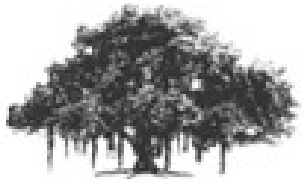


IIIT-H

Web Mining
Lecture 11: Social Influence Analysis (Part 1)

Manish Gupta

4th Sep 2013

Slides borrowed (and modified) from
http://keg.cs.tsinghua.edu.cn/jietang/publications/WSDM13-tutorial-social-influence-analysis.pptx

Recap of Lecture 10: Social Network Analysis (Part 2)

- Preferential Attachment Model
- Copying Model, Forest Fire Model
- Model with Network Components
- Evolving Network Model
- ~~Compressible Graph Model~~

Announcements

- Midsem Exam: Sep 6, 1:30pm-3pm
  - No cheating
  - Allowed 1 A4 size cheat sheet (both sides)
  - Same format as Assignment 2
  - Covers content covered up to Aug 22 (last lecture)
  - Make assumptions if question is unclear
  - All the best!
- Project guidelines will be up soon.

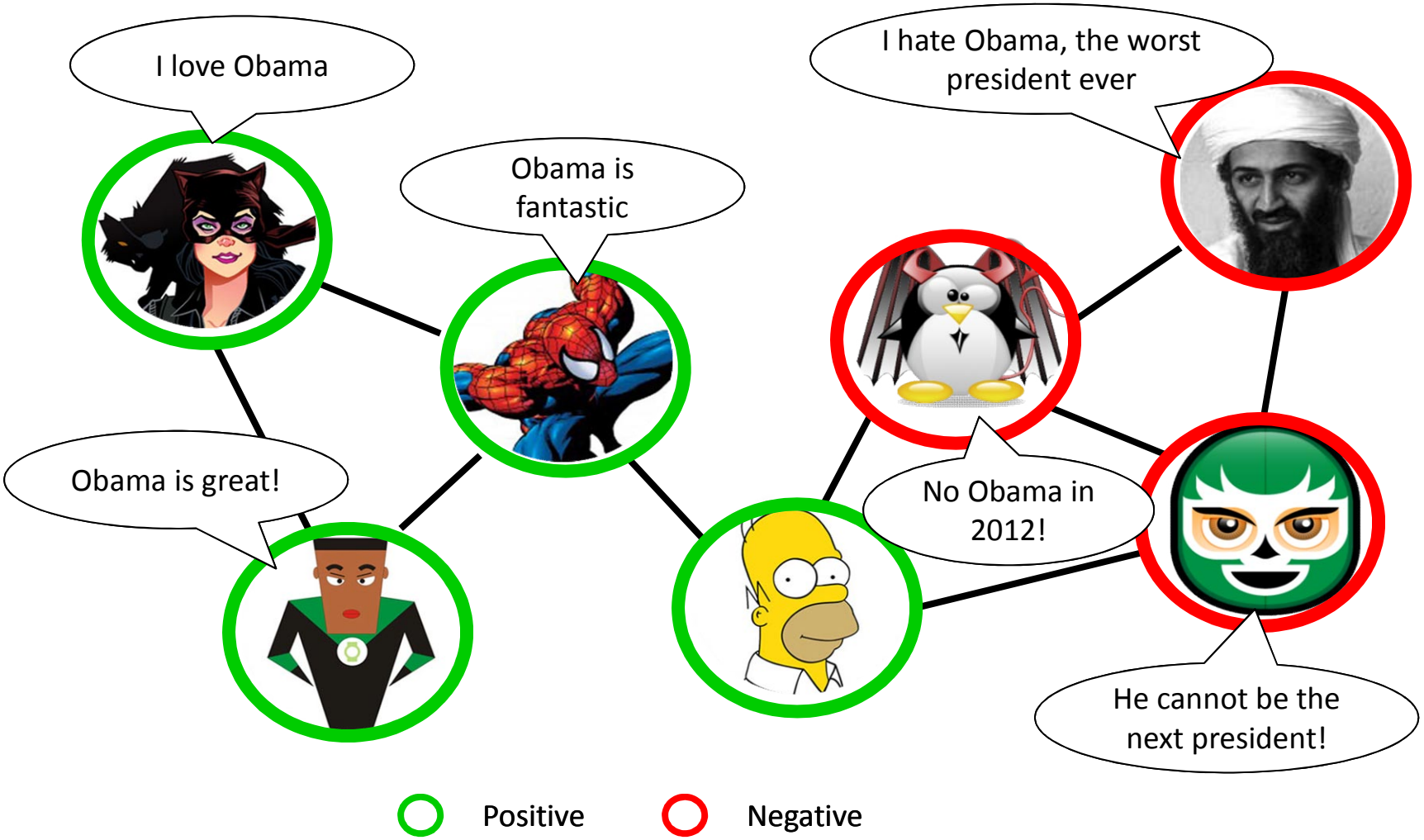
Today's Agenda

- Introduction to Social Influence Analysis
  - Definition of Social Influence
  - Does Social Influence really matter?
  - Homophily
  - Influence and Selection
  - Types of Social Influence
- Existential Tests for Social Influence
  - Randomization Test
  - Shuffle Test
  - Reverse Test
- Measuring Social Influence Analysis
  - Reachability-based methods
  - Structure Similarity
  - Structure + Content Similarity
  - Action-based methods

Today's Agenda

- **Introduction to Social Influence Analysis**
  - **Definition of Social Influence**
  - **Does Social Influence really matter?**
  - **Homophily**
  - **Influence and Selection**
  - **Types of Social Influence**
- Existential Tests for Social Influence
  - Randomization Test
  - Shuffle Test
  - Reverse Test
- Measuring Social Influence Analysis
  - Reachability-based methods
  - Structure Similarity
  - Structure + Content Similarity
  - Action-based methods

“Love Obama” or “Hate Obama”?



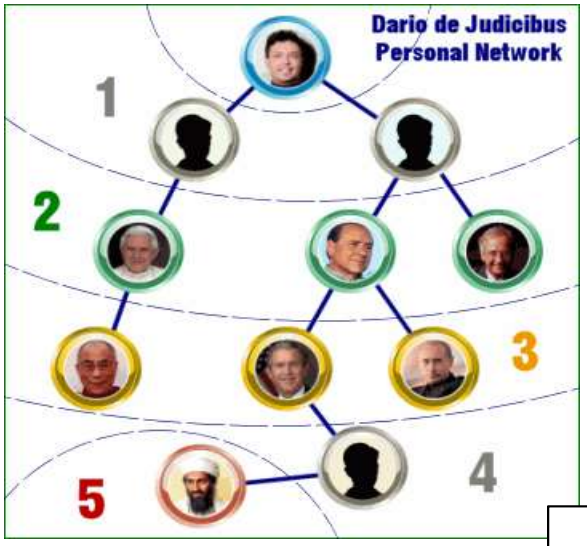
What is Social Influence?

- Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally.[1]
  - **Informational social influence**: to accept information from another;
  - **Normative social influence**: the influence of other people that leads us to conform in order to be liked and accepted by them.

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

Three Degree of Influence

Six degree of separation[1]



Three degree of Influence[2]

You can **influence** >1,000,000 persons in the world, according to the Dunbar's number[3].
Humans can comfortably maintain 150 stable relationships

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

[2] 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

[3] R. Dunbar. Neocortex size as a constraint on group size in primates. Human Evolution, 1992, 20: 469–493.

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.

[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.

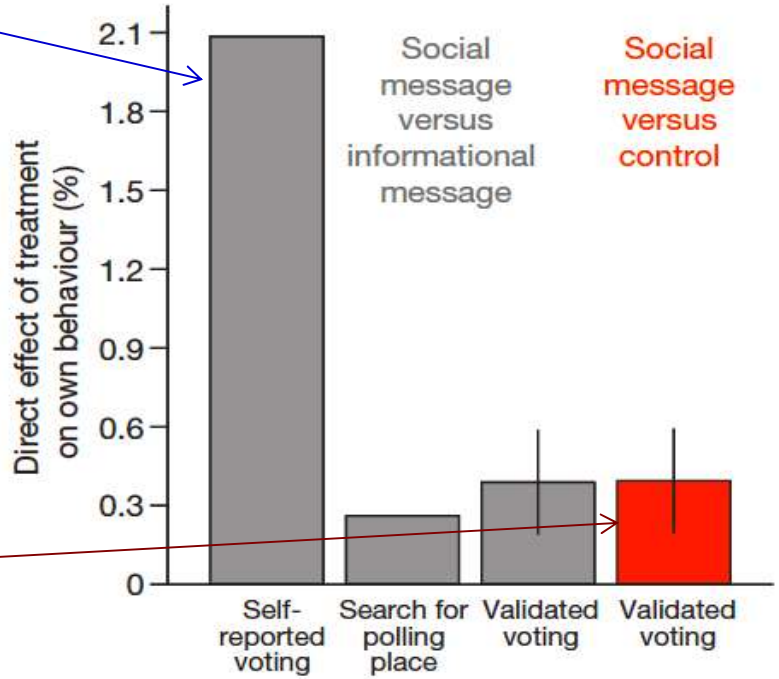
Case 1: Social Influence and Political Mobilization

Social msg group v.s.
Info msg group

Result: The former were 2.08% (*t*-test, *P*<0.01) more likely to click on the "I Voted" button

Social msg group v.s.
Control group

Result: The former were 0.39% (*t*-test, *P*=0.02) more likely to **actually vote** (via examination of public voting records)



[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.

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

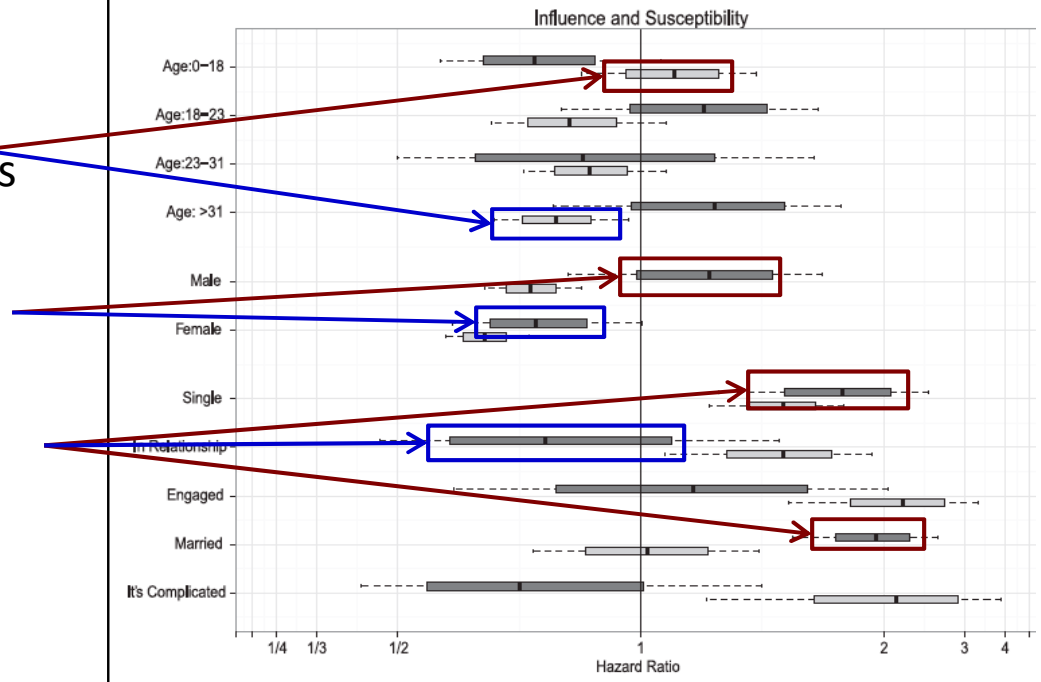
[2] Why I Deleted My Klout Profile, by Pam Moore, at Social Media Today, originally published November 19, 2011; retrieved November 26 2011

Case 3: Influential vs. Susceptible[1]

- Study of product adoption for 1.3M FB users

Results:

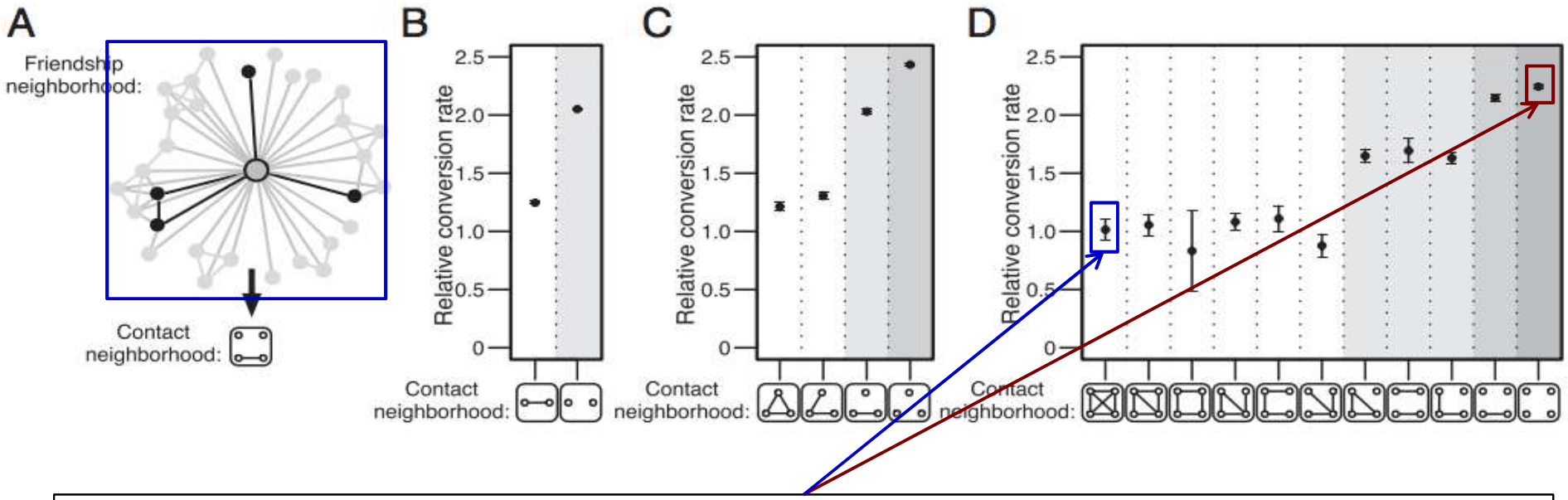
- Younger users are more (18%, $P<0.05$) susceptible to influence than older users
- Men are more (49%, $P<0.05$) influential than women
- Single and Married individuals are significantly more (>100%, $P<0.05$) influential than those who are in a relationship
- Married individuals are the least susceptible to influence



[1] S. Aral and D Walker. Identifying Influential and Susceptible Members of Social Networks. Science, 337:337-341, 2012.

Case 4: Who Influenced you and How?

