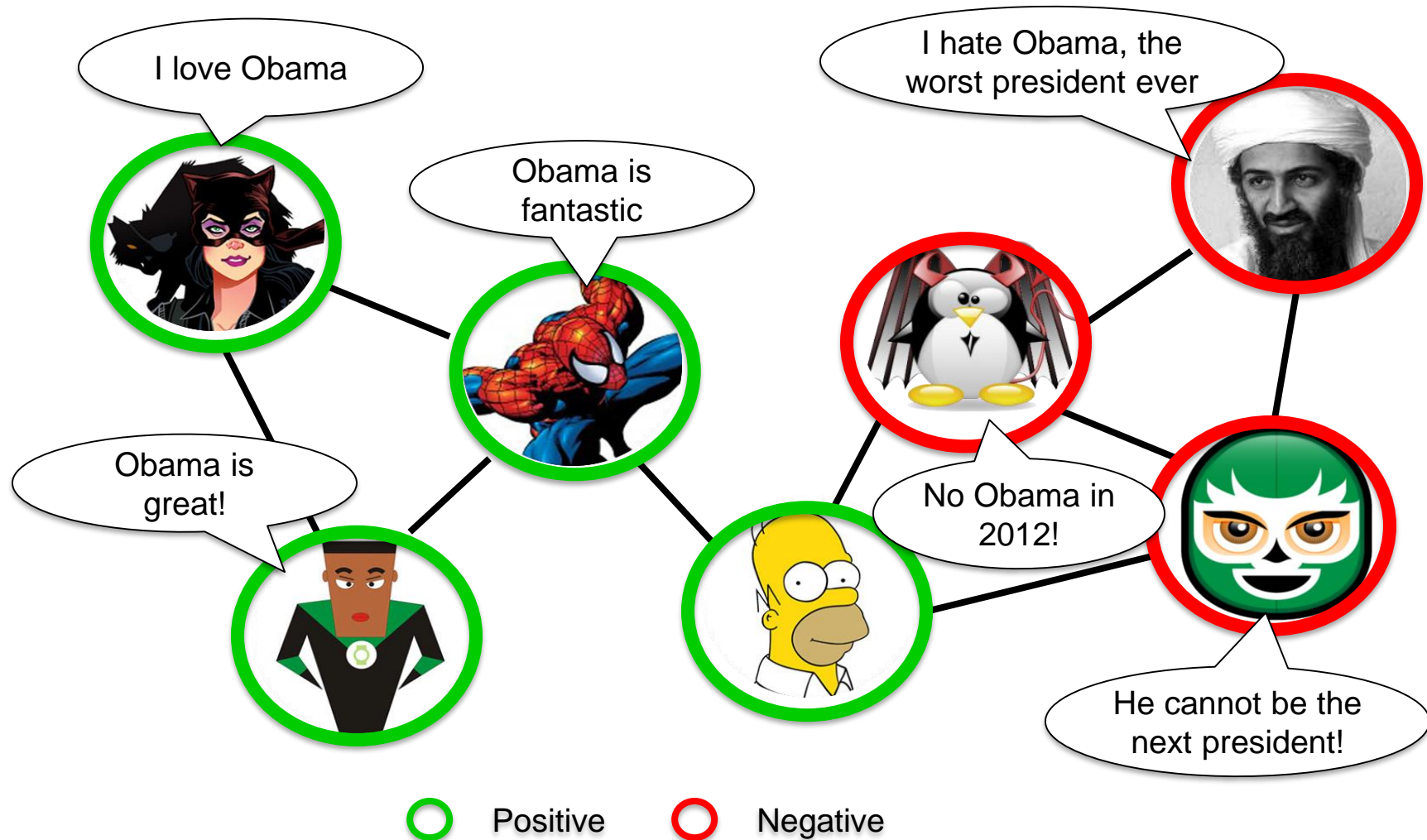
Three Degree of Influence

- Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally.[1]

Six degree of separation[2]	Three degree of Influence[3]

Dario de Judicibus Personal Network

You can **influence** >1,000,000 persons in the world, according to the Dunbar's number[4].
Humans can comfortably maintain 150 stable relationships

[1] http://en.wikipedia.org/wiki/Social_influence

[2] S. Milgram. The Small World Problem. Psychology Today, 1967, Vol. 2, 60–67

[3] J.H. Fowler and N.A. Christakis. The Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study. British Medical Journal 2008; 337: a2338
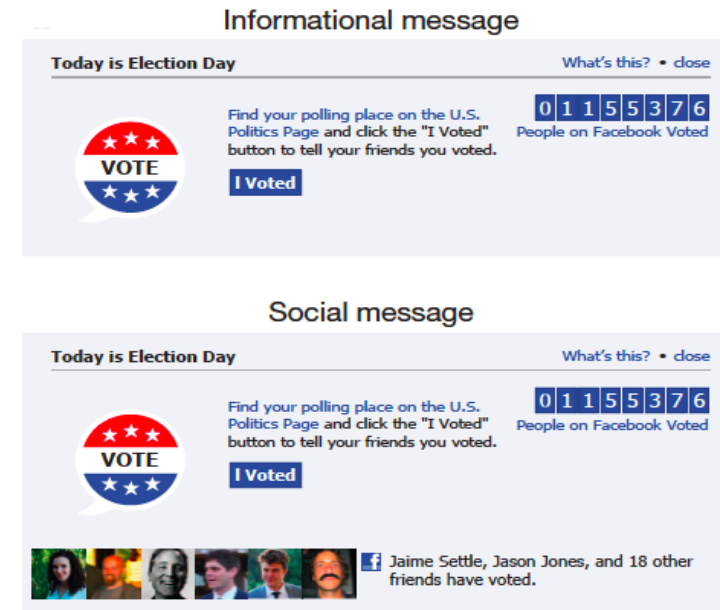
[4] R. Dunbar. Neocortex size as a constraint on group size in primates. Human Evolution, 1992, 20: 469–493.

CSE 7306c

Does Social Influence really matter?

- ## Case 1: Social influence and political mobilization[1]

  - Will online political mobilization really work?

A controlled trial (with 61M users on FB)

- Social msg group: was shown a msg that indicates one's friends who have made the votes.

- Informational msg group: was shown a msg that indicates how many others voted.

- Control group: did not receive any msg.

Informational message

Today is Election Day | What's this? • close

VOTE

Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

I Voted

0 1 1 5 5 3 7 6
People on Facebook Voted

Social message

Today is Election Day | What's this? • close

VOTE

Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

I Voted

0 1 1 5 5 3 7 6
People on Facebook Voted

f Jaime Settle, Jason Jones, and 18 other friends have voted.

Social msg group were 2.08% more likely to click on the "I Voted" button compared to the info msg group. Also, social msg group were 0.39% more likely to **actually vote** (via examination of public voting records) than control group.

CSE 7306c

Case 2: Klout[1]—Social Media Marketing

- Toward measuring real-world influence
  - Twitter, Facebook, G+, LinkedIn, etc.
  - Klout generates a score on a scale of 1-100 for a social user to represent her/his ability to engage other people and inspire social actions.
- Though controversial[2], in May 2012, Cathay Pacific opened SFO lounge to Klout users
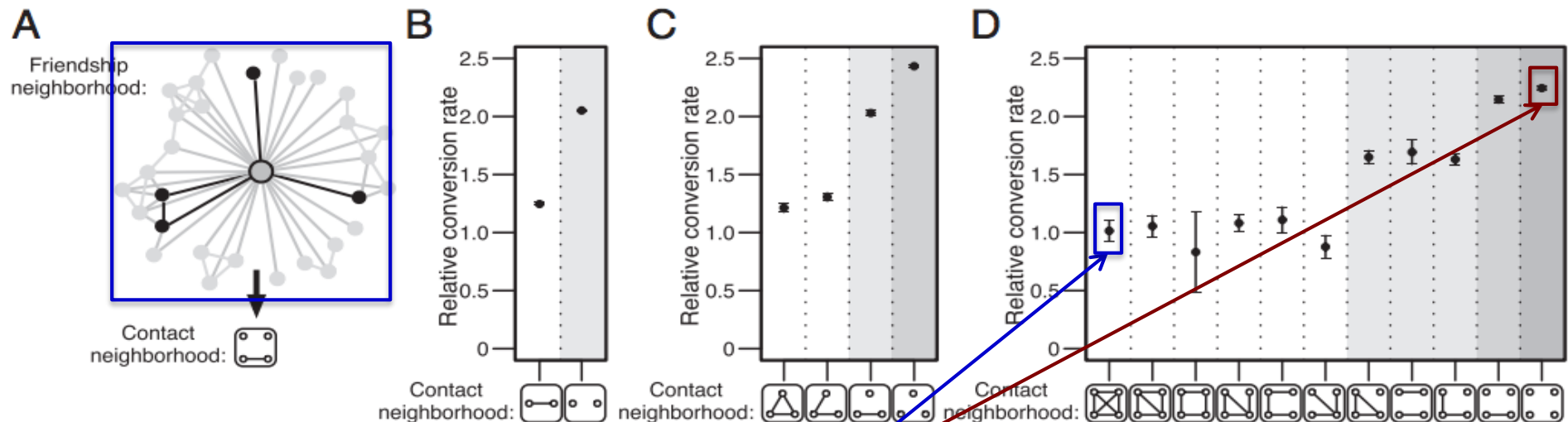 - A high Klout score gets you into Cathay Pacific's SFO lounge

[1] http://klout.com

[2] Why I Deleted My Klout Profile, by Pam Moore, at Social Media Today, originally published November 19, 2011; retrieved November 26 2011

Case 3: Who Influenced you and How?