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

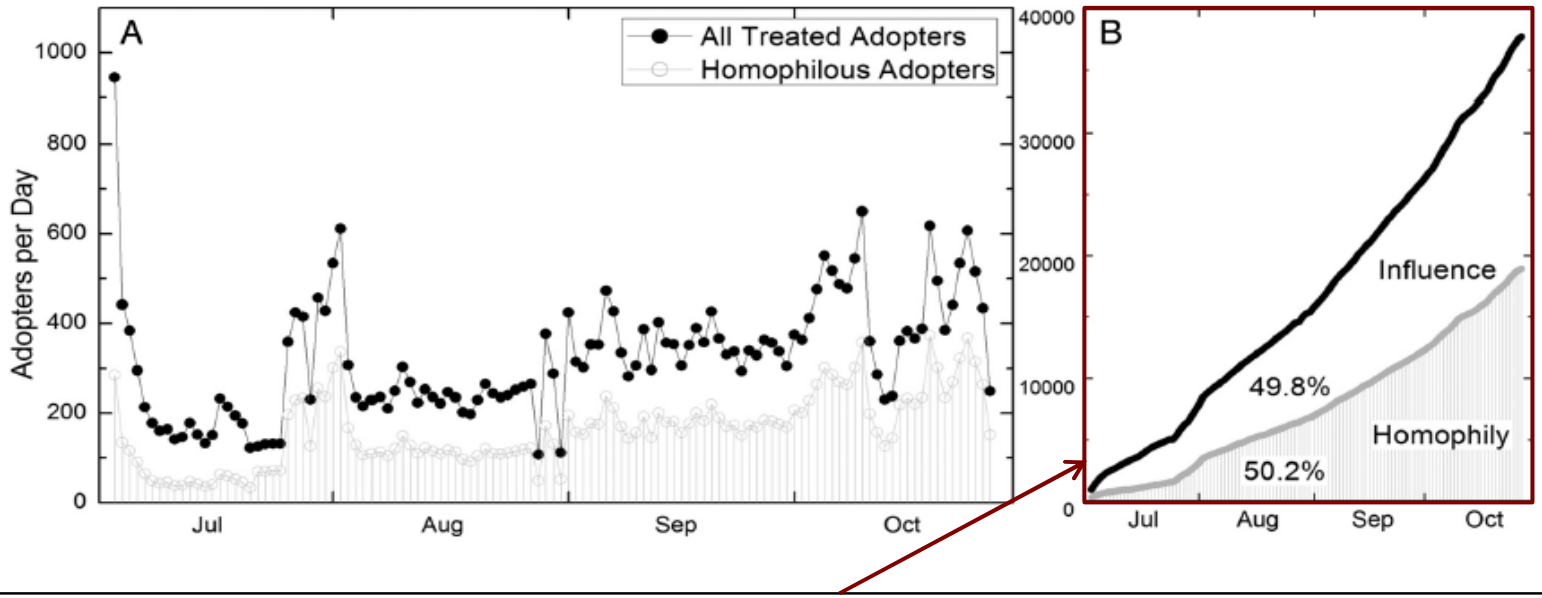


Results: Your behavior is influenced by the "structural diversity" (the number of connected components in your ego network) instead of the number of your friends.

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

Case 5: Influence and Correlation

- "Break" the myth of social influence

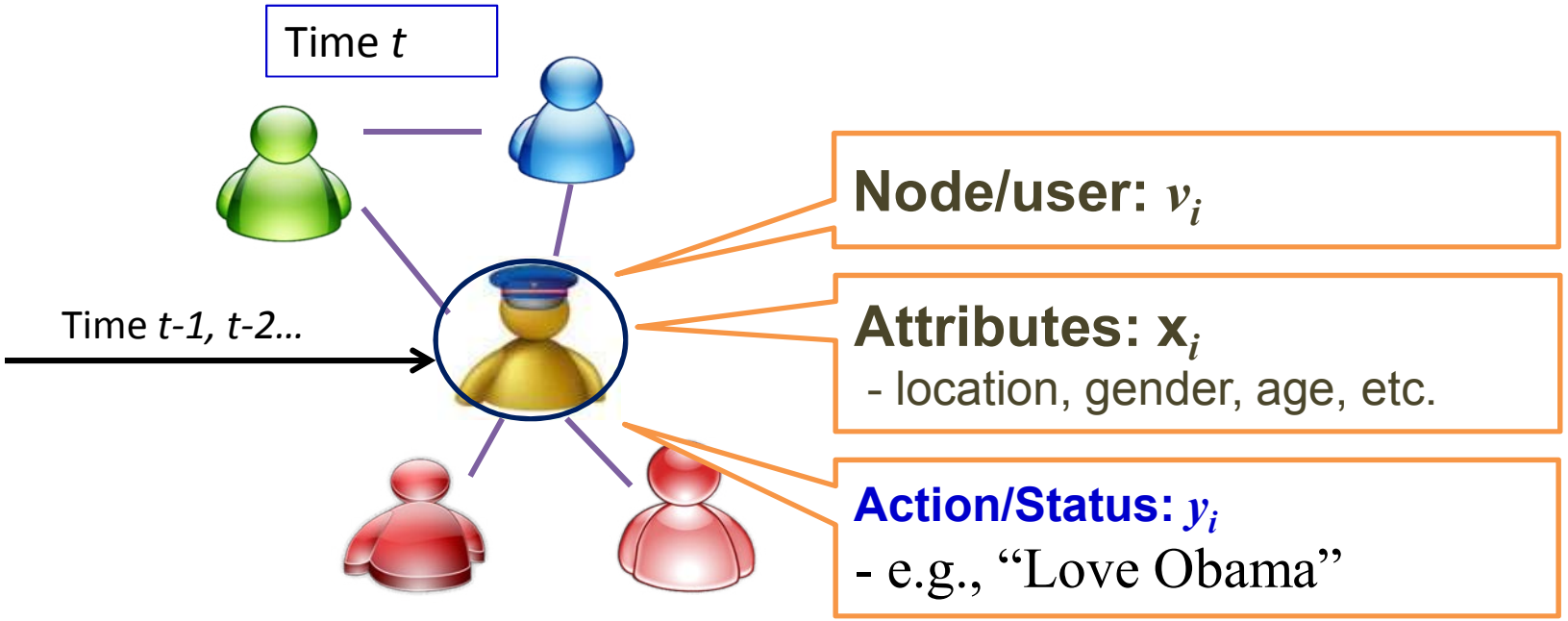


Results:

- Homophily explains >50% of the perceived behavioral contagion
- Previous methods overestimate peer influence by 300-700%

[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.

Notations



$$G = (V, E, X, Y)$$

G^t — the superscript t represents the time stamp

$e_{ij}^t \in E^t$ — represents a link/relationship from v_i to v_j at time t

Homophily

- Homophily
  - A user in the social network tends to be similar to their connected neighbors.
- Originated from different mechanisms
  - Social influence
    - Indicates people tend to follow the behaviors of their friends
  - Selection
    - Indicates people tend to create relationships with other people who are already similar to them
  - Confounding (hidden) variables
    - Other unknown variables exist, which may cause friends to behave similarly with one another.

Influence and Selection[1]

$$Selection = \frac{p(e_{ij}^{t} = 1 \mid e_{ij}^{t-1} = 0, \langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \rangle > \varepsilon)}{p(e_{ij}^{t} = 1 \mid e_{ij}^{t-1} = 0)}$$

Similarity between user *i* and *j* at time *t*-1 is larger than a threshold

There is a link between user *i* and *j* at time *t*

- Denominator: the conditional probability that an unlinked pair will become linked
- Numerator: the same probability for unlinked pairs whose similarity exceeds the threshold

$$Influence = \frac{p(\langle \mathbf{x}_i^{t}, \mathbf{x}_j^{t} \rangle > \langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \rangle \mid e_{ij}^{t} = 1, e_{ij}^{t-1} = 0)}{p(\langle \mathbf{x}_i^{t}, \mathbf{x}_j^{t} \rangle > \langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \rangle \mid e_{ij}^{t-1} = 0)}$$

- Denominator: the probability that the similarity increase from time *t*-1 to time *t* between two nodes that were not linked at time *t*-1
- Numerator: the same probability that became linked at time *t*

- A Model is learned through matrix factorization/factor graph

[1] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. In KDD'09, pages 747–756, 2009.

Other Related Concepts: Hazard Ratio

- **Hazard Ratio**
  - Chance of an event occurring in the treatment group divided by its chance in the control group
  - Example:

$$\frac{\text{Chance of users to buy iPhone with >=1 iPhone user friend(s)}}{\text{Chance of users to buy iPhone without any iPhone user friend}}$$

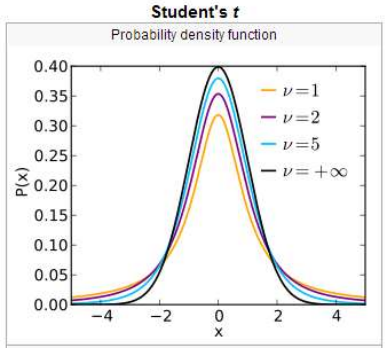
  - Measuring instantaneous chance by *hazard rate h(t)*

$$h(t) = \lim_{\Delta t \to 0} \frac{\text{observed events in interval}[t, t+\Delta t]/N(t)}{\Delta t}$$

    - N(t) is the number at risk at the beginning of the interval
  - The hazard ratio is the relationship between the instantaneous hazards in two groups
  - Proportional hazards models (e.g. Cox-model) could be used to report hazard ratio.

Other Related Concepts: *t*-test

- A *t*-test is usually used when the test statistic follows a Student's *t* distribution if the null hypothesis is supported.
- To test if the difference between sets of data is significant
- Welch's *t*-test
  - Calculate *t*-value



sample mean

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_{\overline{x}_1 - \overline{x}_2}}, \quad s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

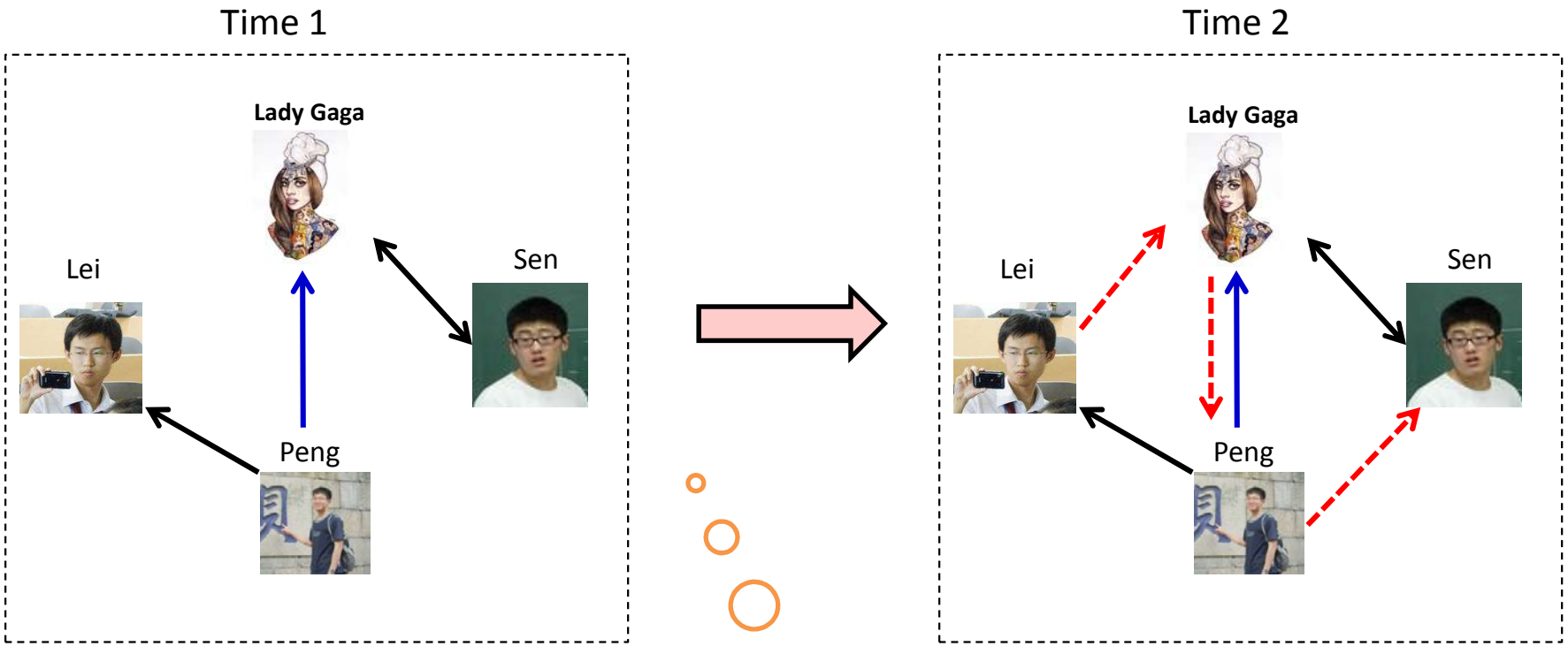
Unbiased estimator of sample variance

#participants in the control group

#participants in the treatment group

  - Find the *p*-value using a table of values from Student's t-distribution
  - If the *p*-value is below chosen threshold (e.g. 0.01) then the two variables are viewed as significant different.

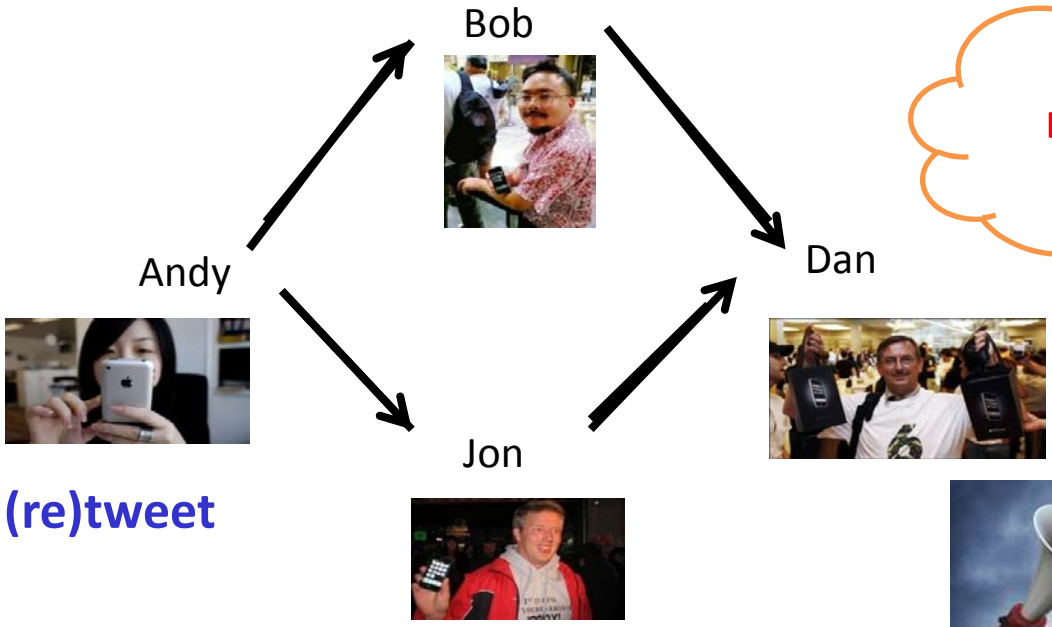
Types of Social Influence: Follow Influence on Twitter



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

新浪微博 weibo.com

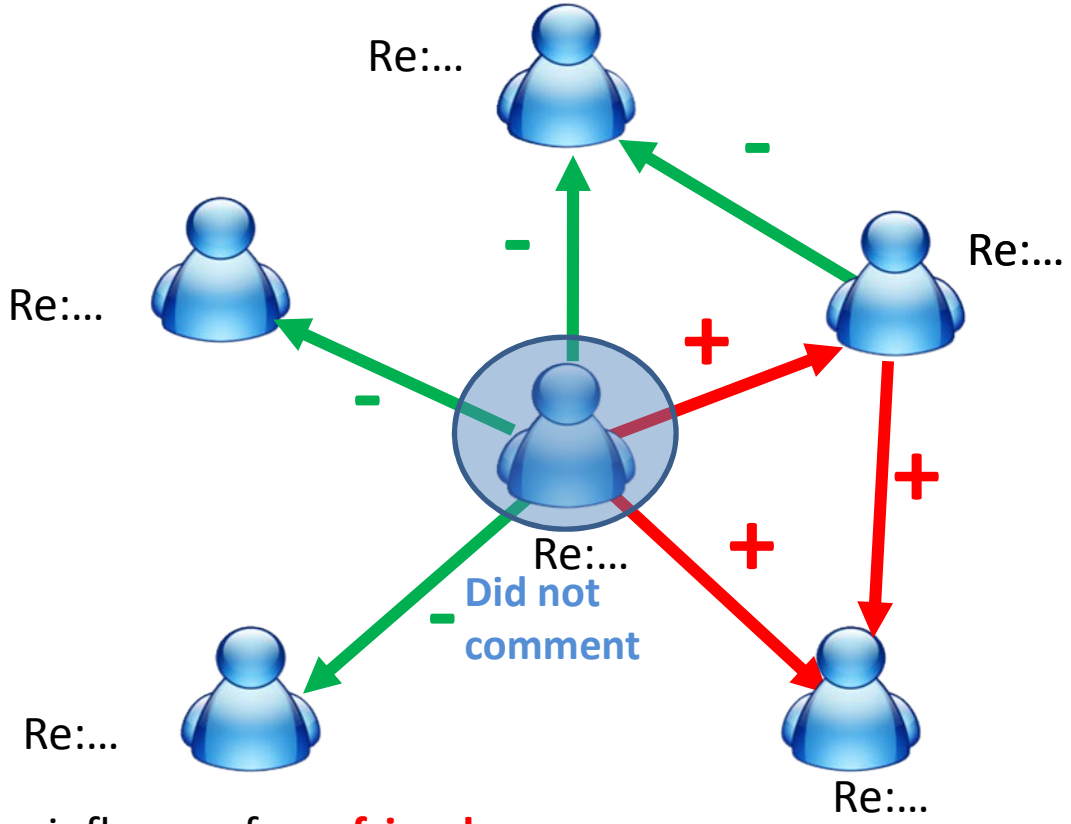


Who will retweet it next?

When you (re)tweet something

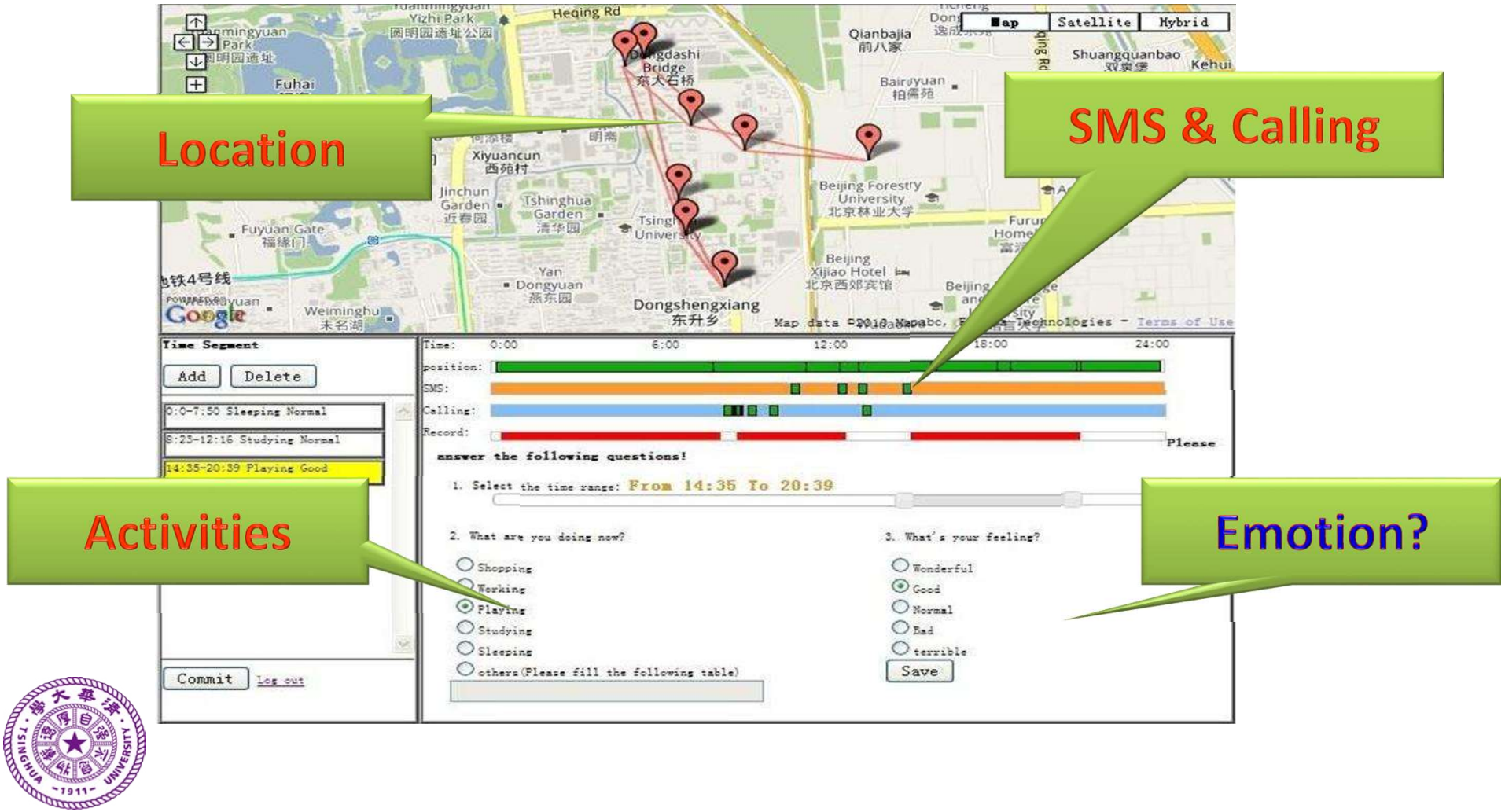
Types of Social Influence: Comment Influence

News: Alan Cox Exits Intel.
Governments Want Private Data

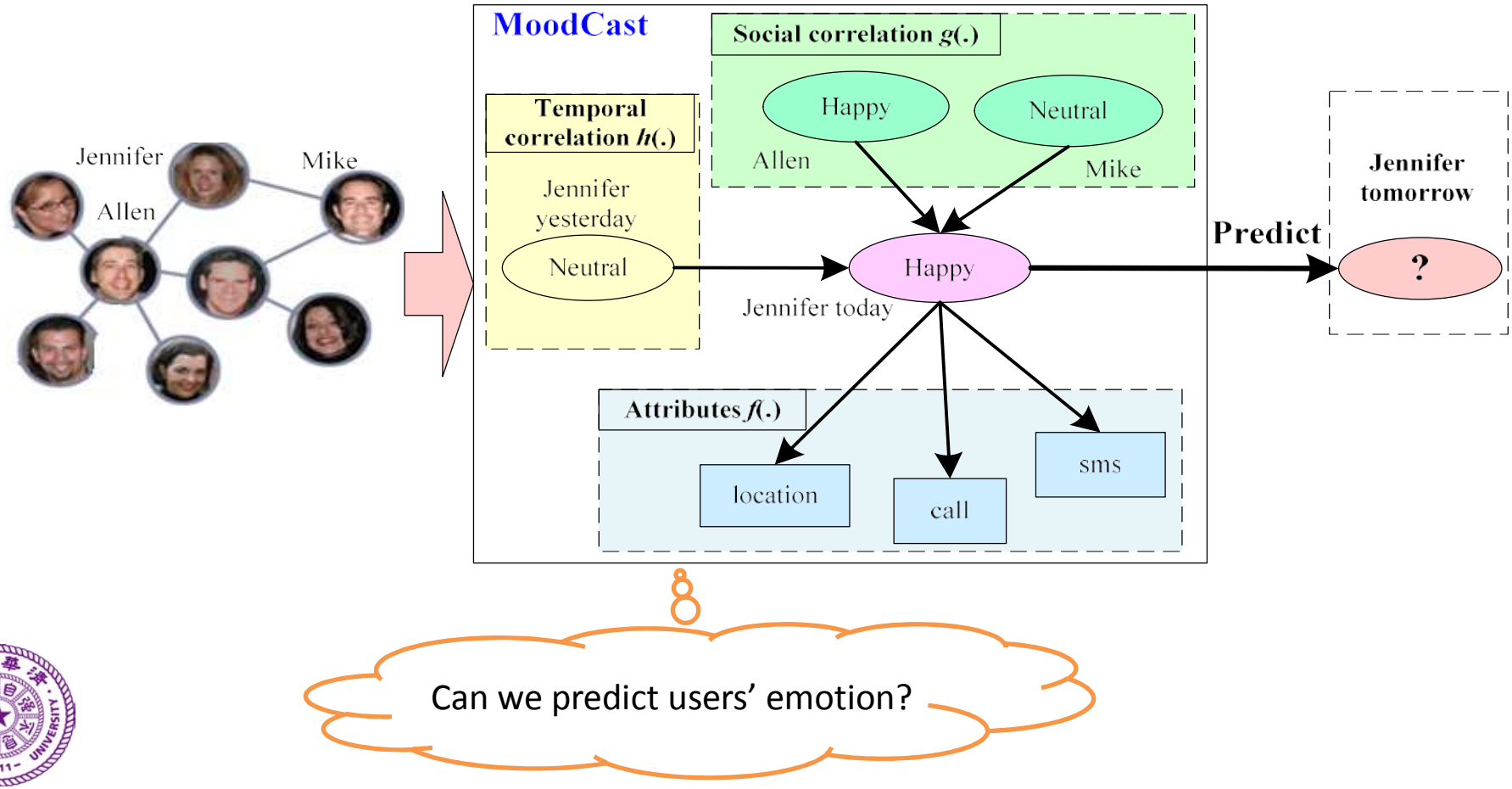


positive influence from **friends**
negative influence from **foes**

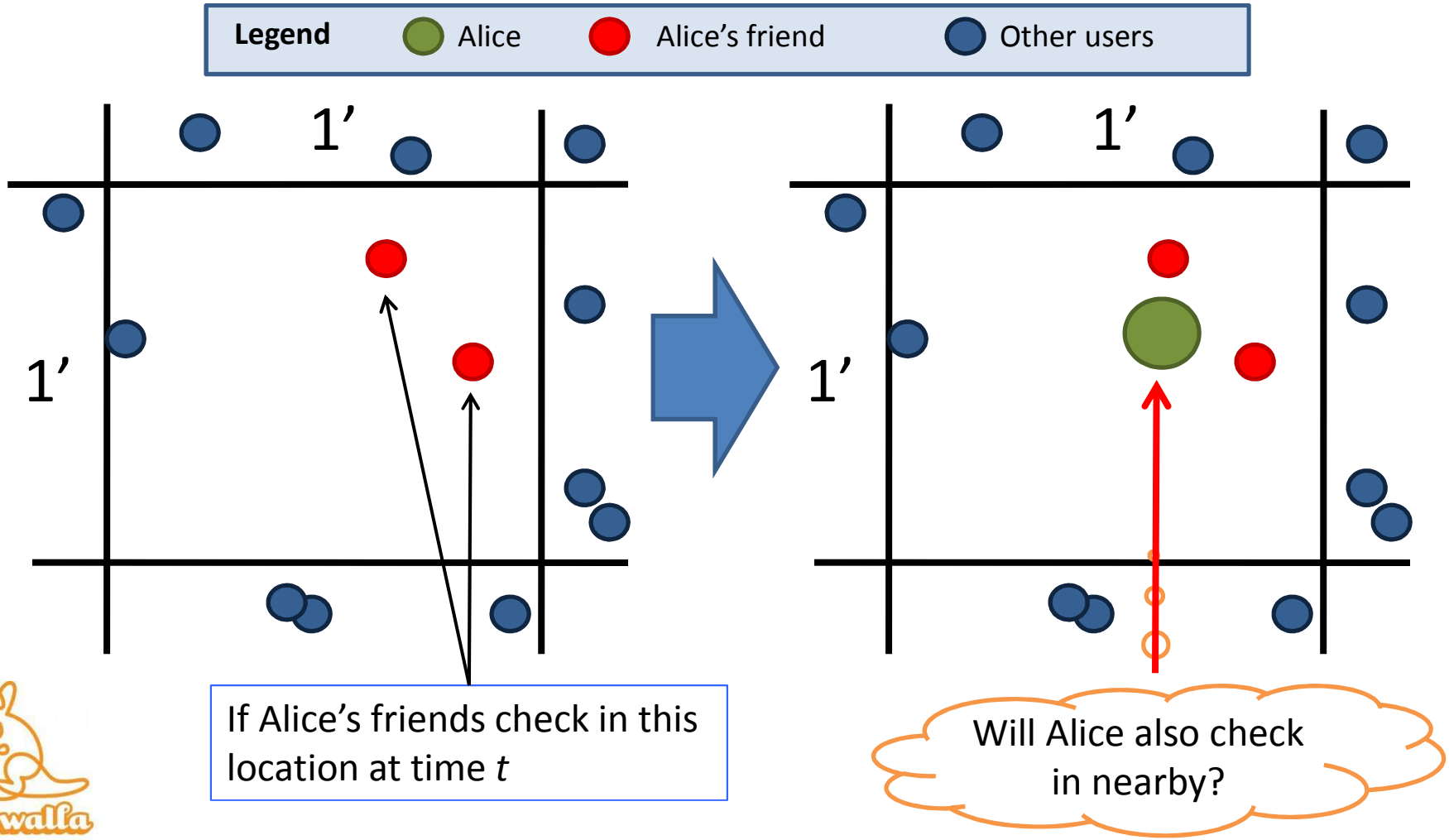
Types of Social Influence: Emotion Influence



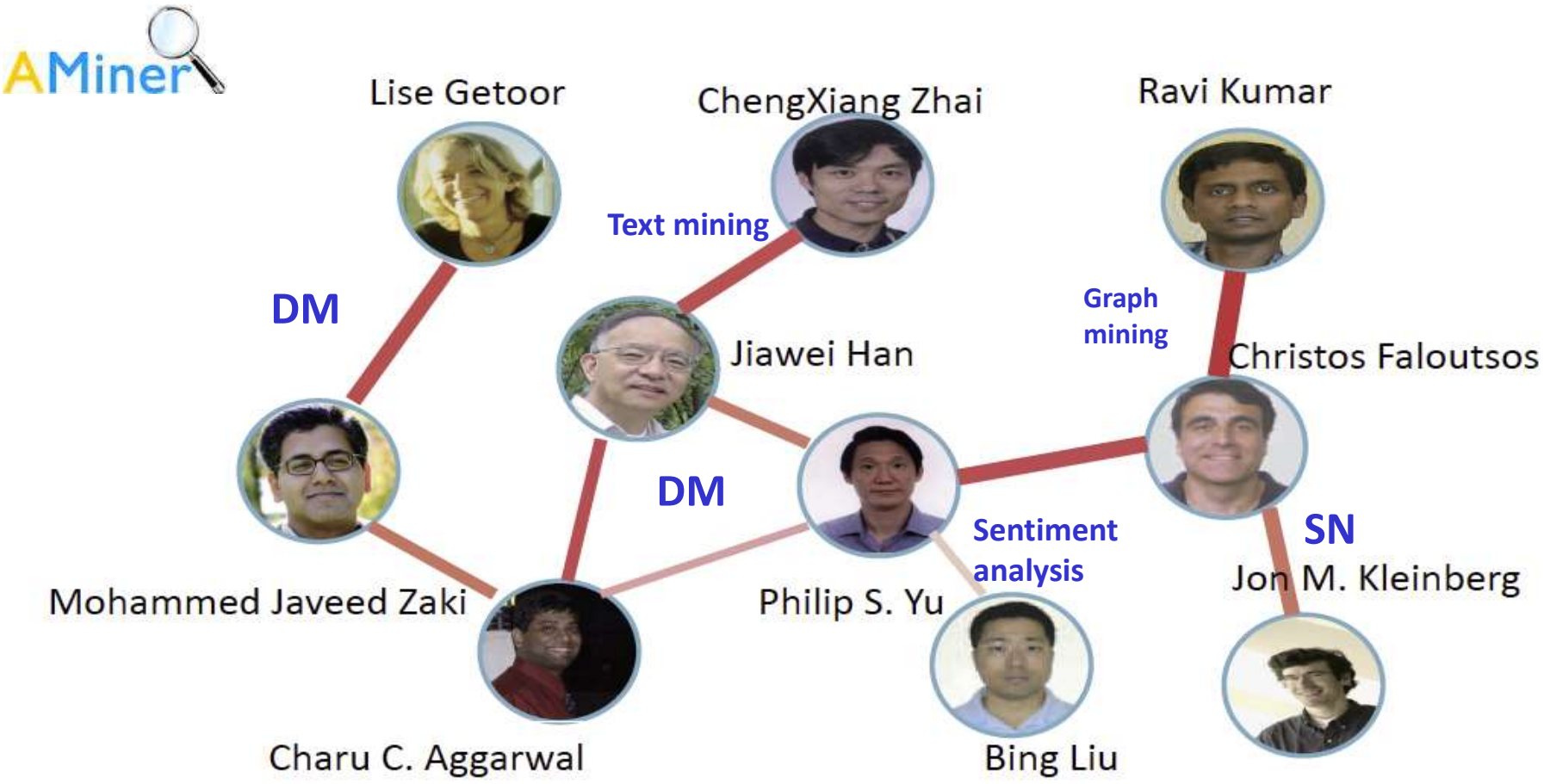
Types of Social Influence: Emotion Influence (cont.)