- Magic: the structural diversity of the ego network[1]


Results: Your behavior is influenced by the "structural diversity" (the number of connected components in your ego network) instead of the number of your friends.

[1] J. Ugandera, L. Backstromb, C. Marlowb, and J. Kleinberg. Structural diversity in social contagion. PNAS, 109 (20):7591-7592, 2012.
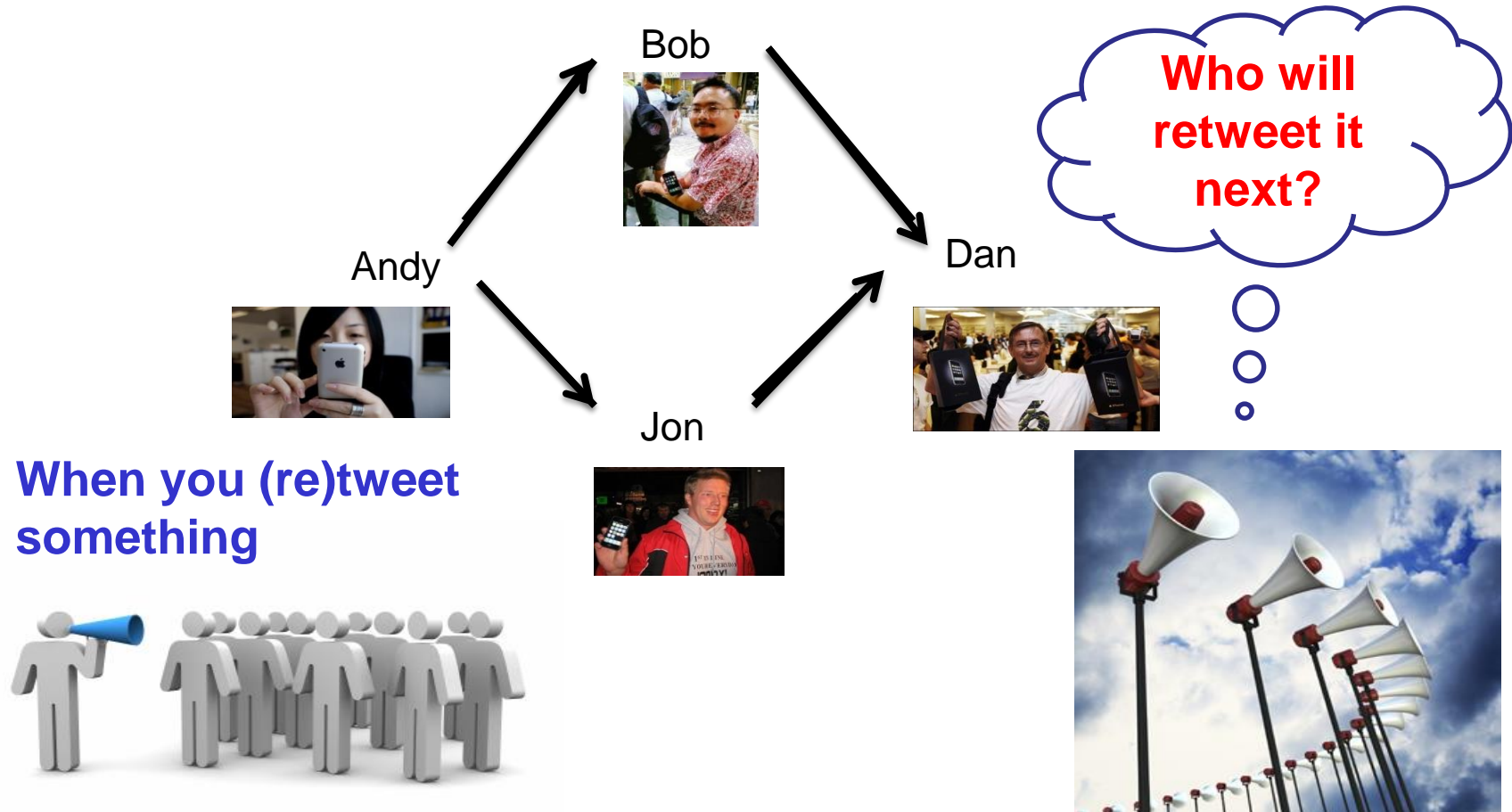
Today's Agenda

- Does social influence really matter?

- **Examples of social influence**

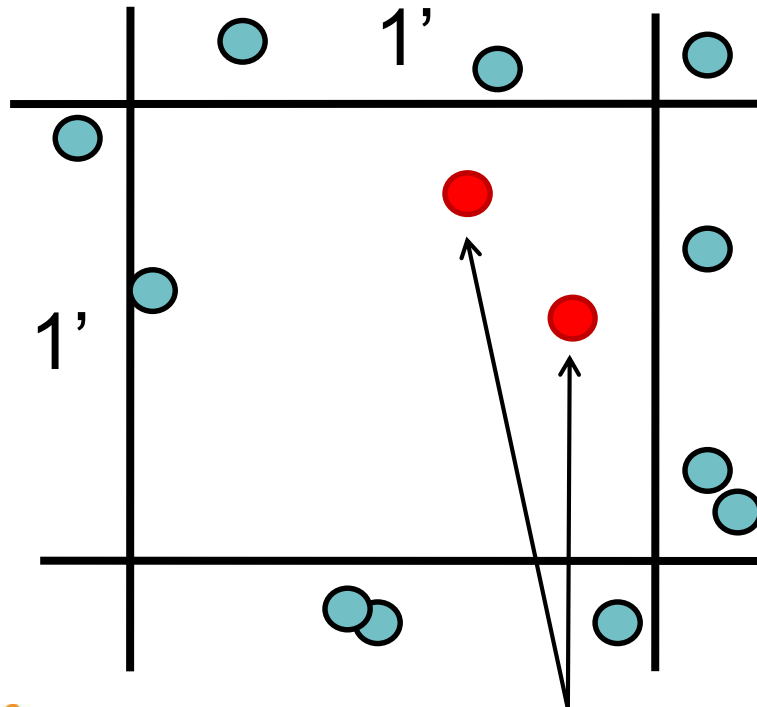# Types of Social Influence: Follow Influence on Twitter
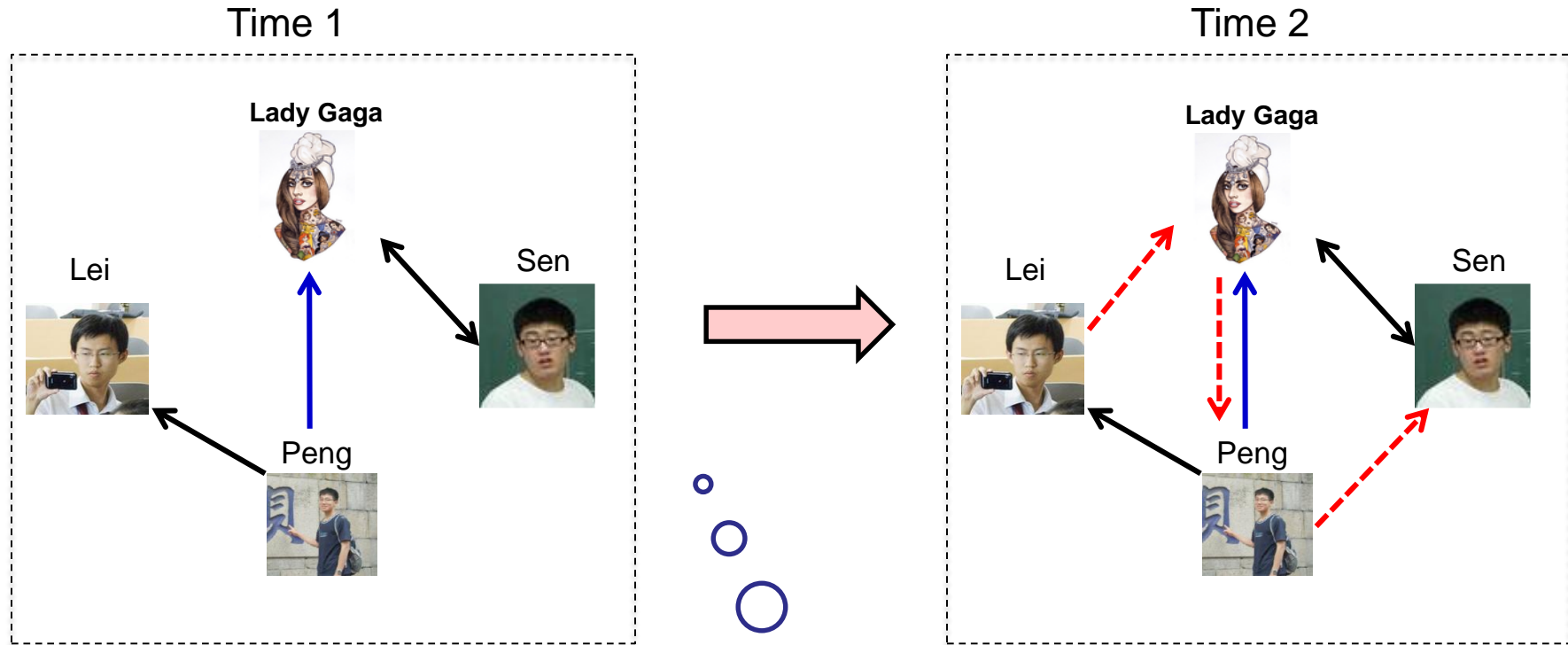
When you **follow** a user in a social network, will the behavior **influence** your friends to also follow her?