Types of Social Influence: Check-in Influence in Gowalla



Types of Social Influence: Correlation & Influence in Academia



Today's Agenda

- Introduction to Social Influence Analysis
  - Definition of Social Influence
  - Does Social Influence really matter?
  - Homophily
  - Influence and Selection
  - Types of Social Influence
- **Existential Tests for Social Influence**
  - **Randomization Test**
  - **Shuffle Test**
  - **Reverse Test**
- Measuring Social Influence Analysis
  - Reachability-based methods
  - Structure Similarity
  - Structure + Content Similarity
  - Action-based methods

Randomization

- Theoretical fundamentals[1, 2]
  - In science, randomized experiments are the experiments that allow the greatest reliability and validity of statistical estimates of treatment effects.
- 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.

[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

RCT in Social Network

- We use RCT to test the influence and its significance in SN.

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

Example: Political mobilization

- There are two kinds of treatments.

A controlled trial

Treatment 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.

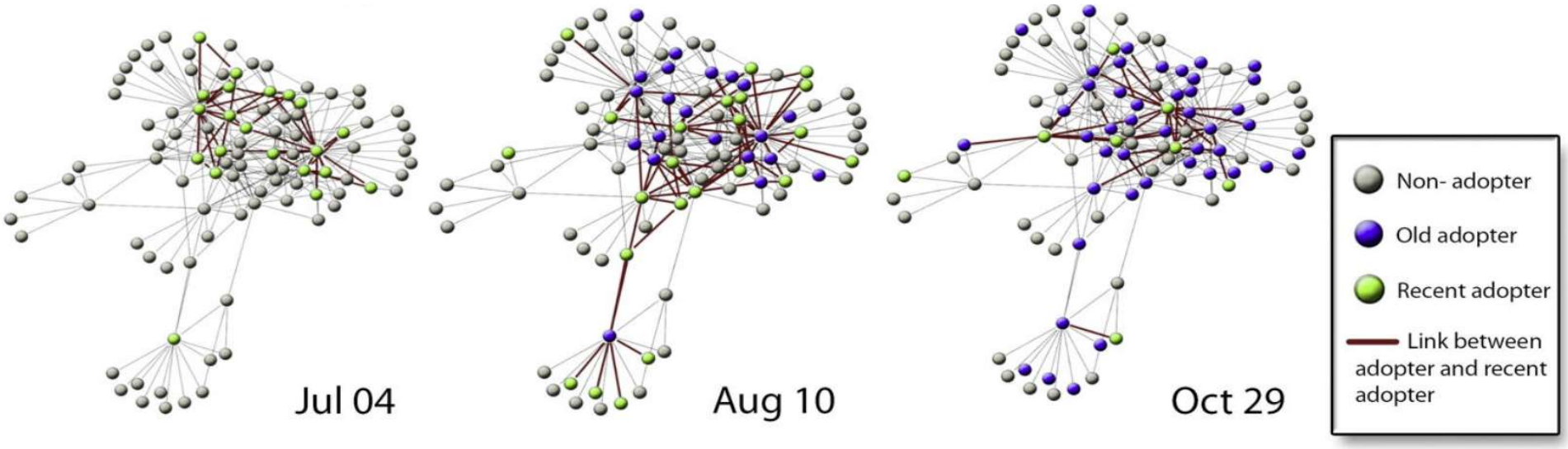
Treatment for Group 1

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.



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*.

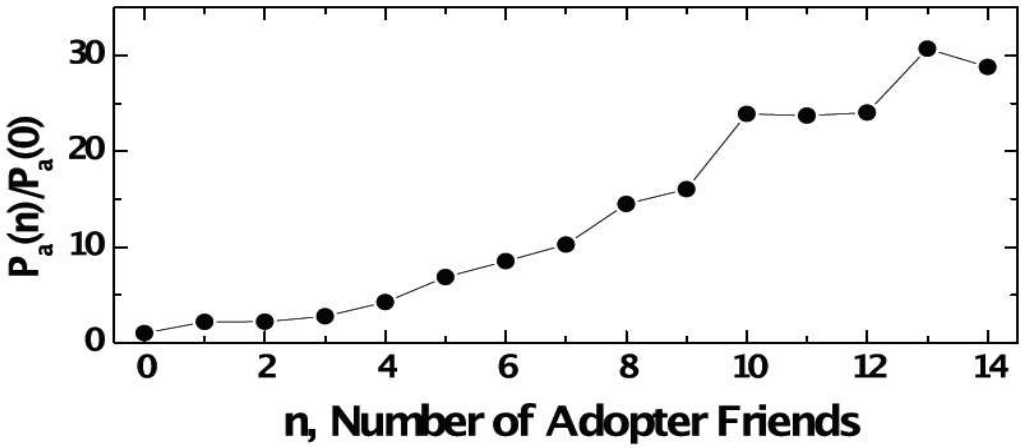
[1] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. PNAS, 106 (51):21544-21549, 2009.

Adoption Diffusion of Y! Go (2)

- Yahoo! Go
  - 27.4 M users, 14 B page views, 3.9 B messages
- The RCT
  - Control seeds: random sample of 2% of the entire network (3.2M nodes)
  - Experimental seeds: all adopters of Yahoo! Go from 6/1/2007 to 10/31/2007 (0.5M nodes)

[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.

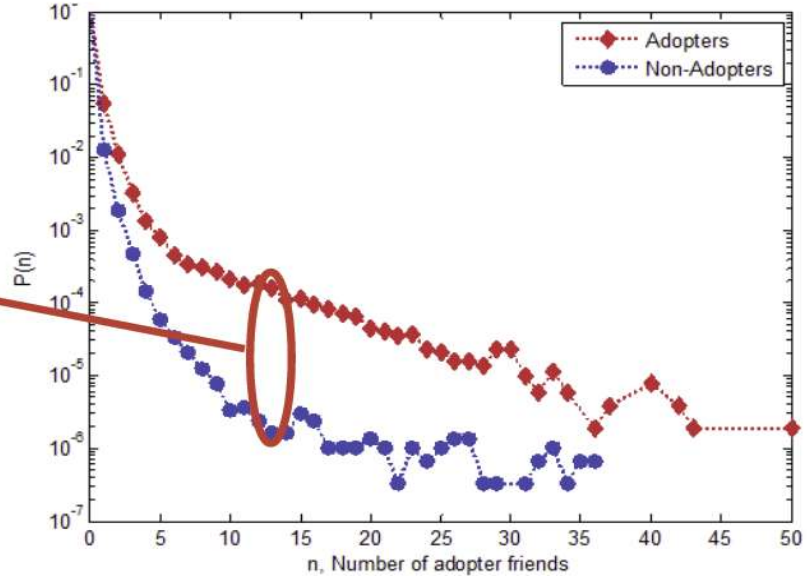
Evidence of Influence?



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.

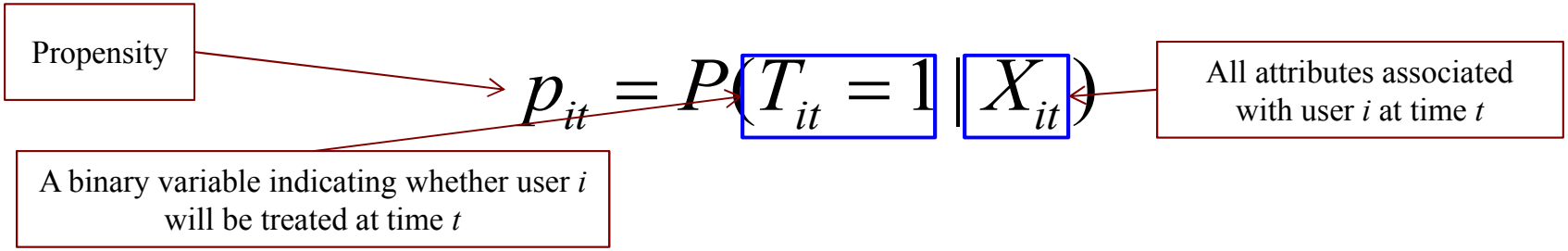
Adopters are 100 times more likely to have 12 adopter friends than non-adopters

Is the setting fair?



Matched Sampling Estimation

- Selection Bias of existing randomized methods
  - Adopters are more likely to have adopter friends than non-adopters (homophily)
- Matched sampling estimation
  - Match the treated (adopter neighbors) observations with untreated who are as likely to have been treated, conditional on a vector of observable characteristics, but who were not treated

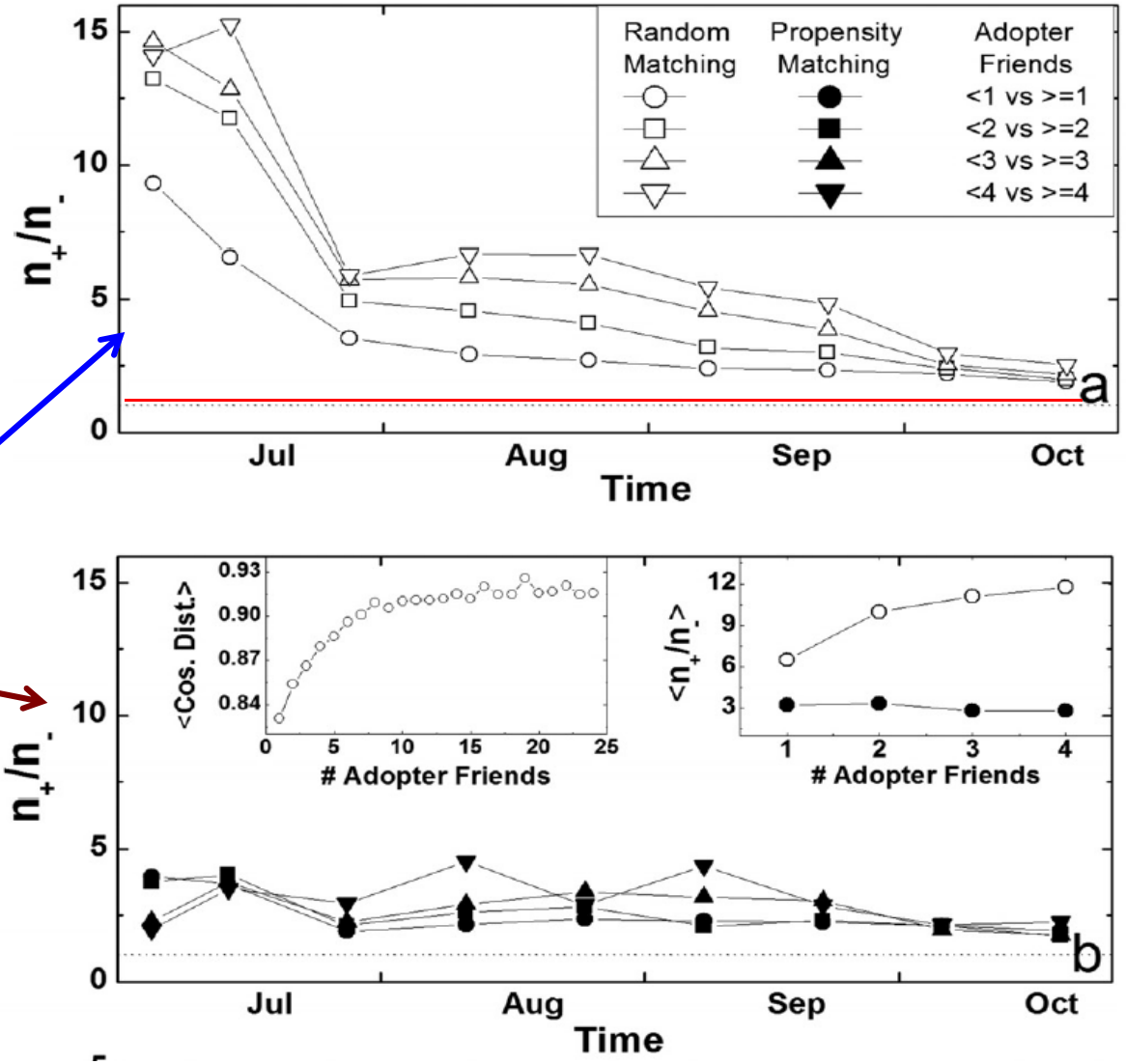


The new RCT:

- **Treatment group:** a user i who have k friends who have adopted the Y! Go at time t;
- **Controlled group:** a matched user j who do not have k friends adopt Y! Go at time t, but is very likely to have k friends to adopt Y!Go at time t, i.e., $|p_{it} - p_{jt}| < \sigma$

Results—Random sampling and Matched sampling

The fraction of observed treated to untreated adopters (n_+/n_-) under:
(a) Random sampling;
(b) Matched sampling.



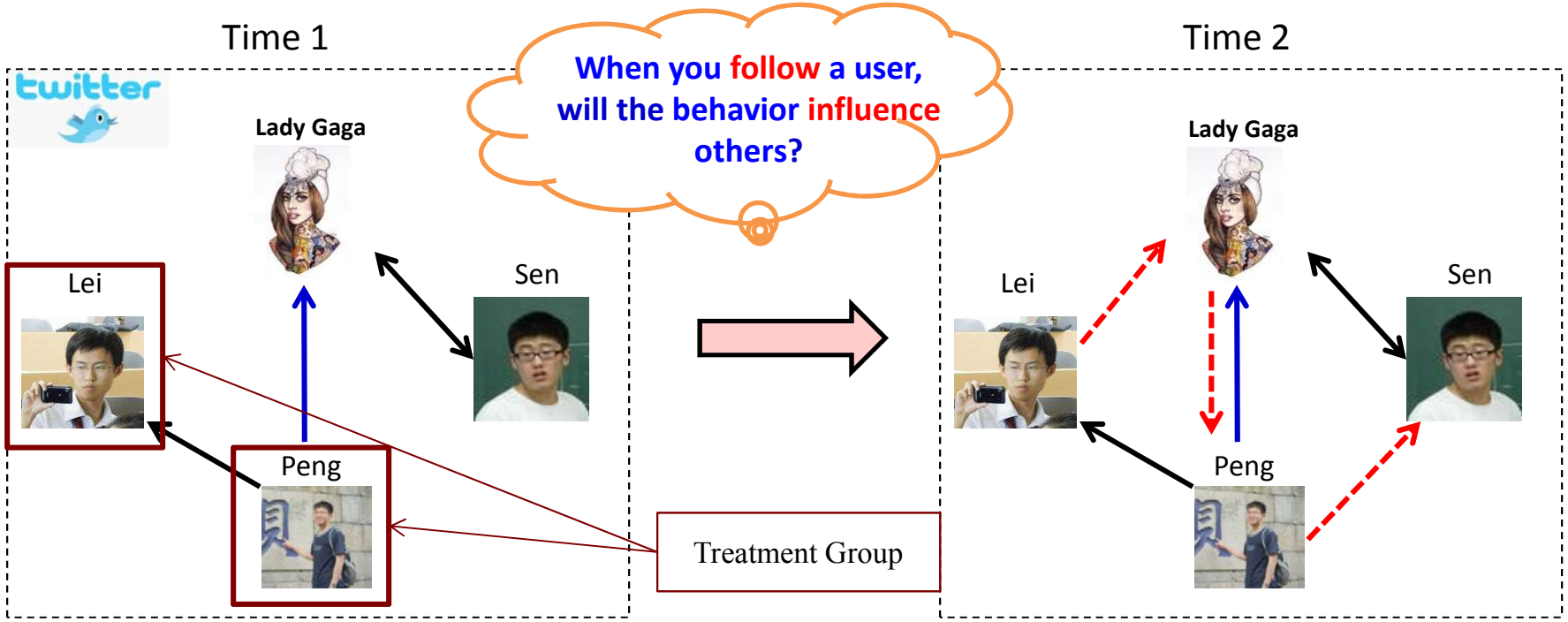
We speculate that exaggerated homophily among early adopters leads to greater upward bias in random matching influence estimates in earlier periods.

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.

Example: Following Influence Test



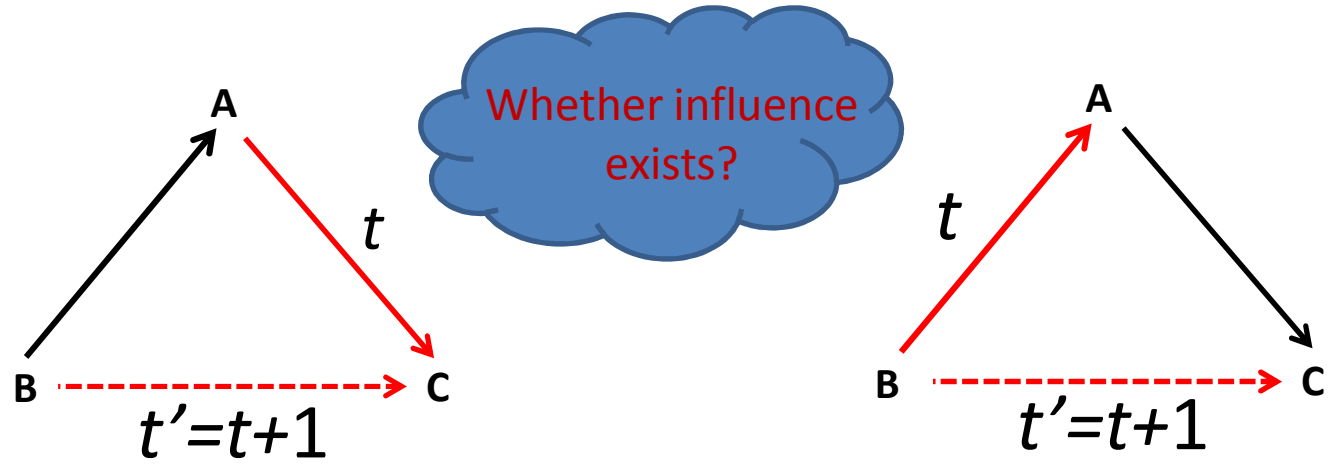
RCT:

- **Treatment group:** people who followed some other people or who have friends following others at time *t*;
- **Controlled group:** people who did not follow anyone and do not have any friends following others at time *t*.

[1] T. Lou, J. Tang, J. Hopcroft, Z. Fang, and X. Ding. Learning to Predict Reciprocity and Triadic Closure in Social Networks. ACM TKDD 2013.

Influence Test via Triad Formation

Two Categories of Following Influences



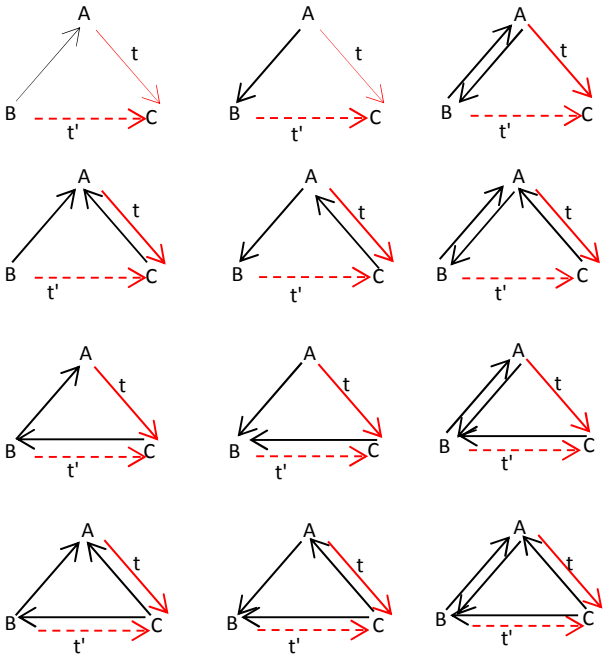
Follower diffusion

Followee diffusion

–>: pre-existed relationships
–>: a new relationship added at *t*
-->: a possible relationship added at *t*+1

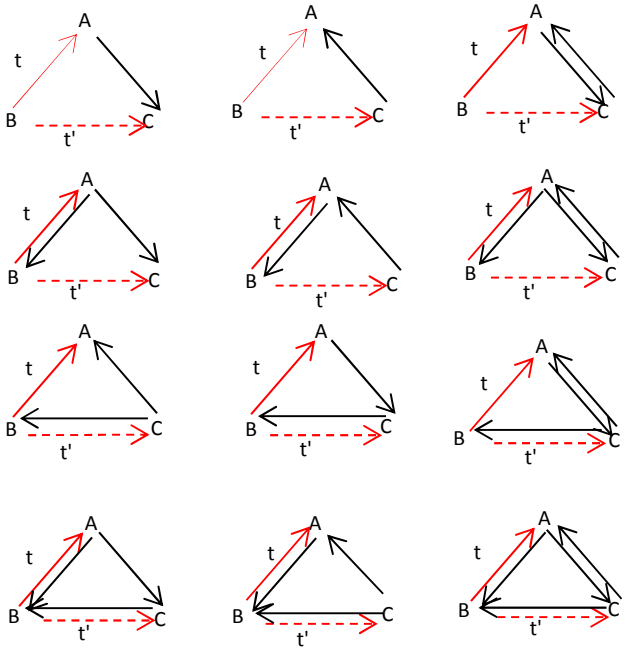
24 Triads in Following Influence

Follower diffusion



12 triads

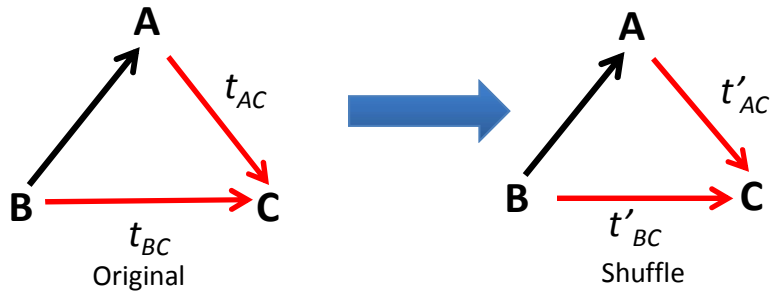
Followee diffusion



12 triads

Timing Shuffle Test

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



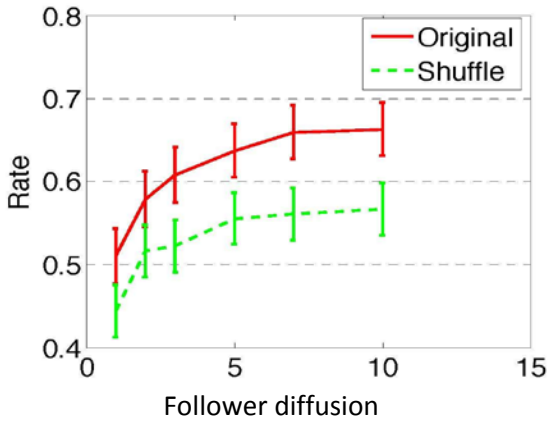
Shuffle test

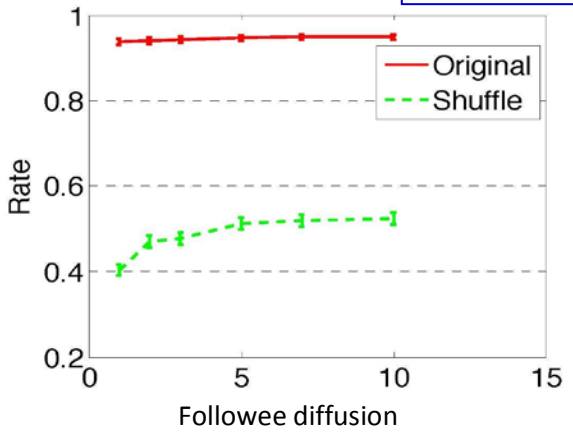
$$Rate = \frac{\#Triad \mid 0 < t_{BC} - t_{AC} < \delta}{\#Triad \mid t_{BC} \text{ and } t_{AC} \text{ exist}}$$

- Compare the rate under the original and shuffled dataset.

t-test, $P<0.01$

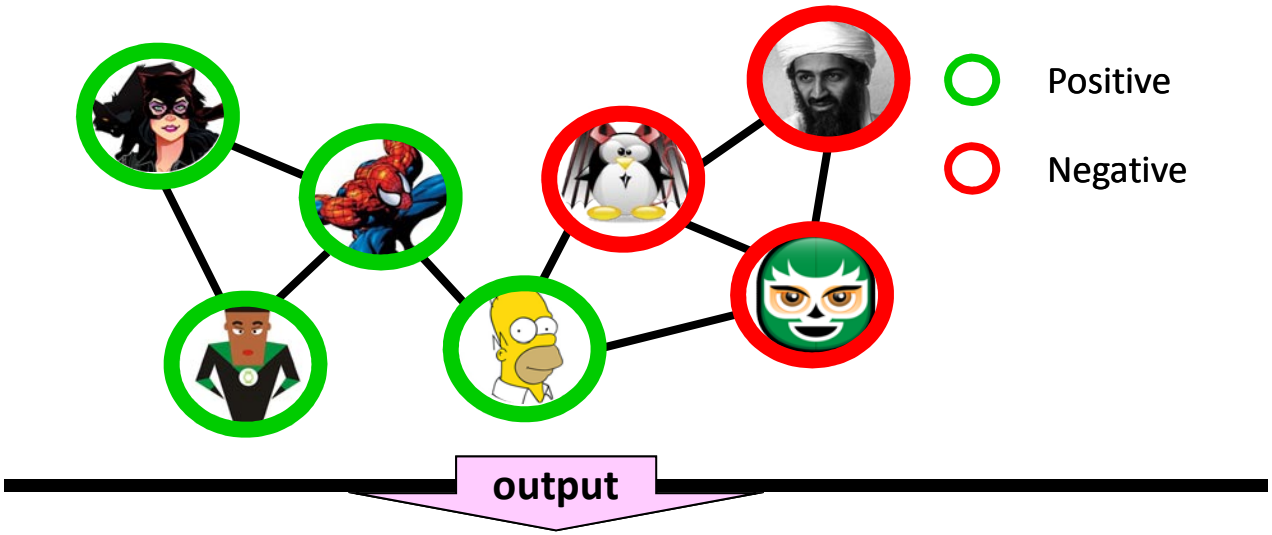
- Result





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

Output of Influence Test



There indeed exists influence!

Today's Agenda

- Introduction to Social Influence Analysis
  - Definition of Social Influence
  - Does Social Influence really matter?
  - Homophily
  - Influence and Selection
  - Types of Social Influence
- Existential Tests for Social Influence
  - Randomization Test
  - Shuffle Test
  - Reverse Test
- **Measuring Social Influence Analysis**
  - **Reachability-based methods**
  - **Structure Similarity**
  - **Structure + Content Similarity**
  - **Action-based methods**

Reachability-based Method[1]

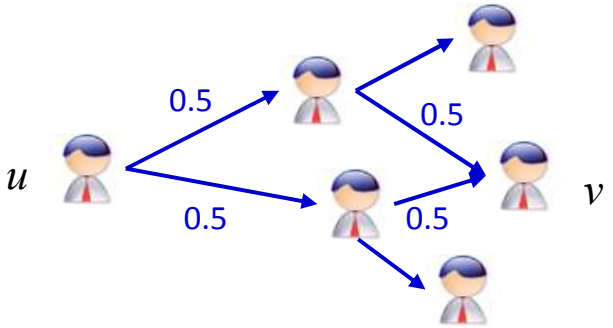
- PageRank, Random Walk with Restart
- Influence of a path

$$\inf(p) = \prod_{v_i \in p} \frac{1}{\text{outdeg}(v_i)}$$

- Influence of user u on v

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

All paths from u to v within path length t



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

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.

Structure Similarity: SimRank

- SimRank is a general similarity measure, based on a simple and intuitive graph-theoretic model (Jeh and Widom, KDD'02).

C is a constant between 0 and 1, e.g., C=0.8

$$sim(u,v) = \frac{C}{|I(u)||I(v)|}\sum_{i=1}^{|I(u)|}\sum_{j=1}^{|I(v)|} sim(I_i(u), I_j(v))$$

$$\text{Initialization}: sim(u,u) = 1, \text{ if } u = v;$$

$$sim(u,v) = 0, \text{if } u \neq v.$$

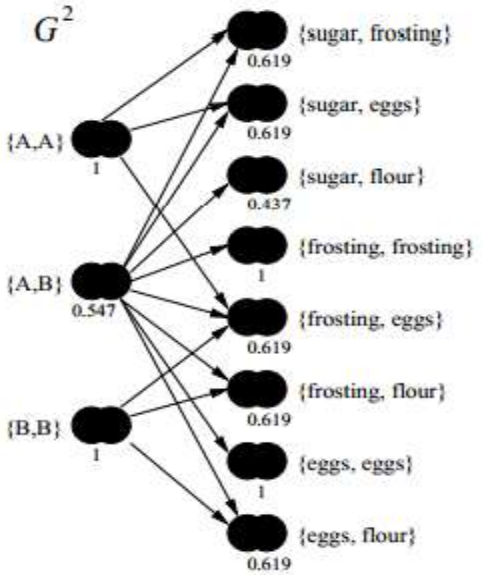
The set of pages which have links pointing to *u*

[1] G. Jeh and J. Widom, SimRank: a measure of structural-context similarity. In KDD, pages 538-543, 2002.

Structure Similarity: Bipartite SimRank

Extend the basic SimRank equation to bipartite domains consisting of two types of objects
{A, B} and {a, b}.

E.g.,
People *A* and *B* are similar if they purchase similar items.
Items *a* and *b* are similar if they are purchased by similar people.



$$sim(A,B) = \frac{C_1}{|O(A)||O(B)|} \sum_{i=1}^{|O(A)|} \sum_{j=1}^{|O(B)|} sim(O_i(A), O_j(B))$$

$$sim(a,b) = \frac{C_2}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} sim(I_i(a), I_j(b))$$

Structure Similarity: MiniMax Variation

In some cases, e.g., we care more about the maximal similarity of two neighbors.

$$sim_A(A,B) = \frac{C_1}{|O(A)|}\sum_{i=1}^{|O(A)|}\max_{j=1}^{|O(B)|} sim(O_i(A),O_j(B))$$

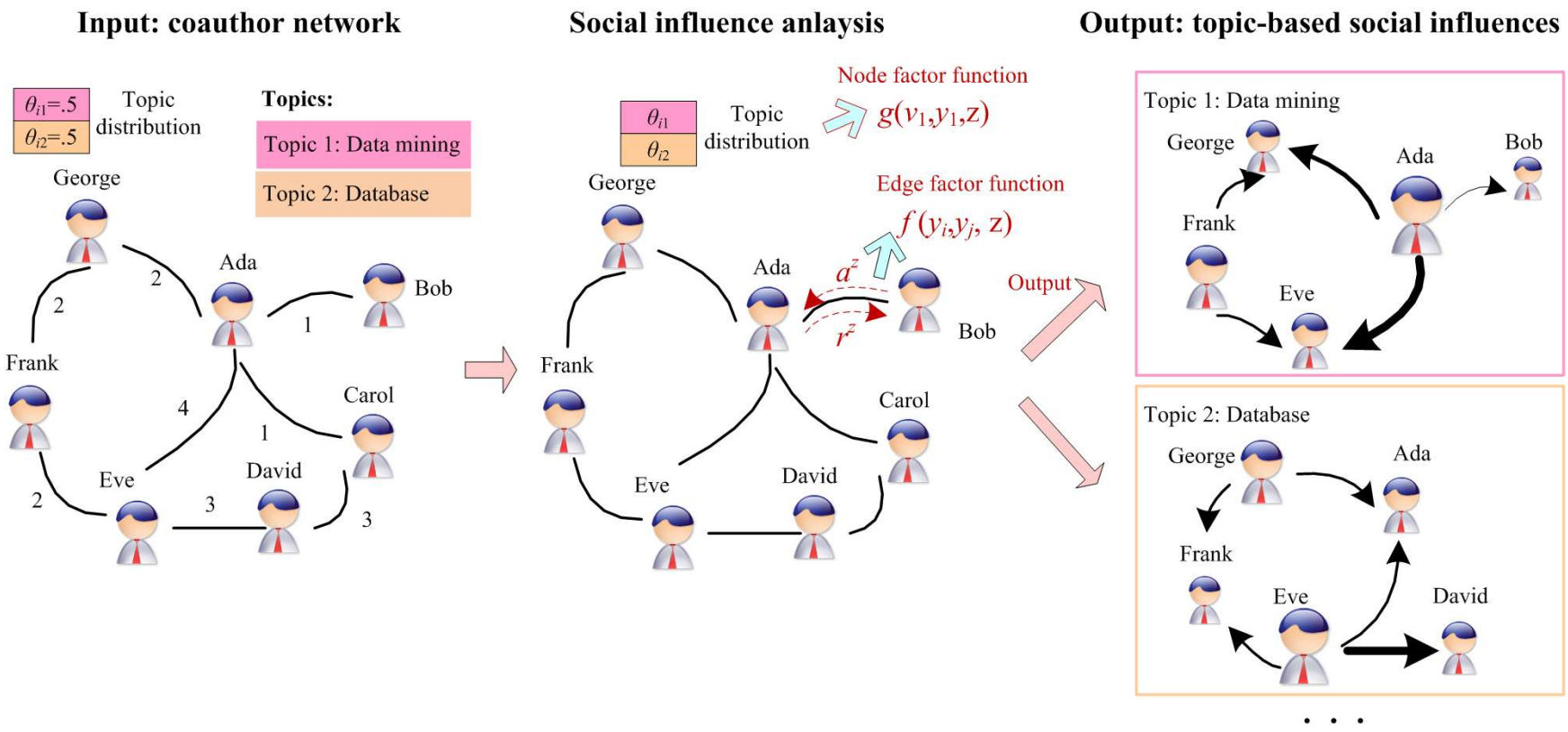
$$sim_B(A,B) = \frac{C_1}{|O(B)|}\sum_{j=1}^{|O(B)|}\max_{i=1}^{|O(A)|} sim(O_i(A),O_j(B))$$

$$sim(A,B) = \min(sim_A(A,B), sim_B(A,B))$$

Note: Again, the method only considers the network information.