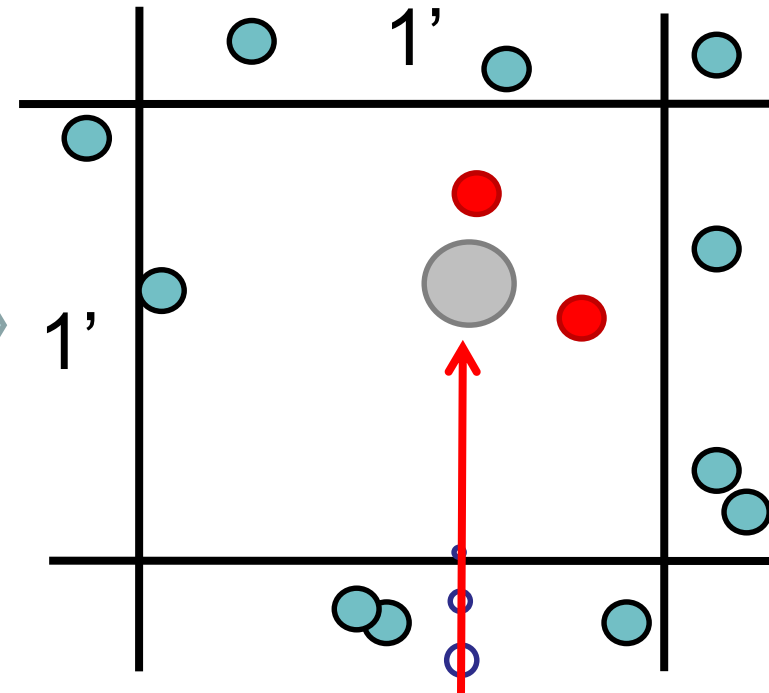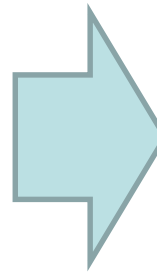# Types of Social Influence: Retweet Influence

Bob

Andy

Dan

Jon

When you (re)tweet something

Who will retweet it next?

Types of Social Influence: Check-in Influence in Gowalla

Legend: ◯ Alice, ● Alice's friend, ◯ Other users

1'

If Alice's friends check in this location at time t

Will Alice also check in nearby?

CSE 7306c

Today's Agenda

- Does social influence really matter?

- Examples of social influence

- **Ways of measuring social influence**

Test for Existence of Social Influence

- Randomized Control Trials (RCT)
  - People are randomly assigned to a "treatment" group or a "controlled" group;
  - People in the treatment group receive some kind of "treatment", while people in the controlled group do not receive the treatment;
  - Compare the result of the two groups, e.g., survival rate with a disease with/without a treatment.

- Two challenges:
  - How to define the treatment group and the controlled group?
  - How to find a real random assignment?

[1] Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66, 5, 688–701.
[2] http://en.wikipedia.org/wiki/Randomized_experiment

CSE 7306c

Example: Political mobilization

- There are two kinds of treatments.

A controlled trial

A controlled trial

Treatment for Group 1

- Social msg group: was shown with msg that indicates one's friends who have made the votes.

Treatment for Group 2

- Informational msg group: was shown with msg that indicates how many others voted.

- Control group: did not receive any msg.

Informational message

Social message

Treatment for Group 1

Treatment for Group 1&2

[1] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. Nature, 489:295-298, 2012.

Adoption Diffusion of Y! Go (1)

Yahoo! Go is a product of Yahoo to access its services of search, mailing, photo sharing, etc.


Jul 04   Aug 10   Oct 29

- Non- adopter
- Old adopter
- Recent adopter
- Link between adopter and recent adopter
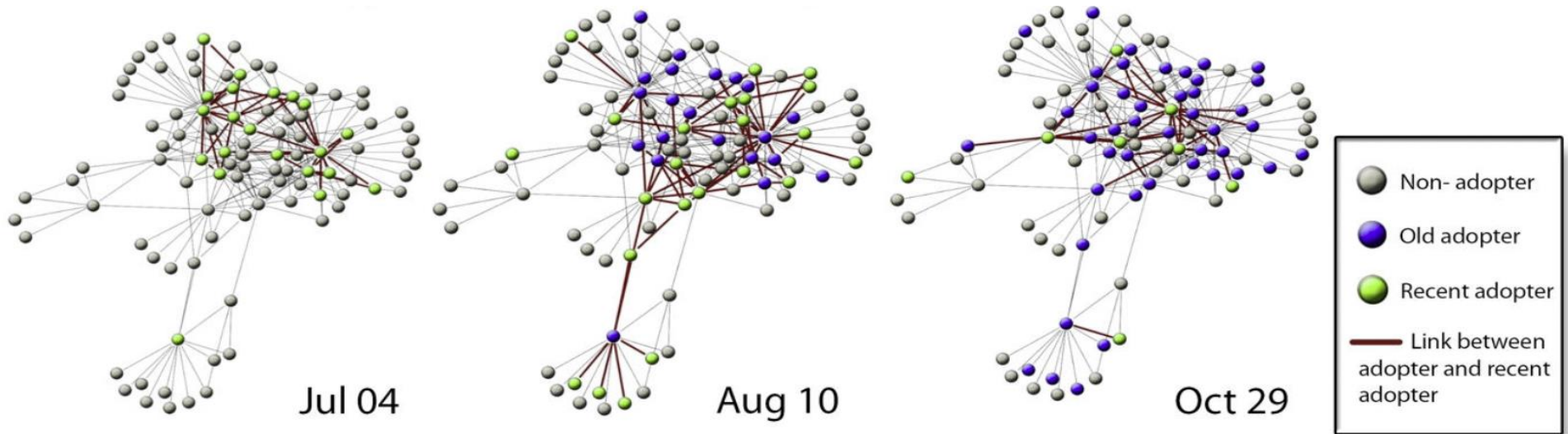
RCT:
- **Treatment group:** people who adopted Y! Go and also have friend(s) who adopted Y! Go at time t;
- **Controlled group:** people who adopted Y! Go and have no friends who adopted Y! Go at time t.

[1] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. PNAS, 106 (51):21544-21549, 2009.