Structure+Content Similarity: Topic-based Social Influence Analysis

- Social network -> Topical influence network



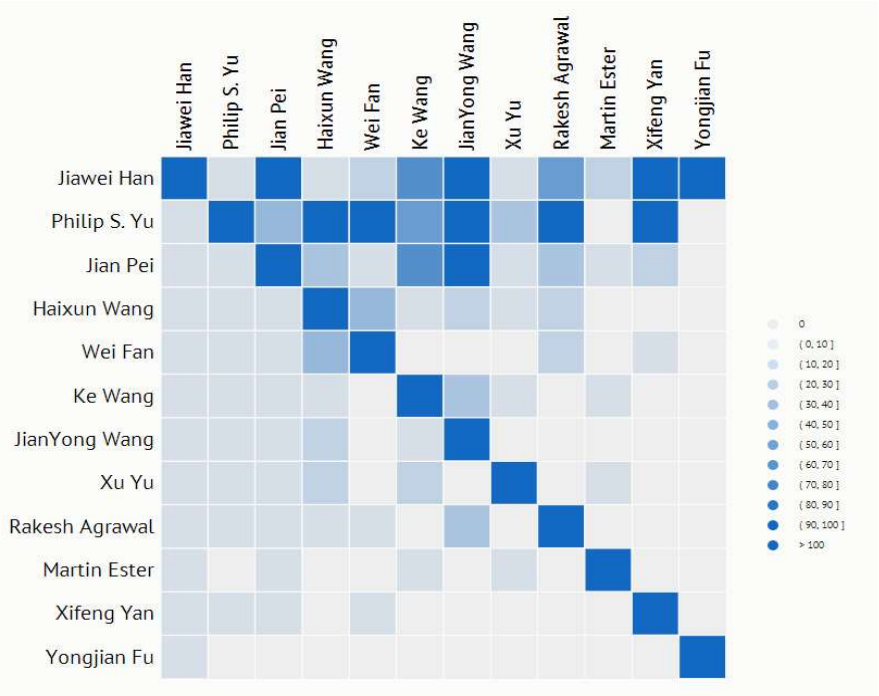
[1] J. Tang, J. Sun, C. Wang, and Z. Yang. Social Influence Analysis in Large-scale Networks. In KDD'09, pages 807-816, 2009.

Structure+Content Similarity: Social Influence Sub-graph on "Data mining"

Table 4: Dynamic influence analysis for Dr. Jian Pei during 2000-2009. Due to space limitation, we only list coauthors who most influence on/by Dr. Pei in each time window.

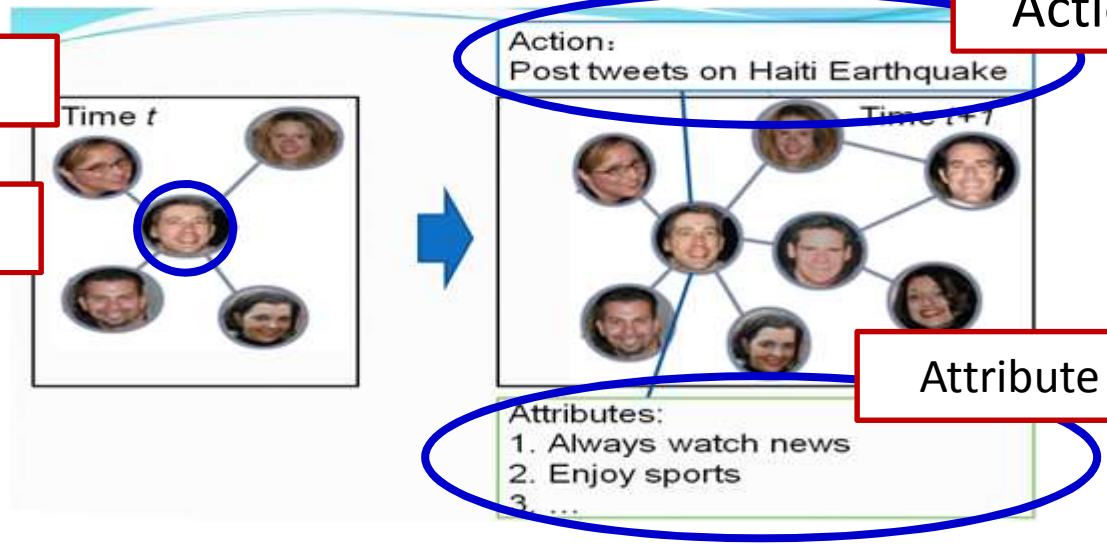
Year	Pairwise	Influence
2000 - 2001	Influence on Dr. Pei	Jiawei Han (0.4961)
	Influenced by Dr. Pei	Jiawei Han (0.0082)
2002 - 2003	Influence on Dr. Pei	Jiawei Han (0.4045), Ke Wang (0.0418), Jianyong Wang (0.019), Xifeng Yan (0.007), Shiwei Tang (0.0052)
	Influenced by Dr. Pei	Shiwei Tang (0.436), Hasan M.Jamil (0.4289), Xifeng Yan (0.2192), Jianyong Wang (0.1667), Ke Wang (0.0687)
2004 - 2005	Influence on Dr. Pei	Jiawei Han (0.2364), Ke Wang (0.0328), Wei Wang (0.0294), Jianyong Wang (0.0248), Philip S. Yu (0.0156)
	Influenced by Dr. Pei	Chun Tang (0.5929), Shiwei Tang (0.5426), Hasan M.Jamil (0.3318), Jianyong Wang (0.1609), Xifeng Yan (0.1458), Yan Huang (0.1054)
2006 - 2007	Influence on Dr. Pei	Jiawei Han (0.1201), Ke Wang (0.0351), Wei Wang (0.0226), Jianyong Wang (0.018), Ada Wai-Chee Fu (0.0125)
	Influenced by Jian Pei	Chun Tang (0.6095), Shiwei Tang (0.6067), Byung-Won On (0.4599), Hasan M.Jamil (0.3433), Jaewoo Kang (0.3386)
2008 - 2009	Influence on Dr. Pei	Jiawei Han (0.2202), Ke Wang (0.0234), Ada Wai-Chee Fu (0.0208), Wei Wang (0.011), Jianyong Wang (0.0095)
	Influenced by Dr. Pei	ZhaoHui Tang (0.654), Chun Tang (0.6494), Shiwei Tang (0.5923), Zhengzheng Xing (0.5549), Hasan M.Jamil (0.3333), Jaewoo Kang (0.3057)

On "Data Mining" in 2009



Action-based Methods: Influence and Action

$G^t = (V^t, E^t, X^t, Y^t)$



Input:

$G^t = (V^t, E^t, X^t, Y^t)$

$t = 1,2,\ldots T$

Output:

$F: f(G^t) \rightarrow Y^{(t+1)}$

Action-based Methods: (a) Learning Influence Probabilities [1]

- **Goal:** Learn user influence and action 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

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

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.

Action-based Methods: (a) Learning Influence Probabilities [1]

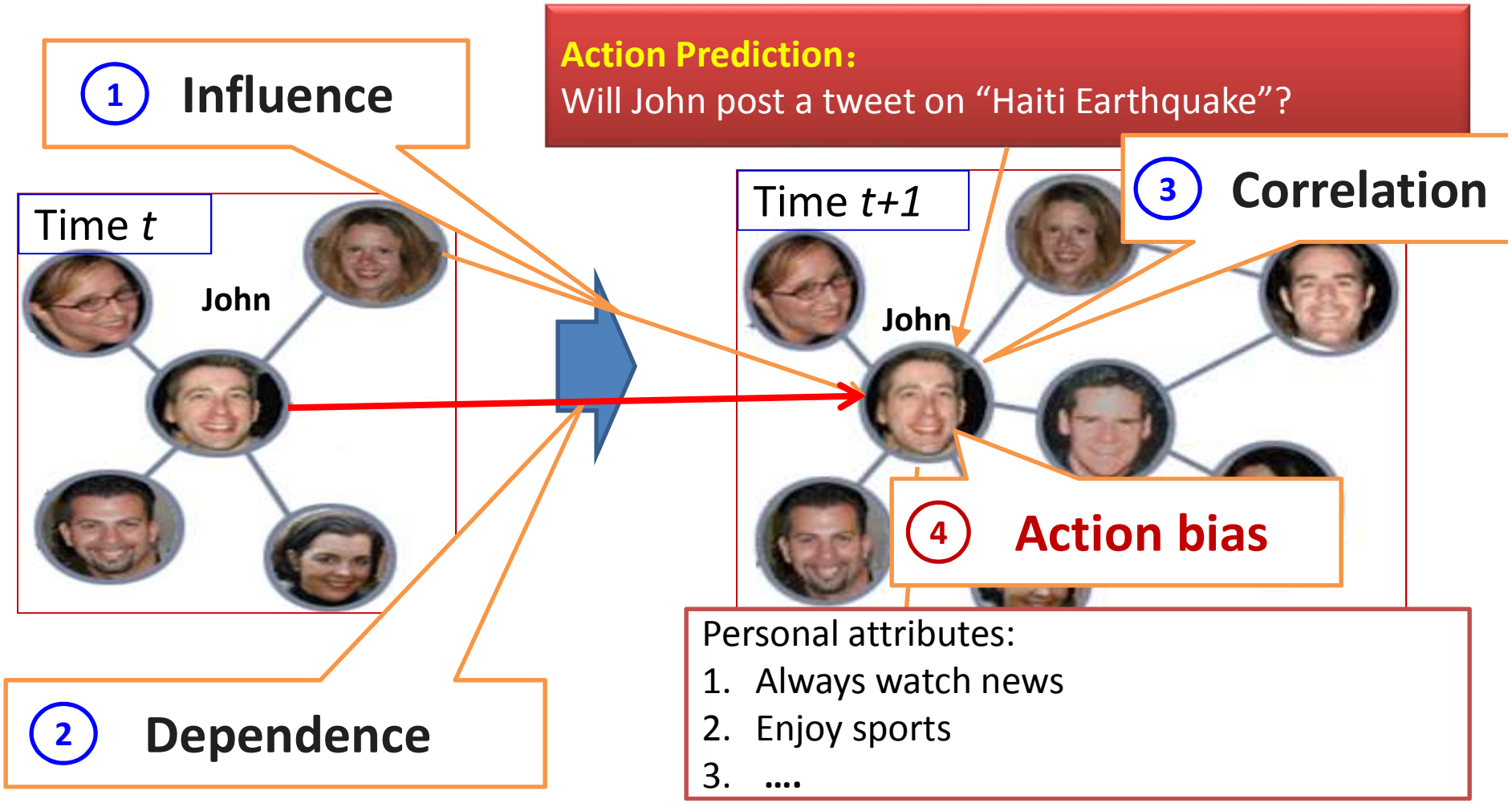
- **Action Influenceability: quantify how influenceable an action is.**

$$\text{influence}(y) = \frac{\left|\{v_j \mid \exists v_i, \Delta t : prop(y, v_i, v_j, \Delta t) \wedge \tau \geq \Delta t \geq 0\}\right|}{\text{number of users performing } y}$$

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; $prop(y, v_i, v_j, \Delta t)$ represents the action propagation score

[1] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In WSDM'10, pages 207–217, 2010.

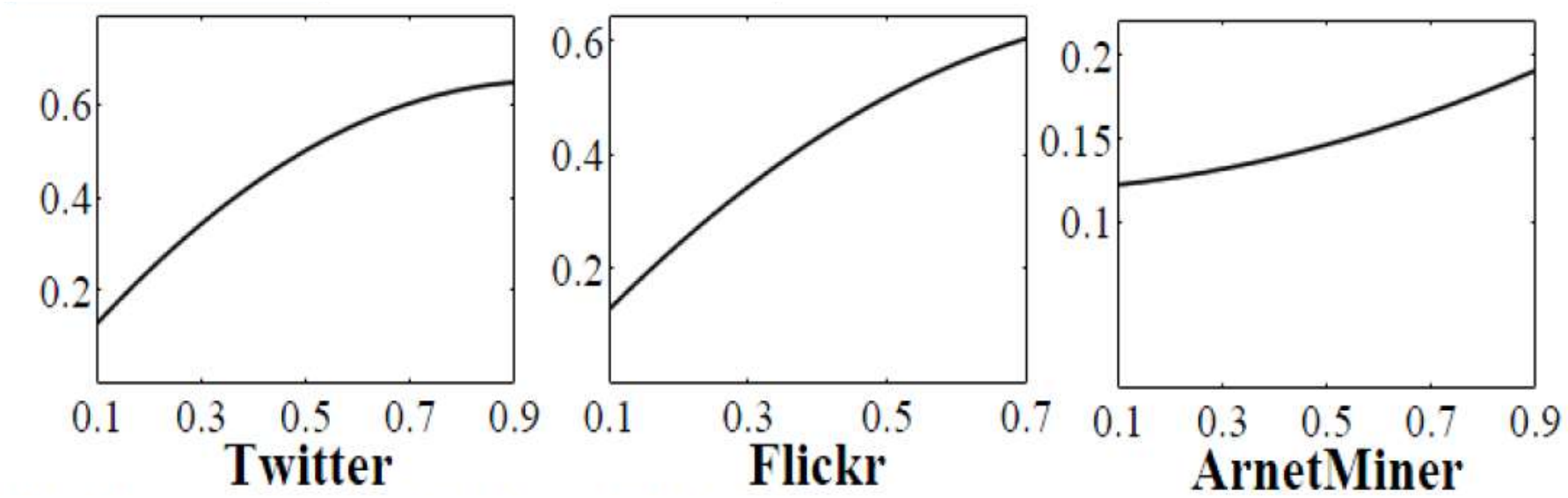
Action-based Methods: (b) Social Influence & Action Modeling[1]



[1] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In KDD'10, pages 807–816, 2010.

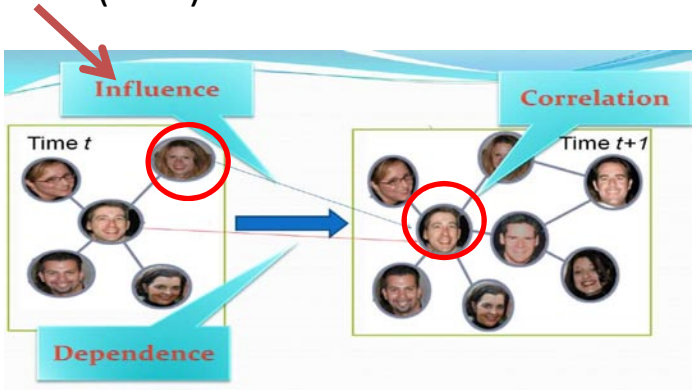
Action-based Methods: Statistical Study of Influence

Y-axis: the likelihood that the user also performs the action at time *t*

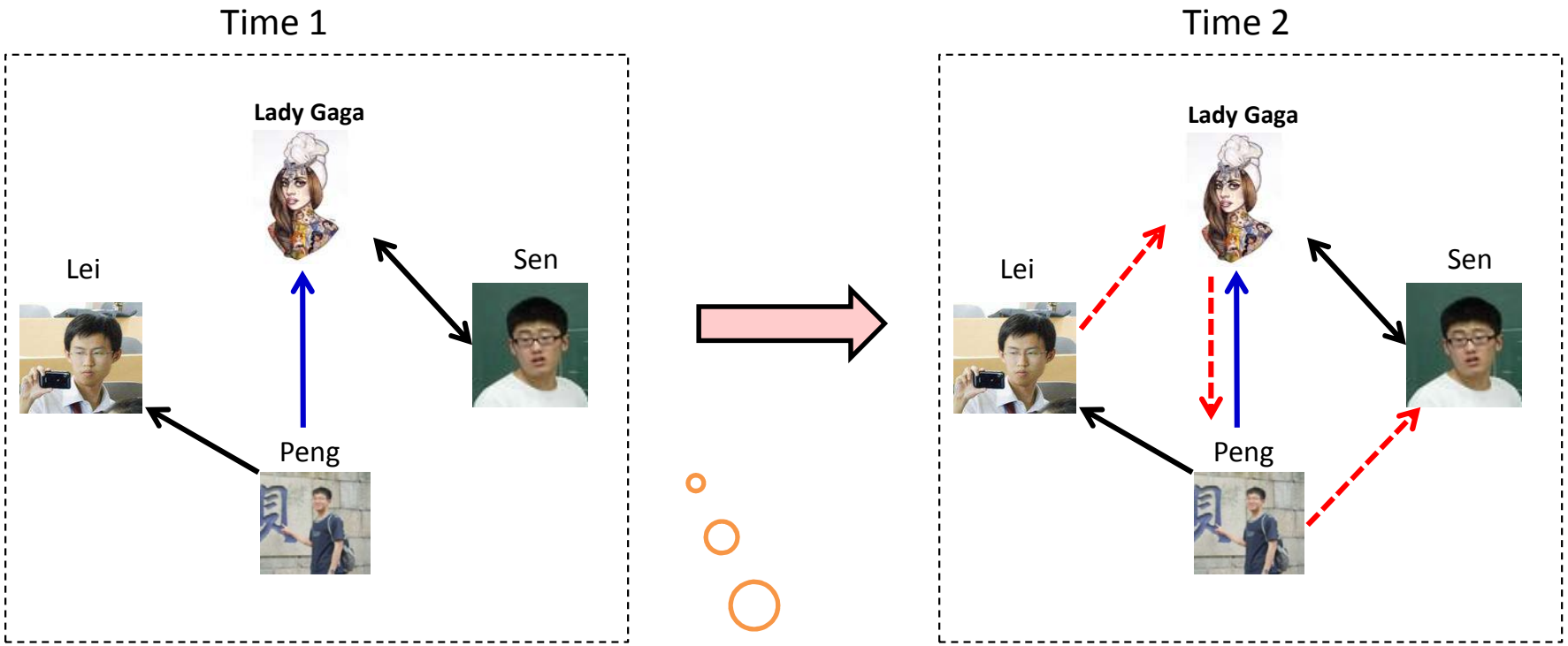


X-axis: the percentage of one's friends who perform an action at time (*t* − *1*)

Twitter Action: Tweet on "Haiti Earthquake"
Flickr Action: Add a picture into favorite list
ArnetMiner Action: Publish on a conference



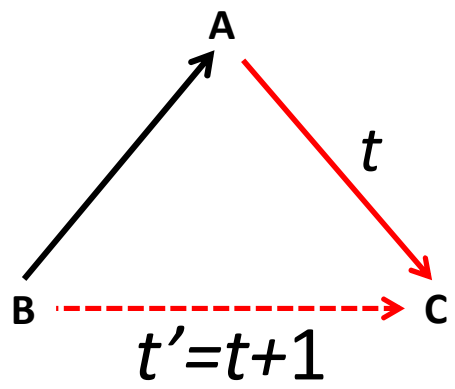
Action-based Methods: Measuring Following Influence



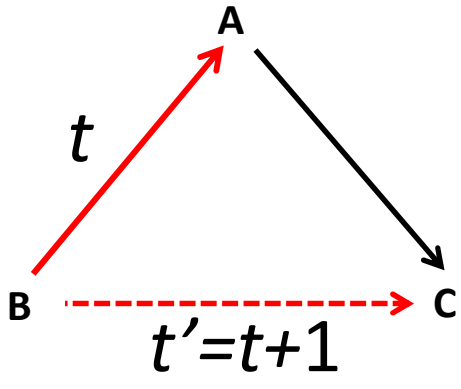
When you follow a user in a social network, will the behavior influence your friends to also follow her?

Recall we defined two kinds of influence..

Two Categories of Following Influences



Follower diffusion

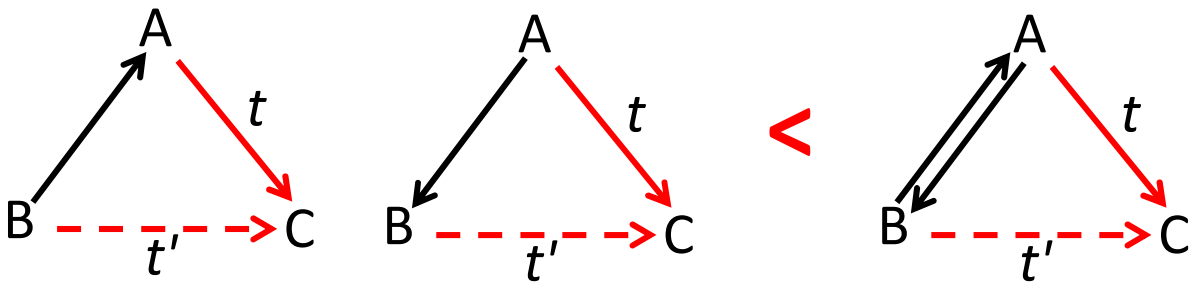


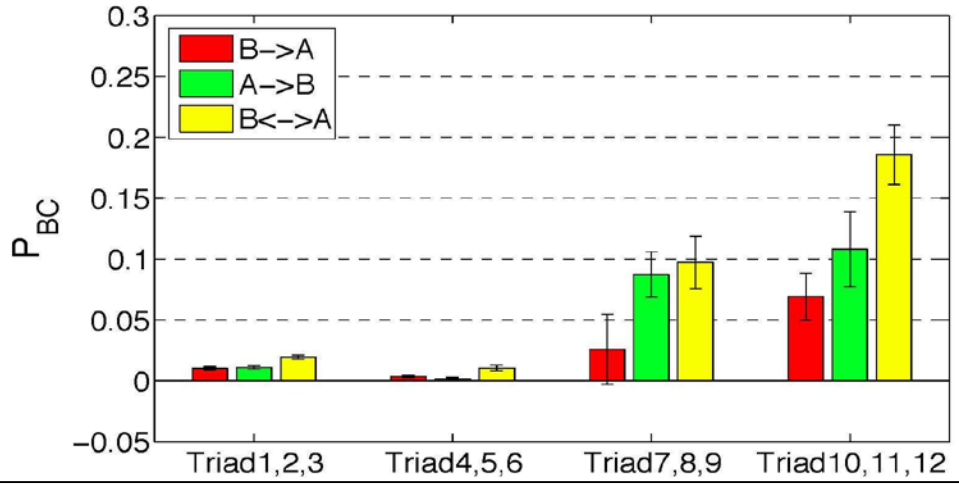
Followee diffusion

–>: pre-existed relationships
–>: a new relationship added at *t*
-->: a possible relationship added at *t*+1

One can use a generative model to learn influence probabilities between 2 users

Follower Diffusion: Power of Reciprocity

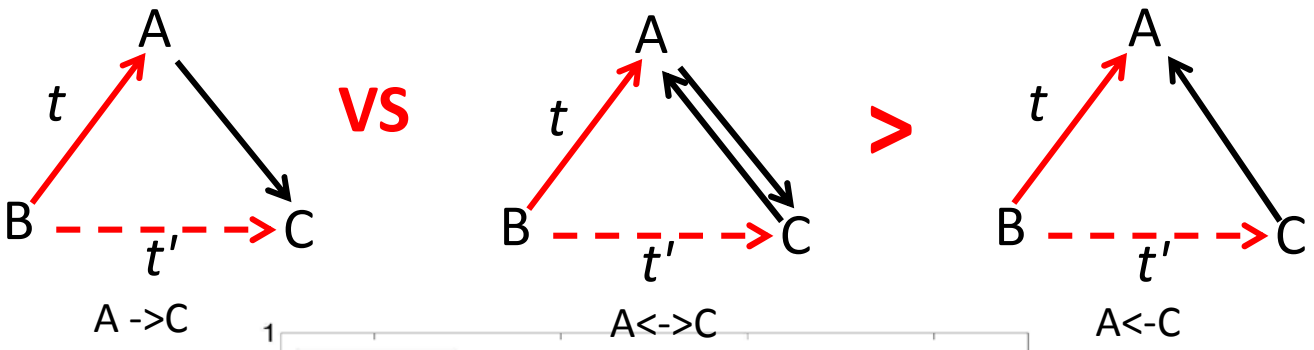


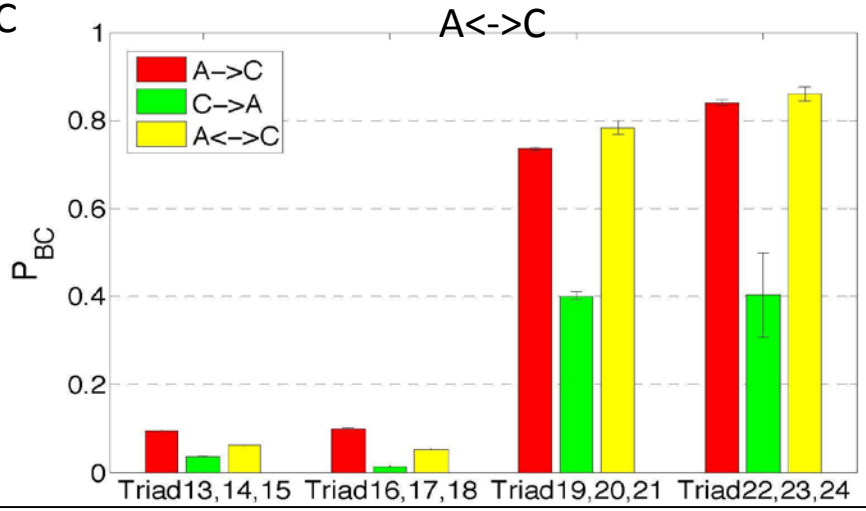


Observation: Following influence is more significant when there is a **reciprocal** relationship between B and A.

Explanation: "intimacy" is one of the three key factors that can increase people's likelihood to respond to social influence(social impact theory)

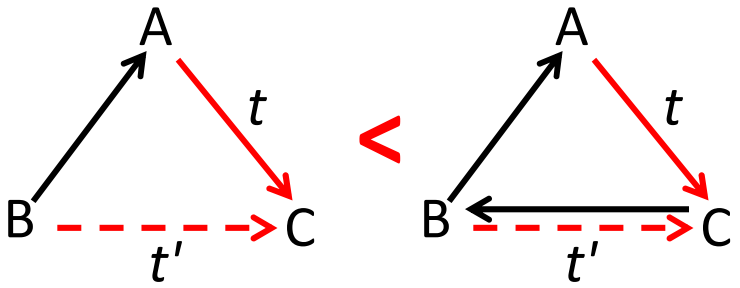
Followee Diffusion: One-way Relationship





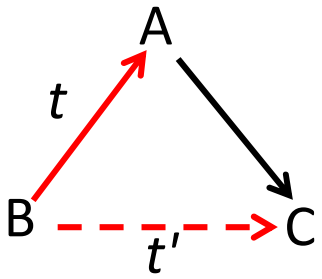
Observation: Following influence is more significant for A->C and A<->C rather than C->A.
Explanation: Users usually prefer to check their followee's followees, from whom they select those they may be interested to follow.

Reversed Relationship

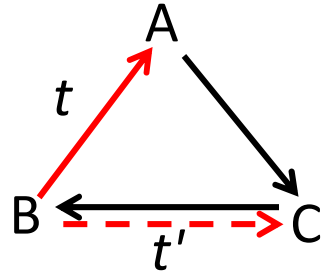


Without C->B

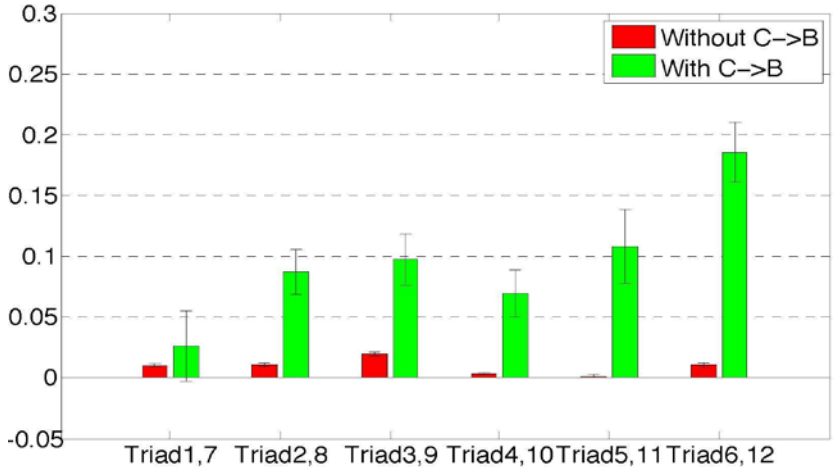
With C ->B

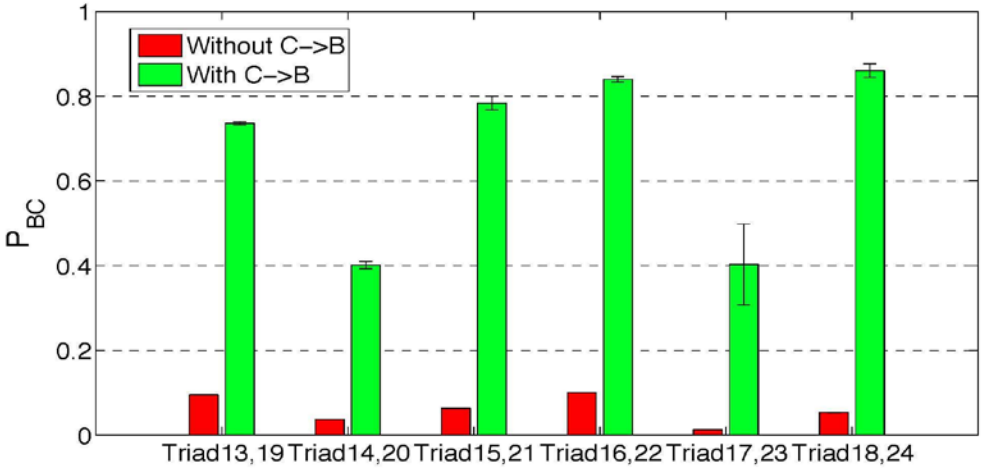


Without C->B



With C ->B

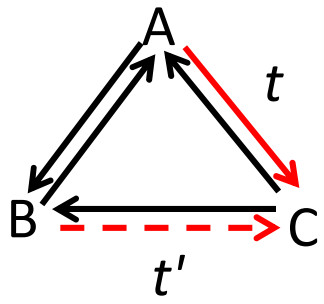




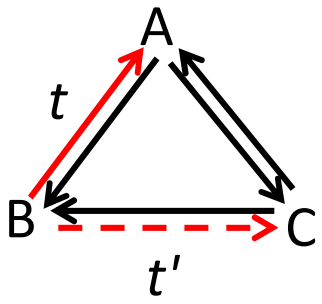
Observation: Following influence is more significant when there is a reversed relationship from C to B.