CSE 7306c

Evidence of Influence?
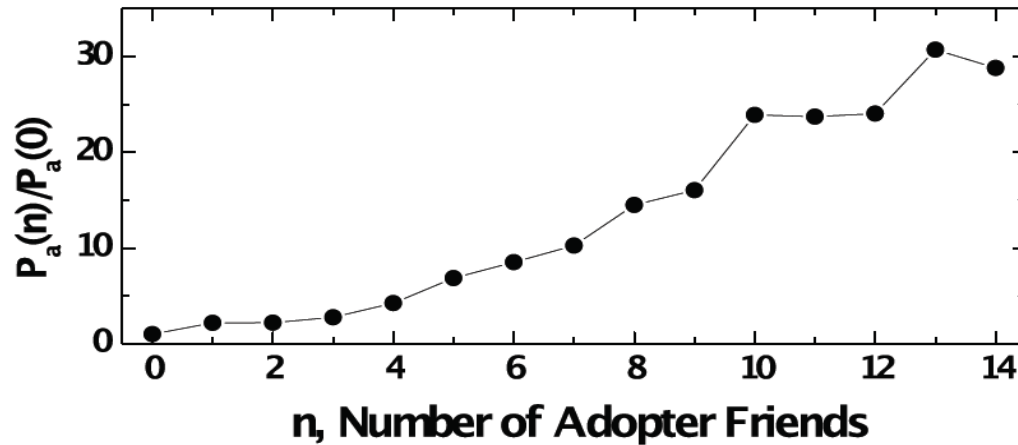
The ratio of the likelihood of adoption given n adopter friends $P_a(n)$ and the likelihood of adoption given 0 adopter friends $P_a(0)$ where the number of adopter friends is assessed at the time of adoption.

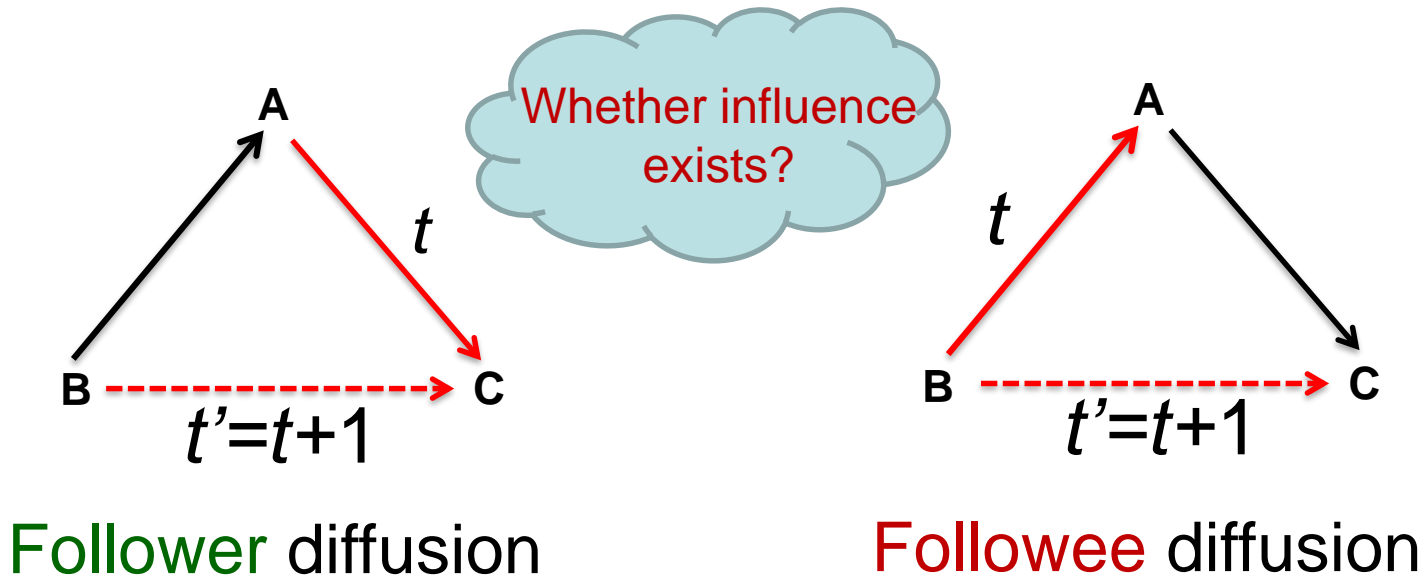**Adopters are 100 times more likely to have 12 adopter friends than non-adopters**

CSE 7306c
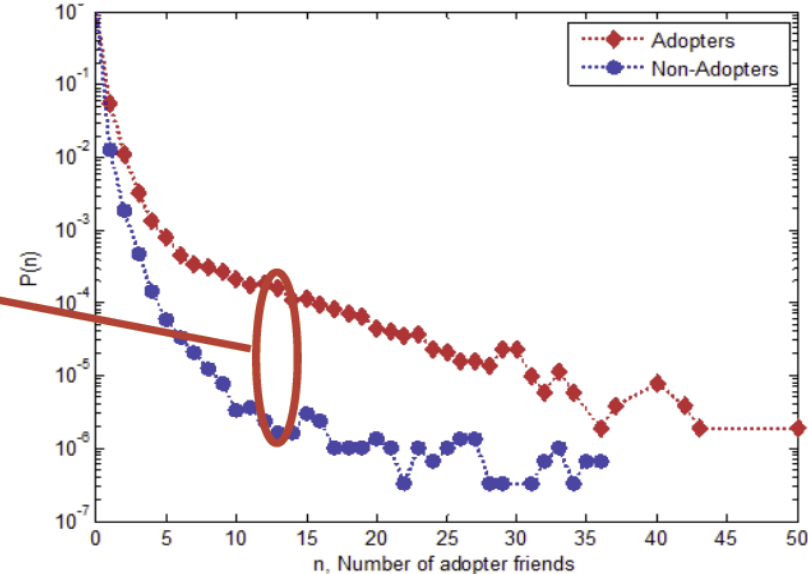
Two More Methods

- **Shuffle test:** shuffle the activation time of users.
  - If social influence does not play a role, then the timing of activation should be independent of the timing of activation of others.

- **Reverse test:** reverse the direction of all edges.
  - Social influence spreads in the direction specified by the edges of the graph, and hence reversing the edges should intuitively change the estimate of the correlation.
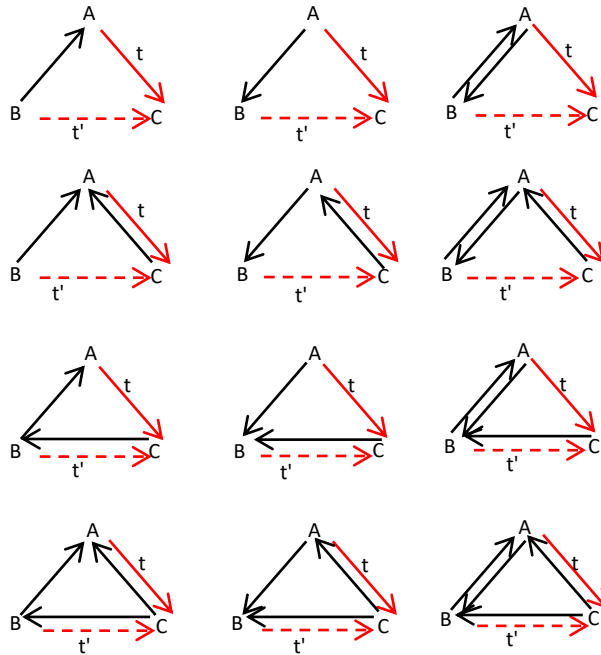
CSE 7306c

Influence Test via Triad Formation

Two Categories of Following Influences

Whether influence exists?

t

$t'=t+1$

Follower diffusion

t

$t'=t+1$

Followee diffusion

—>: pre-existed relationships
—>: a new relationship added at t
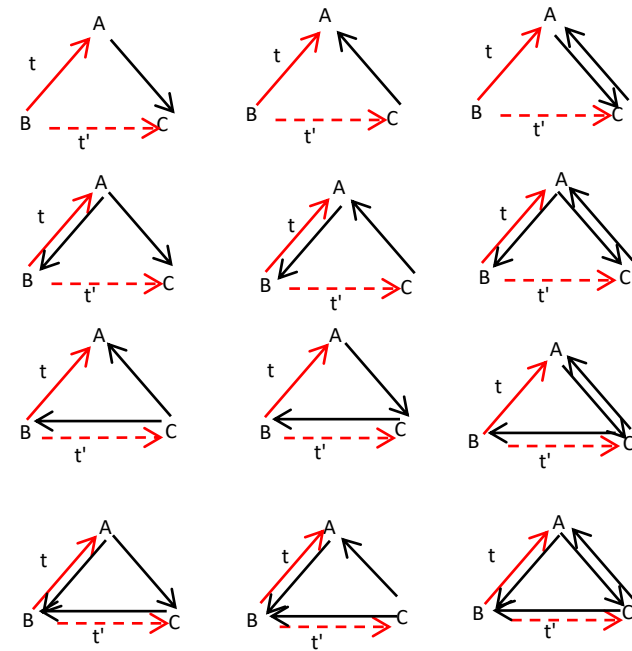-->: a possible relationship added at $t+1$