Explanation: Users are highly encouraged to follow their followers.

Social Theories: Structural Balance[1]



Follower diffusion



Followee diffusion

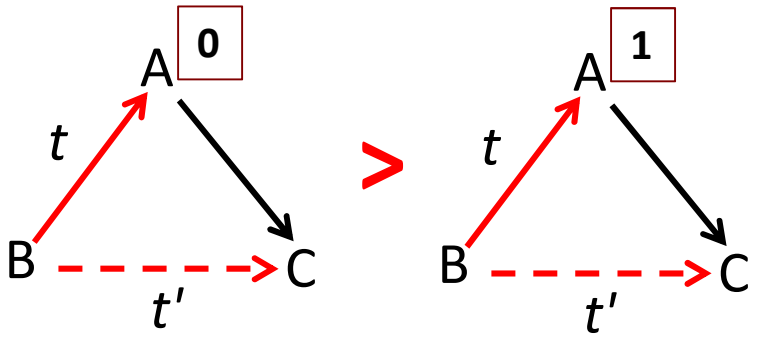
Social Balance: my friend's friend is also my friend
The probabilities of B following C in the two triads are higher than others in their respective categories.

Explanation: Users have tendency to form a balanced triad

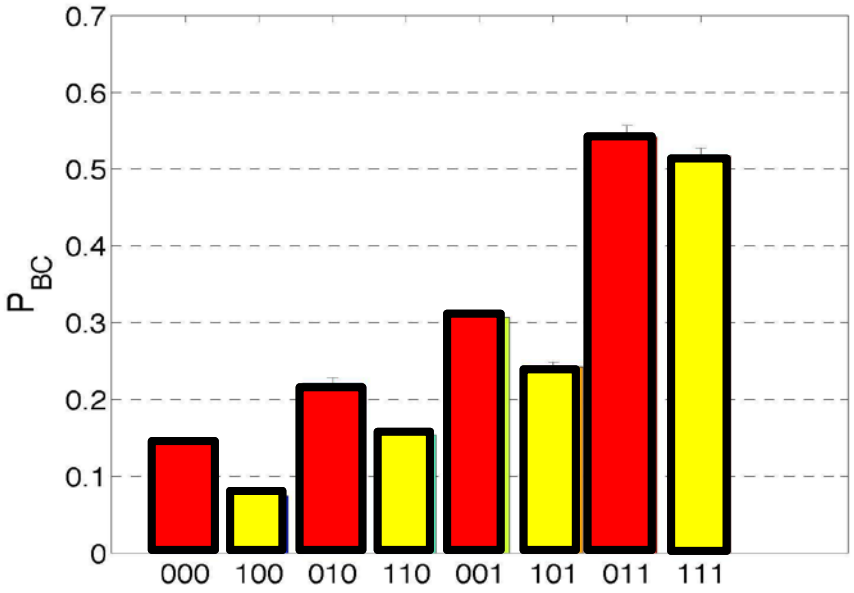
Fritz Heider (1958). The Psychology of Interpersonal Relations. John Wiley & Sons.

Social Theories: Social Status

Followee diffusion: P(0XX) > P(1XX)



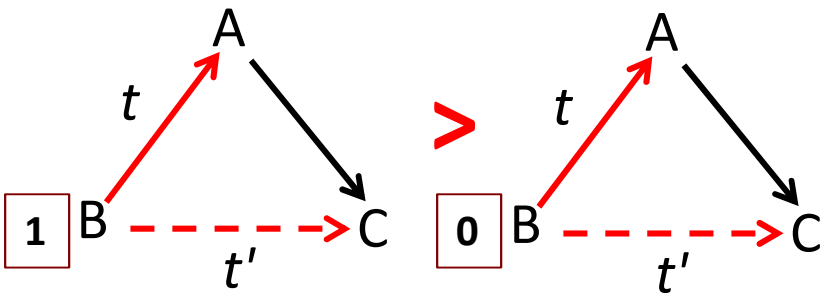
1: Elite user
0: Low-status user



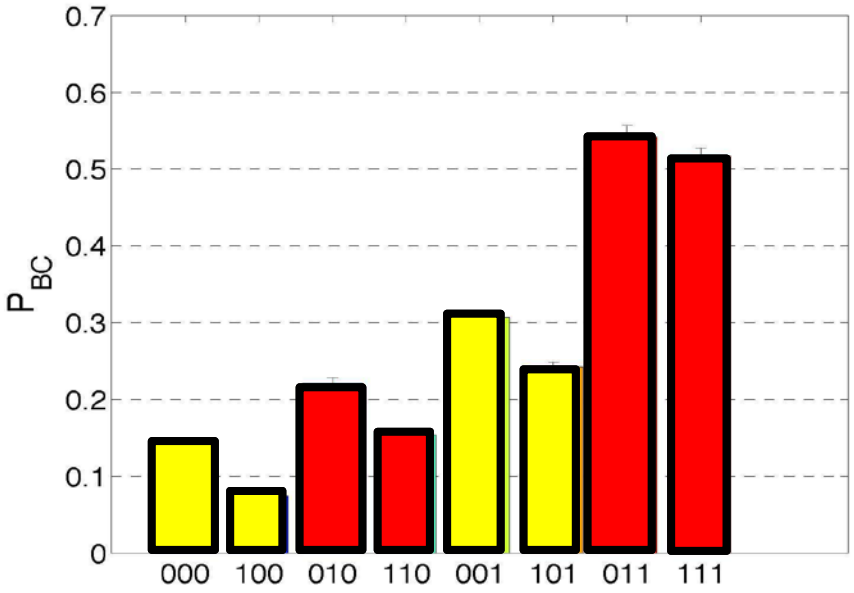
- Low-status users act as a bridge to connect users so as to form a closure triad.
- The likelihood of 0XX is 1.4 times of 1XX.

Social Theories: Social Status

Followee diffusion: P(X1X) > P(X0X)



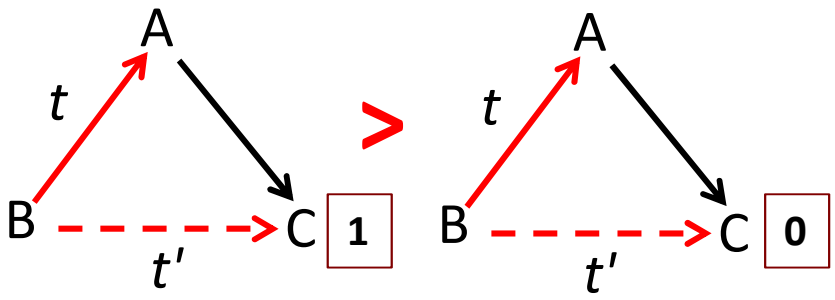
1: Elite user
0: Low-status user



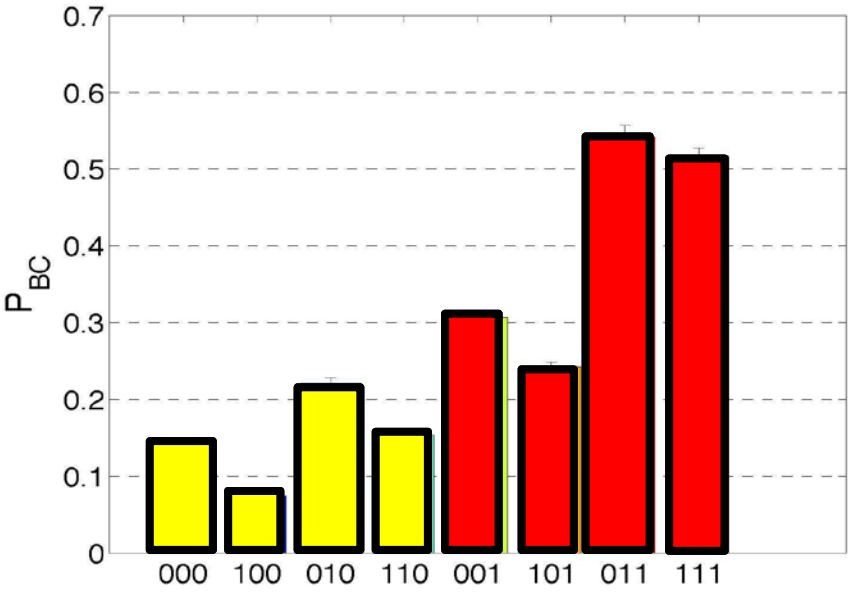
- Elite users play a more important role to form the triadic closure.
- The likelihood of X1X is almost double the probability of X0X.

Social Theories: Social Status

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



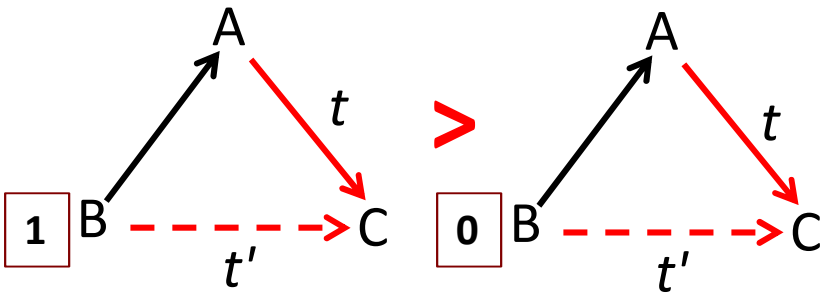
1: Elite user
0: Low-status user



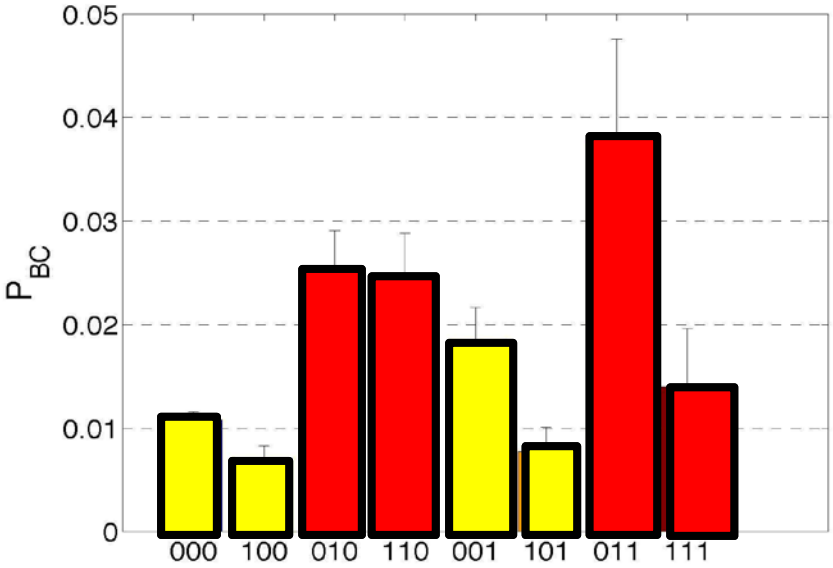
- The rich gets richer.
- The likelihood of XX1 is nearly 2 times higher than that of XX0.
- This phenomenon validates the mechanism of preferential attachment.

Social Theories: Social Status

Follower diffusion: P(X1X) > P(X0X)

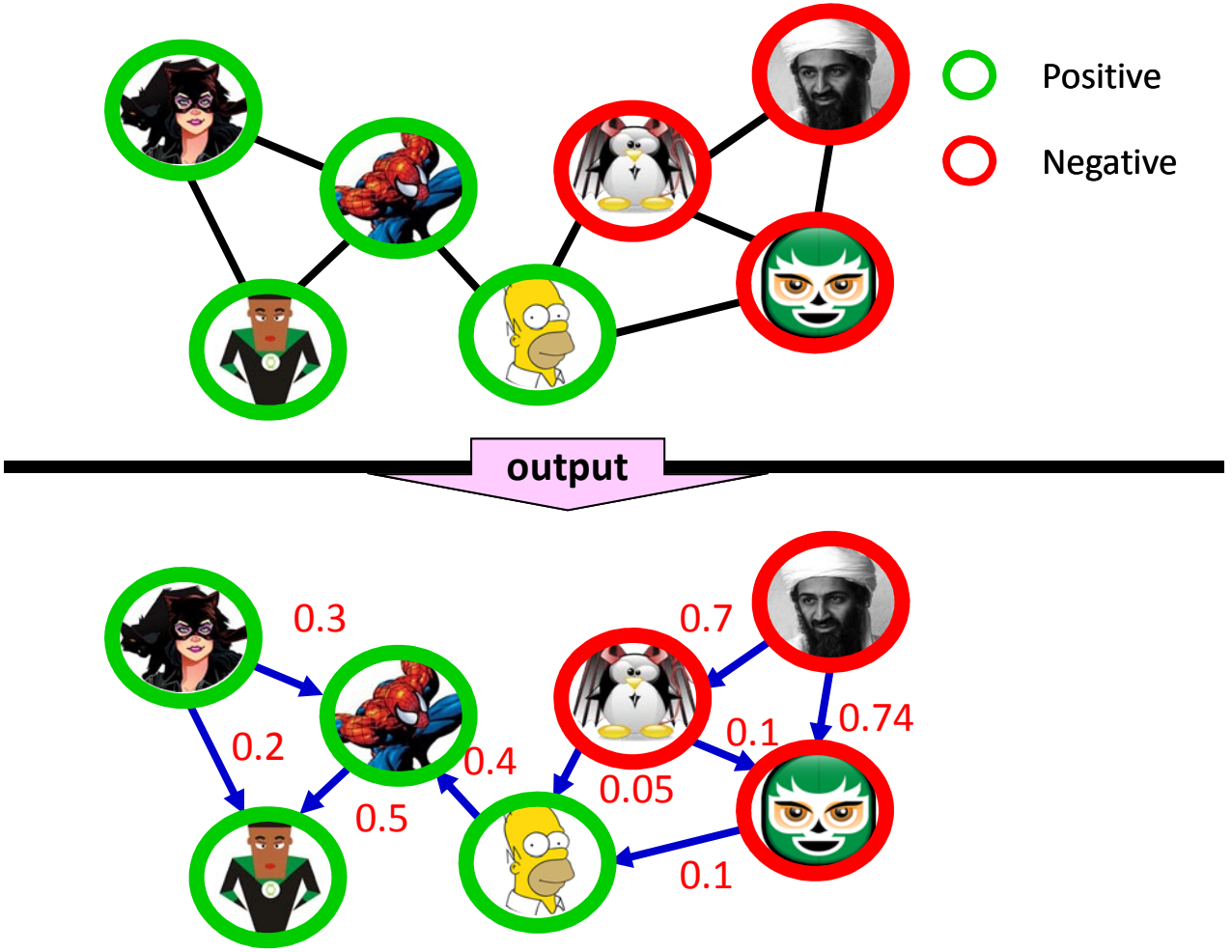


1: Elite user
0: Low-status user



- Elite users play a more important role to form the triadic closure.
- The likelihood of X1X is almost double the probability of X0X.

Output of Measuring Influence



Take-away Messages

- We discussed basics of social influence and observed examples where social influence is prominent.
- We discussed the following Existential Tests for Social Influence
  - Randomization Test
  - Shuffle Test
  - Reverse Test
- We studied the following methods for measuring Social Influence Analysis
  - Reachability-based methods
  - Structure Similarity
  - Structure + Content Similarity
  - Action-based methods

Further Reading

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

Preview of Lecture 12: Social Influence Analysis (Part 2)

- Models for Social Influence Analysis
  - Decision based models
  - Probabilistic Models
- Influence Maximization
- Applications of Social Influence Analysis
  - Social advertising
  - Opinion leader finding
  - Social recommendation
  - Emotion analysis

Disclaimers

- This course represents opinions of the instructor only. It does not reflect views of Microsoft or any other entity (except of authors from whom the slides have been borrowed).
- Algorithms, techniques, features, etc mentioned here might or might not be in use by Microsoft or any other company.
- Lot of material covered in this course is borrowed from slides across many universities and conference tutorials. These are gratefully acknowledged.

Thanks!