24 Triads in Following Influence
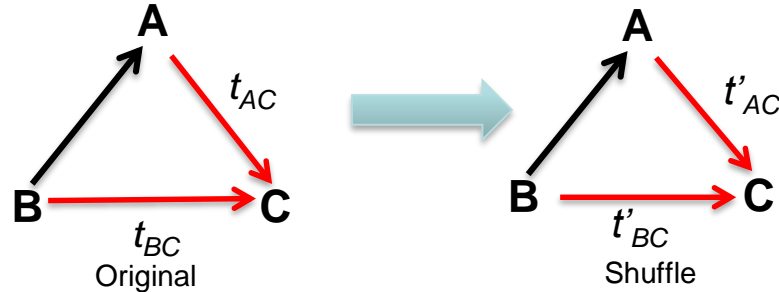
Follower diffusion

12 triads

Followee diffusion

12 triads

Timing Shuffle Test

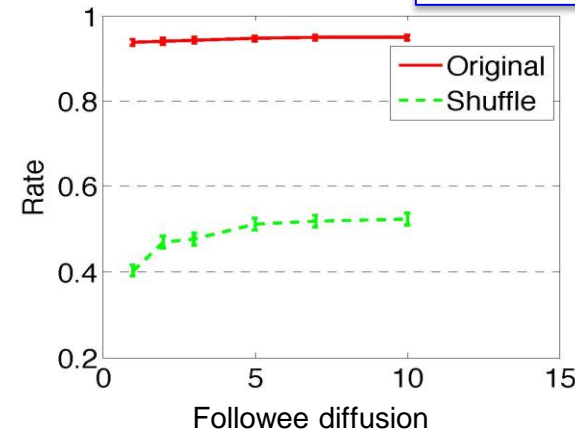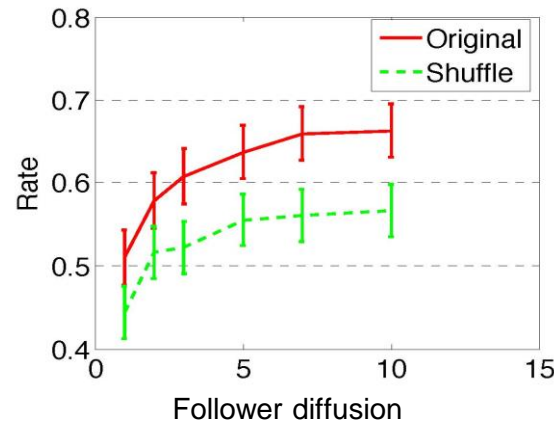- Method: Shuffle the timing of all the following relationships.

Shuffle test

$$Rate = \frac{\#Triad \quad | \quad 0 < t_{BC} - t_{AC} < d}{\#Triad \quad | \quad t_{BC} \quad and \quad t_{AC} \quad exist}$$

- Compare the rate under the original and shuffled dataset.

t-test, *P*<0.01

- Result

[1] A. Anagnostopoulos, R. Kumar, M. Mahdian. Influence and correlation in social networks. In KDD, pages 7-15, 2008.
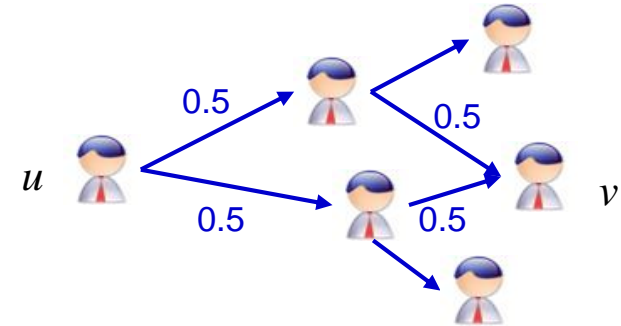
Reachability-based Method[1]

- PageRank, Random Walk with Restart

- Influence of a path

$$\text{inf}(p) = \widetilde{\bigcirc}_{v_i \hat{\text{l}} \, p} \frac{1}{\text{outdeg}(v_i)}$$

- Influence of user u on v

Influence(u, v)
=0.5*0.5+0.5*0.5

$$\text{influence}(u,v) = \lim_{t \to \infty} \sum_{p \in path_t(u,v)} \text{inf}(p)$$

All paths from u to v within path length t

Note: The method only considers the network information and does not consider the content information

[1] G. Jeh and J. Widom. Scaling personalized web search. In WWW '03, pages 271-279, 2003.

Action-based Methods: Learning Influence Probabilities [1]

- **Goal:** Learn user influence from historical actions

- **Assumption**
  - If user v_i performs an action y at time t and later his friend v_j also perform the action, then there is an influence from v_i to v_j

- **User Influenceability: quantifies how influenceable a user is.**

Action propagation from v_i to v_j within Δt | threshold

$$\text{influence}(v_i) = \frac{\left| \{y \mid \$v_j, \mathrm{D}t : \boxed{prop(y, v_i, v_j, \mathrm{D}t)} \wedge \boxed{t} \geq \mathrm{D}t \geq 0\} \right|}{Y_{v_i}}$$

where $\Delta t = t_j - t_i$ is the difference between the time when v_j performing the action and the time when user v_i performing the action, given $e_{ij}=1$.
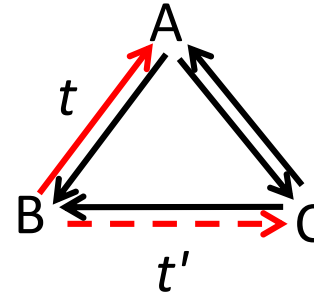
[1] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In WSDM'10, pages 207–217, 2010.

CSE 7306c

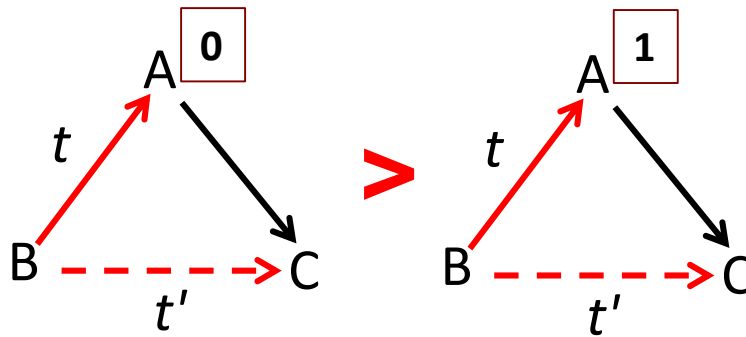# Social Theories: Structural Balance[1]
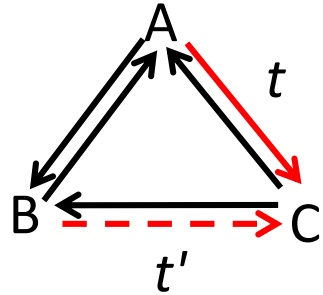
Follower diffusion

Followee diffusion

Social Balance: my friend's friend is also my friend

Explanation: Users have tendency to form a balanced triad

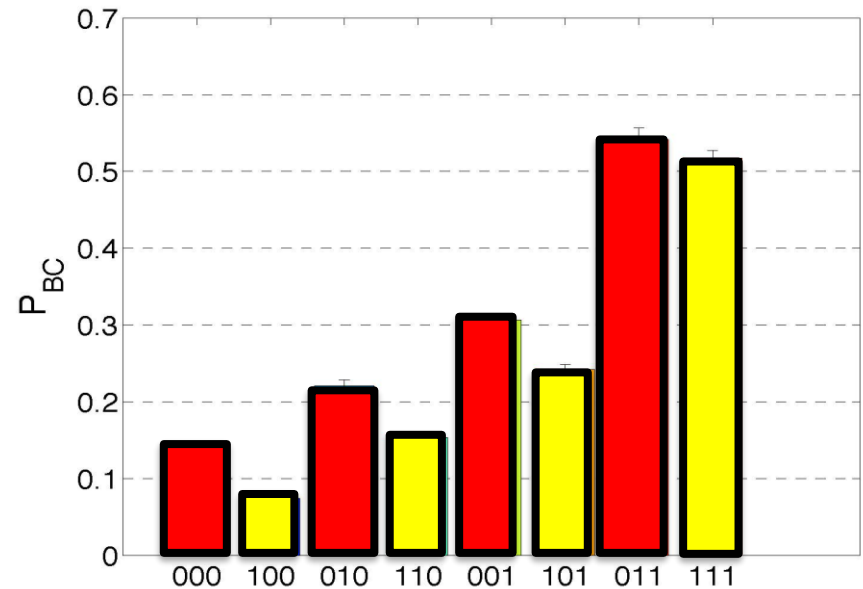Fritz Heider (1958). The Psychology of Interpersonal Relations. John Wiley & Sons.

Social Theories: Social Status
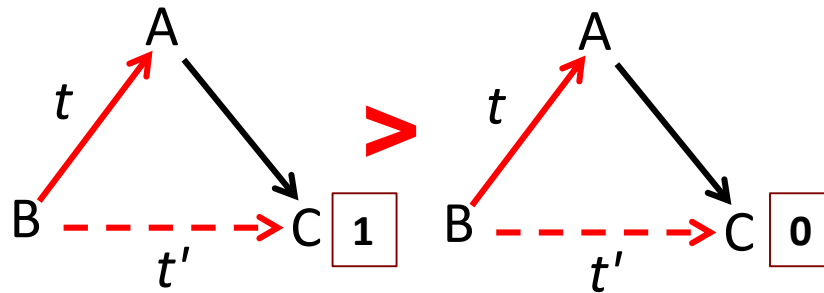
Followee diffusion: $P(0XX) > P(1XX)$

1: Elite user
0: Low-status user

- Low-status users act as a bridge to connect users so as to form a closure triad.
- The likelihood of 0XX is 1.4 times of 1XX.

Social Theories: Social Status

Followee diffusion: P(XX1) > P(XX0)

1: Elite user
0: Low-status user

- The rich gets richer.
- The likelihood of XX1 is nearly 2 times higher than that of XX0.
- This phenomenon validates the mechanism of preferential attachment.
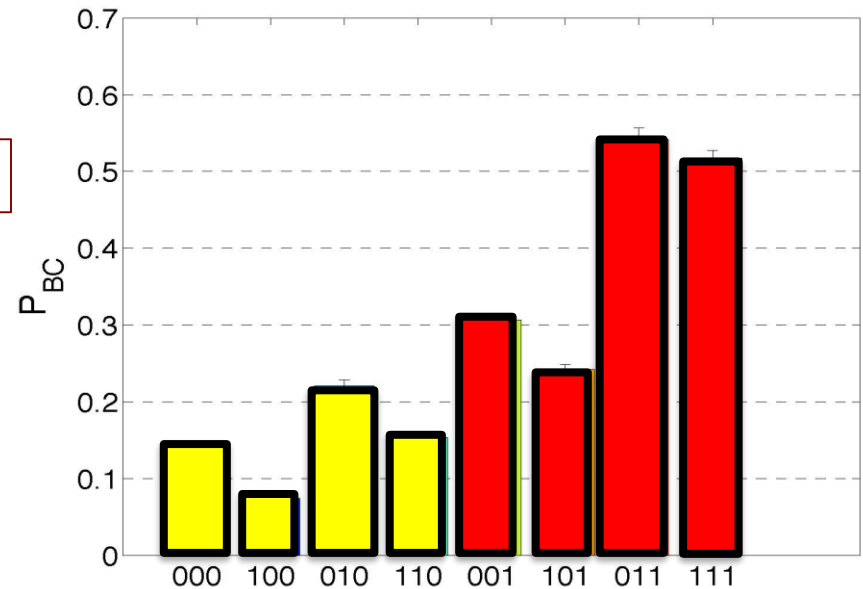
Output of Measuring Social Influence
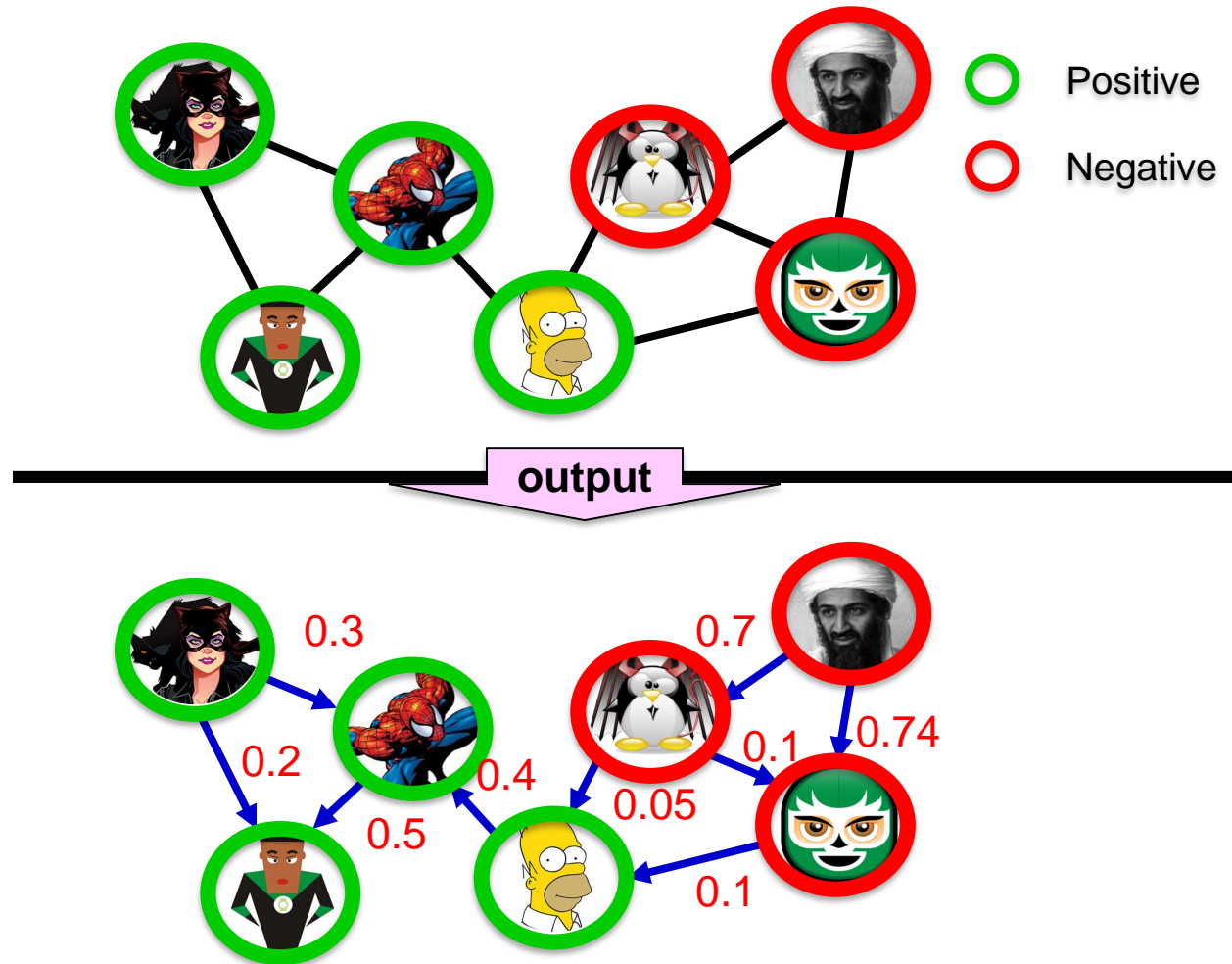
Positive
Negative

output

0.3
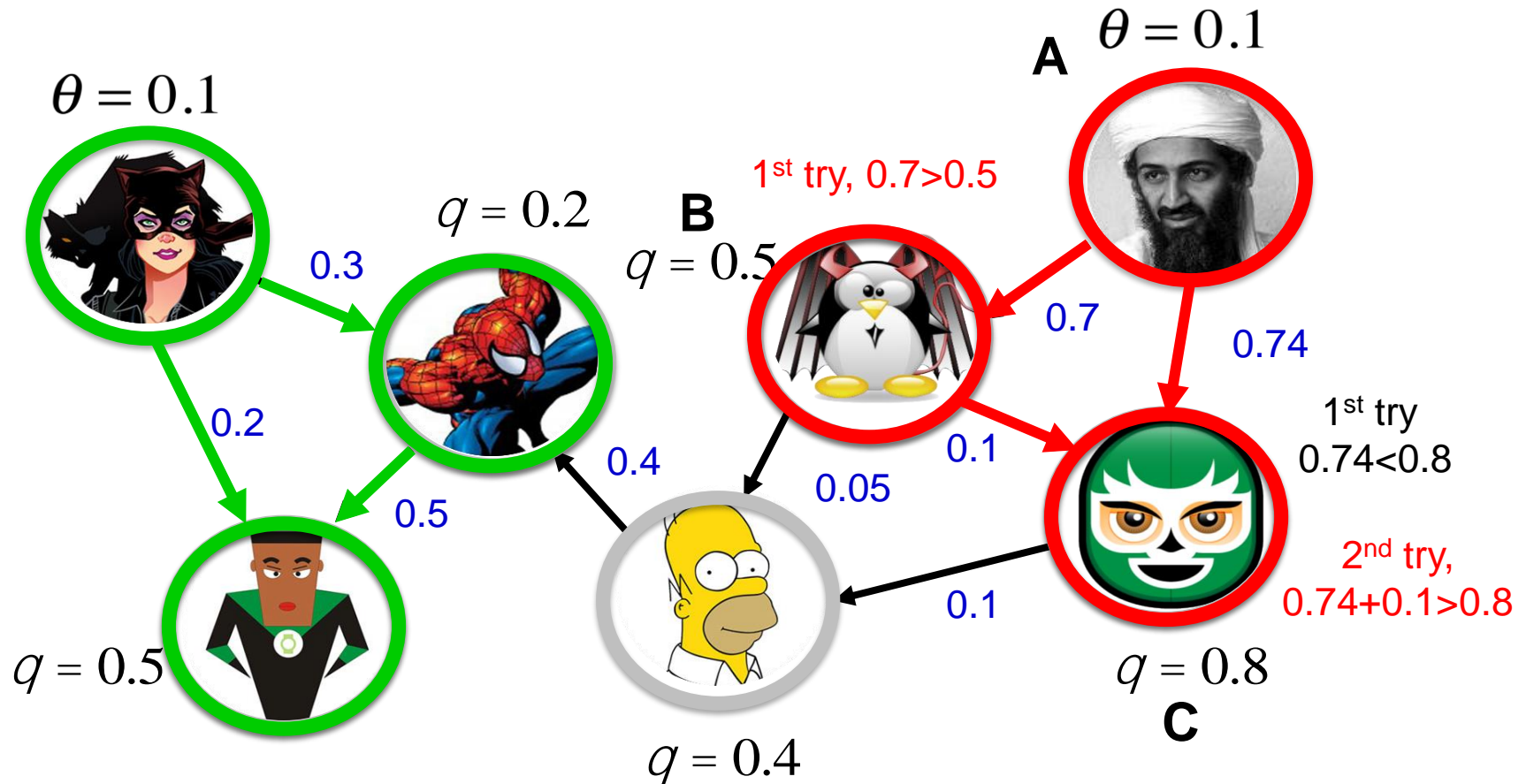0.2
0.4
0.5
0.7
0.1
0.74
0.05
0.1

CSE 7306c

Today's Agenda

- Does social influence really matter?

- Examples of social influence

- Ways of measuring social influence

- **Models for social influence analysis**

Models for Social Influence Analysis

- ## Decision based models

  - Models of product adoption, decision making
    - A node observes decisions of its neighbors and makes its own decision
    - E.g., You join demonstrations if at least *k (threshold)* of your friends do so too
    - Linear Threshold Model

- ## Probabilistic Models

  - Models of influence or disease spreading
    - An infected node tries to "push" the contagion to an uninfected node
    - E.g., You "catch" a disease with some prob. from each active neighbor in the network
    - Cascade Model

Linear Threshold Model: An Example

$\theta = 0.1$

A $\theta = 0.1$

1st try, 0.7>0.5

$q = 0.2$

B $q = 0.5$

0.3

0.7

0.74

0.2

0.4

0.5

0.1

1st try 0.74<0.8

0.05

0.1

$q = 0.5$

$q = 0.4$

0.1

$q = 0.8$

2nd try, 0.74+0.1>0.8

C

Probabilistic Models: Independent Cascade Model

- Initially some nodes *S* are active

- Each edge *(u,v)* has probability (weight) p_{uv}

- When node v becomes active
  - It activates each out-neighbor *v* with prob. p_{uv}

- Activations spread through the network

- Independent cascade model is simple but requires many parameters!
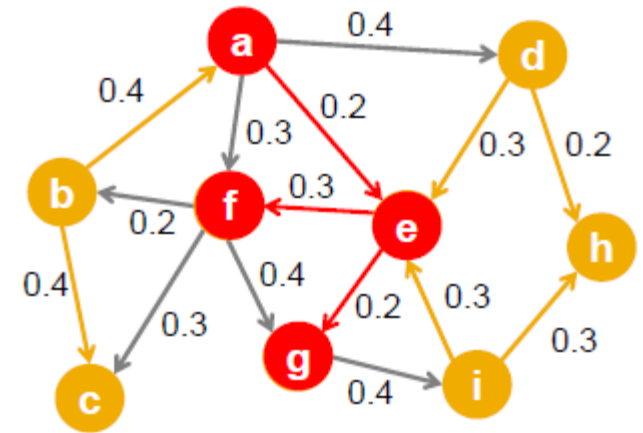
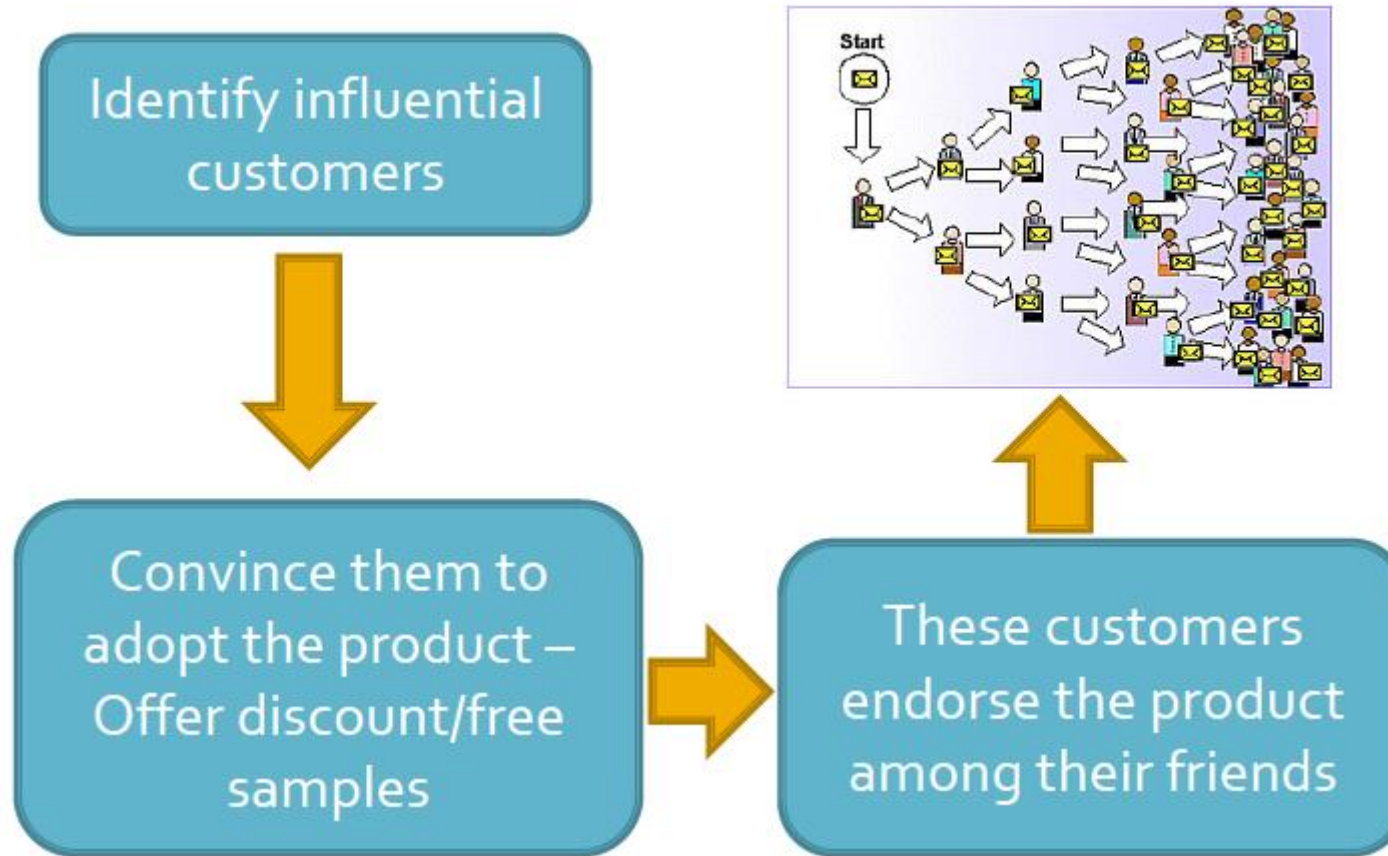# Today's Agenda

- Does social influence really matter?

- Examples of social influence

- Ways of measuring social influence

- Models for social influence analysis

- **The influence maximization problem**

Why Maximize Influence?

Identify influential customers

↓

Convince them to adopt the product – Offer discount/free samples

→

These customers endorse the product among their friends

↑

Who is influential? Influence Spread
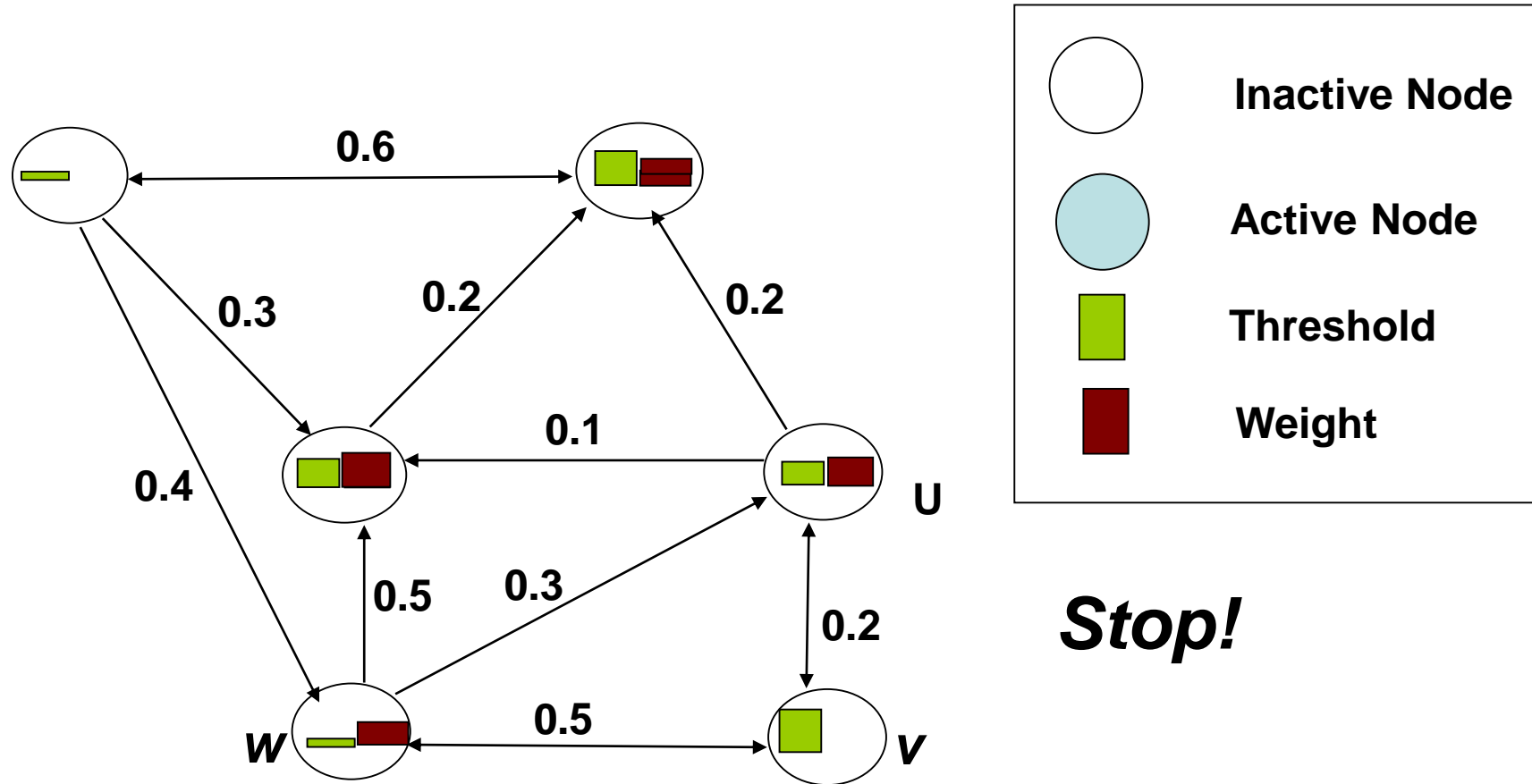
Source: David Kempe's slides

Influence Maximization

- Influence of a set of nodes A, i.e., $\sigma(A)$, is the number of active nodes at the end of the process starting from A.
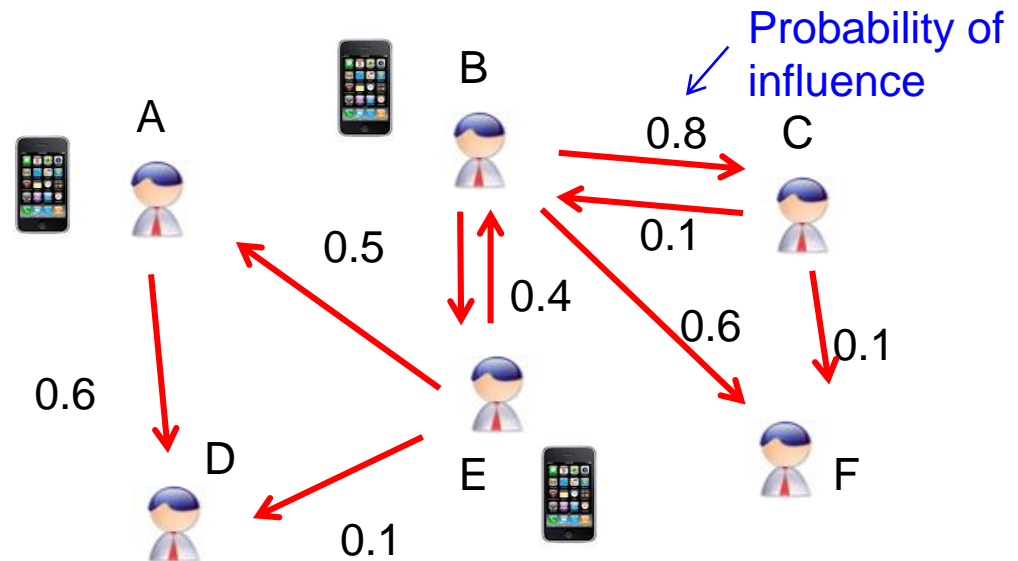
- Problem Definition:
  - Given a parameter k, find a k-node set A to maximize $\sigma(A)$.

- It is an NP-hard problem.

Probability of influence

Influence Maximization Problem

- Find an approximation algorithm for the influence maximization problem.

- Using a greedy algorithm, one can obtain a $(1-1/e)$ optimal result

- Greedy Algorithm

  – Start with an empty set A

  – Choose an element that provides the largest marginal increase in the function value $\sigma(A)$.

  – Until $|A| = k$

Algorithms

- General Greedy
  - In each round, the algorithm adds one vertex into the selected set *S* such that this vertex together with current set *S* maximizes the influence spread.
    - Involves running random cascade model from each node

- Low-distance Heuristic: nodes with the shortest paths to other nodes as seed nodes
  - Individuals are more likely to be influenced by those who are closely related to them.

- High-degree heuristic
  - Nodes with more neighbors would arguably tend to impose more influence upon its direct neighbors.

- Degree Discount Heuristic
  - If *u* has been selected as a seed, then when considering selecting *v* as a new seed based on its degree, we should not count the edge *v->u* and also edges from v to u's influence set.

CSE 7306c

Take-away Messages

- We discussed basics of social influence and observed examples where social influence is prominent.

- We discussed various tests to test for existence of social influence.

- We studied the decision based and probabilistic models for Social Influence Analysis

- We also visited the Influence Maximization and studied various approximate and heuristic solutions

Further Reading

- Jimeng Sun, Jie Tang: A Survey of Models and Algorithms for Social Influence Analysis. 177-214. Social Network Data Analytics. Springer. 2011

- Chapter 19 and 21 of "Easley-Kleinberg" book
  - http://www.cs.cornell.edu/home/kleinber/networks-book/

- R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. Nature, 489:295-298, 2012.

- P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'01), pages 57–66, 2001.

International School of Engineering

Plot 63/A, 1st Floor, Road # 13, Film Nagar, Jubilee Hills, Hyderabad - 500 033

For Individuals:	+91-9502334561/63 or 040-65743991
For Corporates:	+91-9618483483
Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering