IIIT-H

Web Mining
Lecture 12: Social Influence Analysis (Part 2)

Manish Gupta

11th Sep 2013

Slides borrowed (and modified) from
http://keg.cs.tsinghua.edu.cn/jietang/publications/WSDM13-tutorial-social-influence-analysis.pptx
http://www.stanford.edu/class/cs224w/slides/

Recap of Lecture 11: Social Influence Analysis (Part 1)

- 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

Networks with Negative Links

Slashdot

Char

stories
submissions
popular
blog
all stories
ask slashdot
book reviews
games
idle
yro
cloud
hardware
linux
management
mobile

Slashdot FAQ

- About Slashdot
- Editorial—Story Submissions, etc.
- Slashdot Interviews
- Tags
- Accounts
- Moderation and Metamoderation
- Karma, Thresholds, Abuse, etc.
- The Firehose

The Friends and Foes System

How do I add users to my friends / foes list (or remove them)?

Search for the user. Click. Add the user as your friend or foe.

You'll find those symbols on users' comments as well, and elsewhere on the site. Basically, anywhere you see the following symbol you can decide to add a user as friend or foe.

To remove a friend or foe, click on their friend/foe symbol and change them back to Neutral.

Where can I see my friends / foes?
Can I see other people's friends and foes?

Announcements

- Project guidelines are up, have a look and let me know if you have any questions
  - Groups of 3
  - Total 12 proposed projects; you can propose a new one broadly in the area of web mining
  - Deadline for choice + optionally your project is Sep 14, 9pm
  - Final allocation of projects by Sep 21
  - Students should ideally start working on the project from Oct 15
  - Mid-project review deadline: Oct 27
  - Final report+presentation+code with posters will be on Nov 20
- Schedule change
  - 25^{th} Sep class moved to 20^{th} Sep 6-7:30pm
  - 28^{th} Sep class moved to 30^{th} Sep 6-7:30pm

Today's Agenda

- 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

Today's Agenda

- **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

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 *k* 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

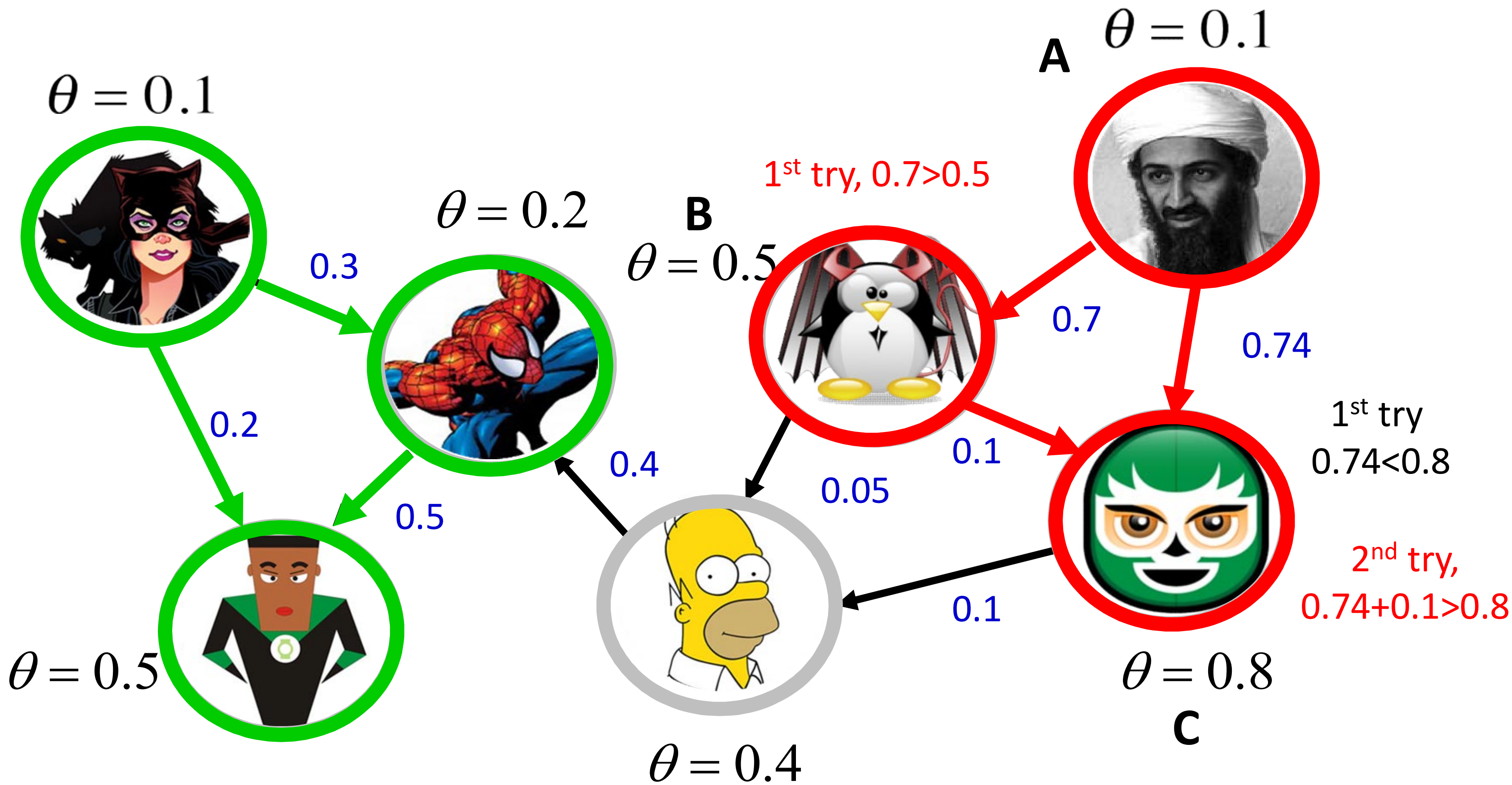
Decision Based Model: Linear Threshold Model (LTM)

- General idea
  - Whether a given node will be active can be based on an arbitrary monotone function of its neighbors that are already active.
- Formalization
  - f_v : map subsets of v's neighbors' influence to real numbers in [0,1]
  - θ_v : a threshold for each node
  - S: the set of neighbors of v that are active in step t-1
  - Node v will turn active in step t if $f_v(S) > \theta_v$
- Specifically, in [Kempe, 2003], f_v is defined as $\sum_{u \in S} b_{v,u}$, where $b_{v,u}$ can be seen as a fixed weight, satisfying

$$\sum_{v \in N(u)} b_{u,v} \leq 1$$

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'03), pages 137–146, 2003.

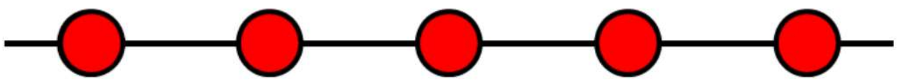
Linear Threshold Model: An example



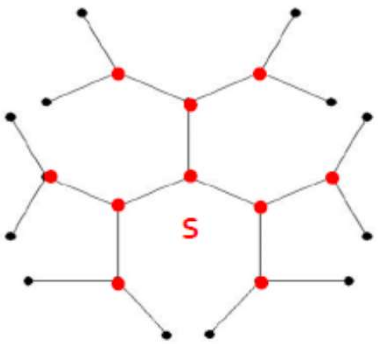
Properties of LTM (Single Contagion)

- Monotonic Spreading
  - Use of the product spreads monotonically (Nodes adopt the product; but don't un-adopt)
- Cascade effect on Infinite Graphs
  - Consider infinite graph *G,* with each node having finite number of neighbors
  - Let S be the seed set of nodes spreading the contagion
  - We say that a finite set S causes a cascade in G with threshold θ if, when S adopts product A, eventually every node adopts A

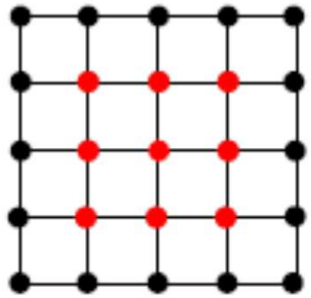
Infinite Path: If $\theta \leq 0.5$, cascade occurs



Infinite Tree: If $\theta \leq 1/3$, cascade occurs

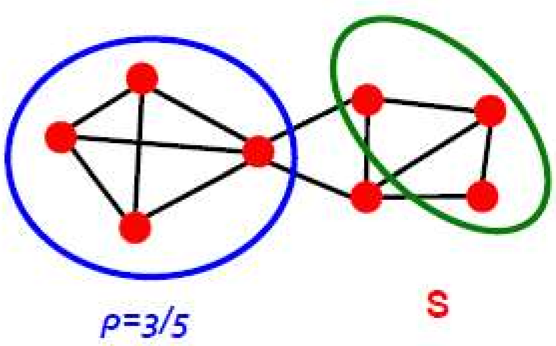


Infinite Grid: If $\theta \leq 1/4$, cascade occurs



Properties of LTM (Single Contagion)

- Cascade capacity of a graph G is the largest θ for which some finite set S can cause a cascade
  - There is no G where cascade capacity>0.5
- What prevents cascades from spreading?
  - Cluster of density ρ is a set of nodes C where each node in the set has at least ρ fraction of edges in C.
  - Let S be an initial set of adopters of product A
  - All nodes apply threshold θ to decide whether to switch to A
  - Two facts
    - If G\S contains a cluster of density >(1-θ) then S can not cause a cascade
    - If S fails to create a cascade, then there is a cluster of density >(1-θ) in G\S



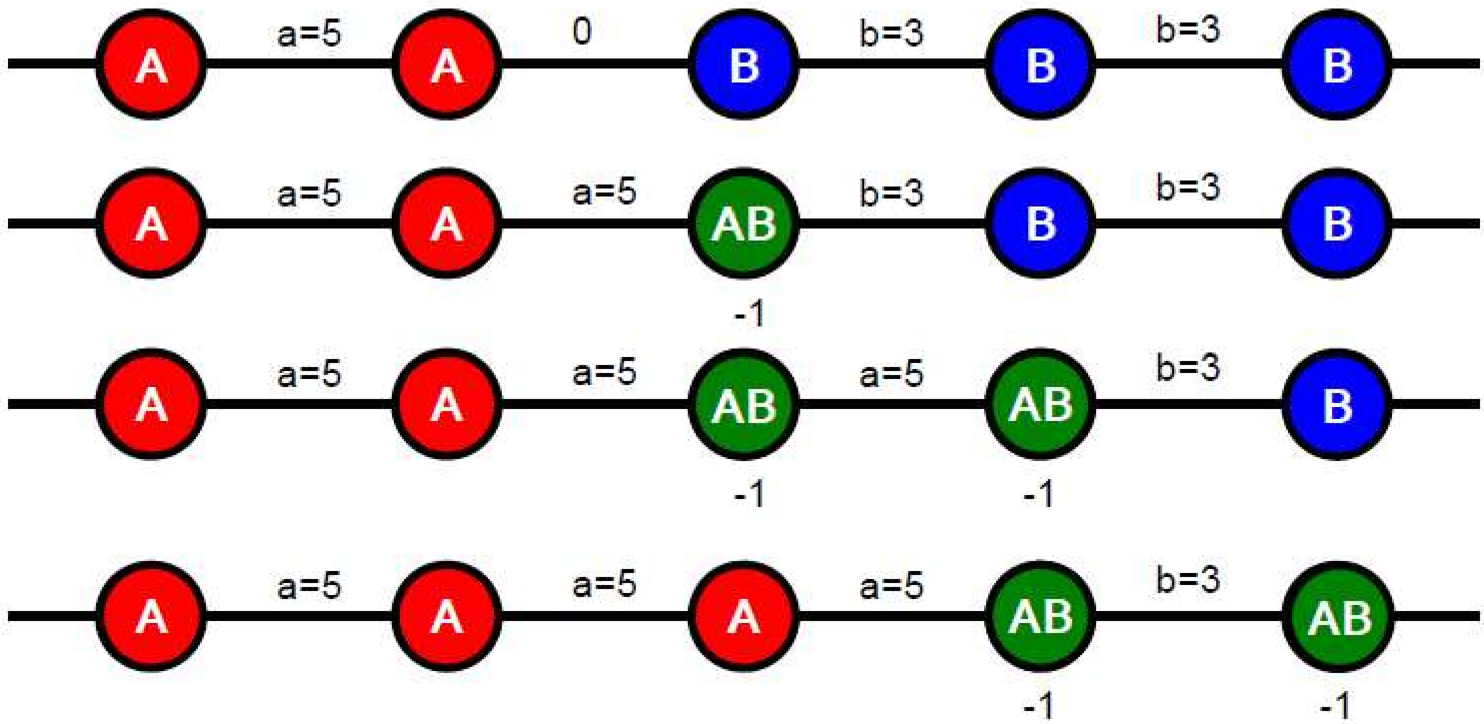
No cascade if $\theta > 2/5$

In Presence of Multiple Contagion

- Till now, behaviors *A* and *B* competed
  - Can only get utility from neighbors of same behavior: *A*-*A* get a, *B*-*B* get *b*, *A*-*B* get *0*
- Let the nodes now have A, B or AB
- *AB*-*A*: gets *a; AB*-*B*: gets *b; AB*-*AB*: gets max(*a, b*); cost *c* for the effort of maintaining both strategies
- Say the graph had all Bs; A is a new/better product

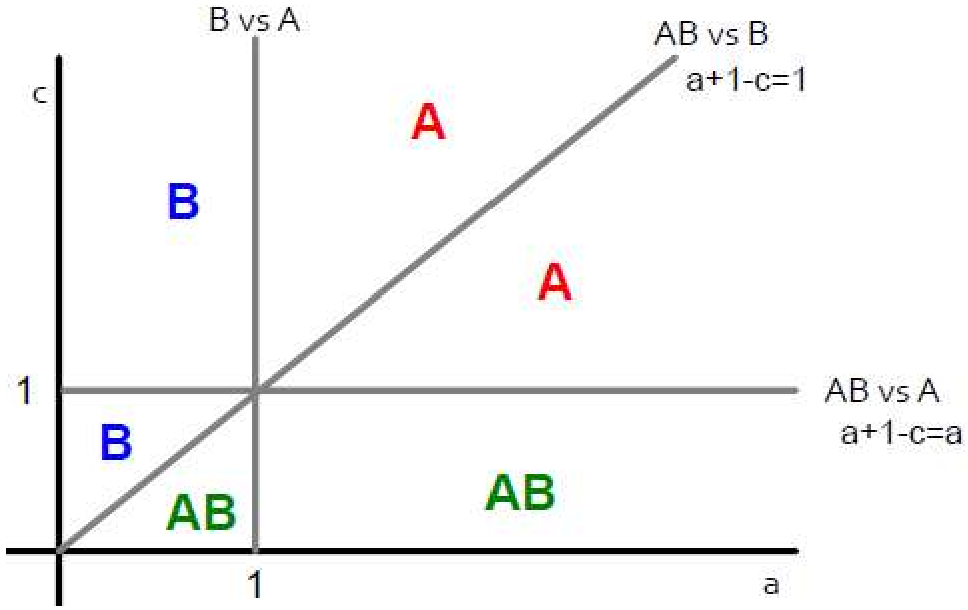
Spread of a New Better Contagion in Presence of Another

- Infinite path start with all Bs; let a=5, b=3, c=1



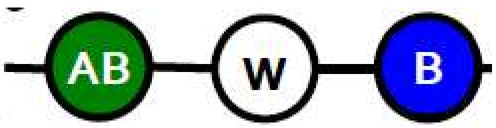
Values of "a" and "c" for Spread of A

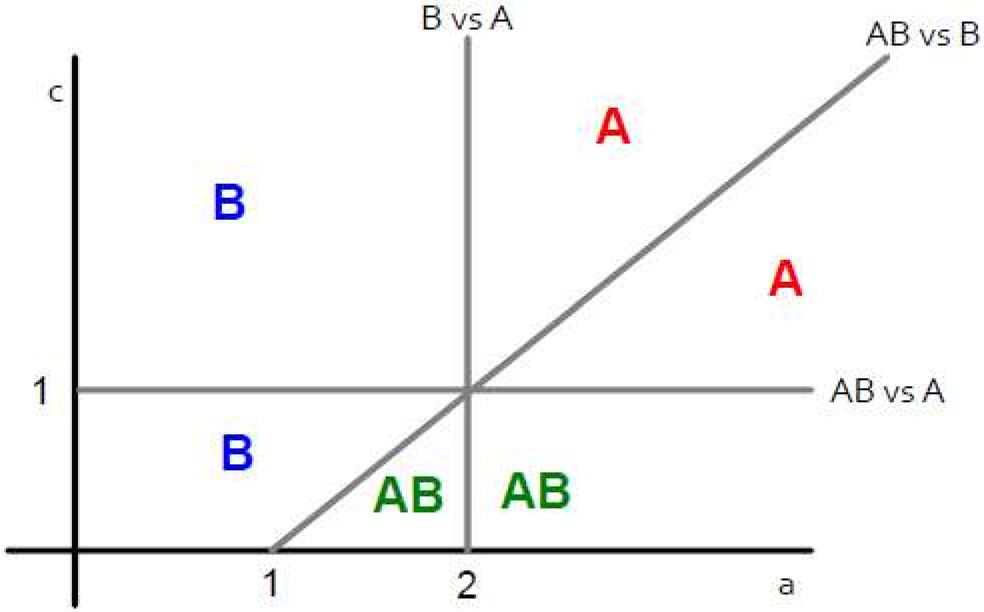
- Infinite path, start with all Bs
- Payoffs for *w*: A:a, B:1, AB:a+1-c
- What does node w in A-w-B do?



Values of "a" and "c" for Spread of A

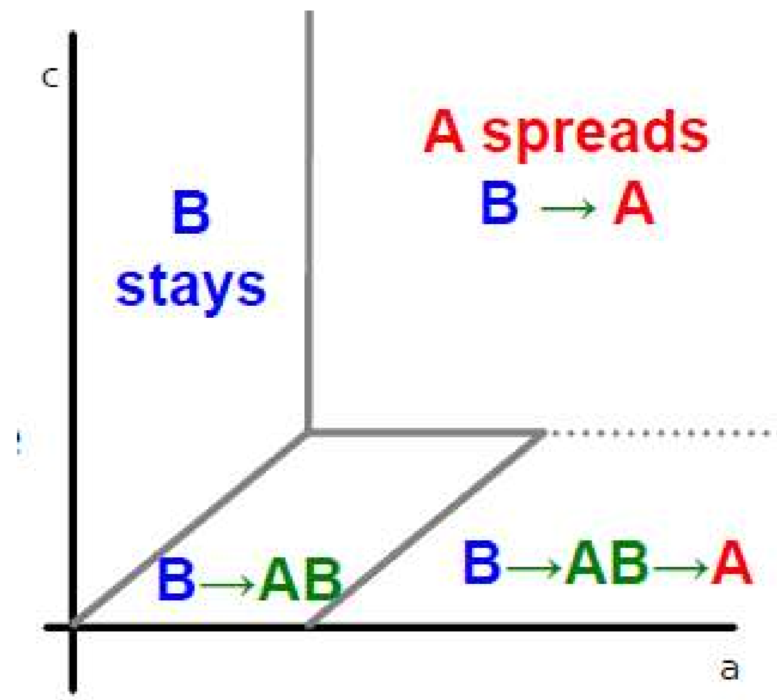
- Same reward structure as before but now payoffs for *w* change: A:a, B:1+1, AB:a+1-c
- Notice: Now also AB spreads
- What does node w in AB-w-B do?





Overall (Joining the Two Pictures)

- You manufacture default B and new/better A comes along
  - Infiltration: If B is too compatible then people will take on both and then drop the worse one (B)
  - Direct conquest: If A makes itself not compatible – people on the border must choose. They pick the better one (A)
  - Buffer zone: If you choose an optimal level then you keep a static “buffer” between A and B



Today's Agenda

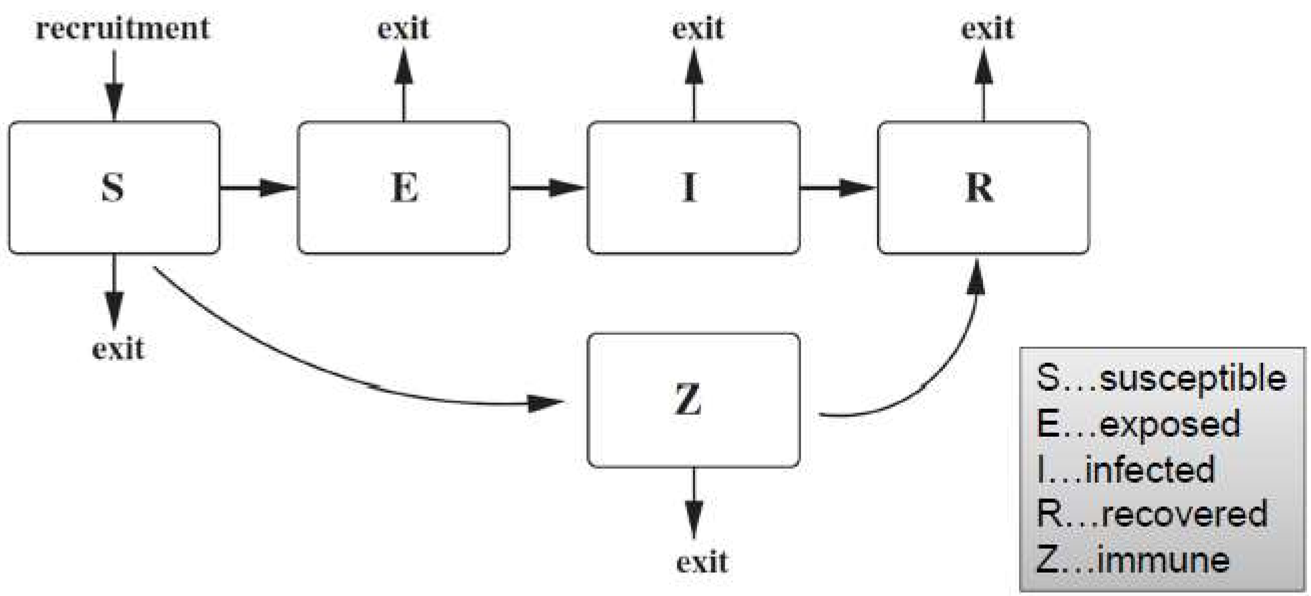
- Models for Social Influence Analysis
  - Decision based models
  - **Probabilistic Models**
- Influence Maximization
- Applications of Social Influence Analysis
  - Social advertising
  - Opinion leader finding
  - Social recommendation
  - Emotion analysis

Epidemic Model based on Trees

- Epidemic Model based on Random Trees
  - A patient meets *d* other people
  - With probability *q* > *0* infects each of them
- For which values of *d* and *q* does the epidemic run forever?
  - P_h= $P[Infected\ node\ at\ depth\ h]$
  - Run forever: $\lim_{h\to\infty} P_h > 0$
  - Die out: $\lim\limits_{h\to\infty} P_h = 0$
  - $P_h = 1 - (1 - q.P_{h-1})^d$
  - One can prove that there will be an epidemic if $qd \geq 1$

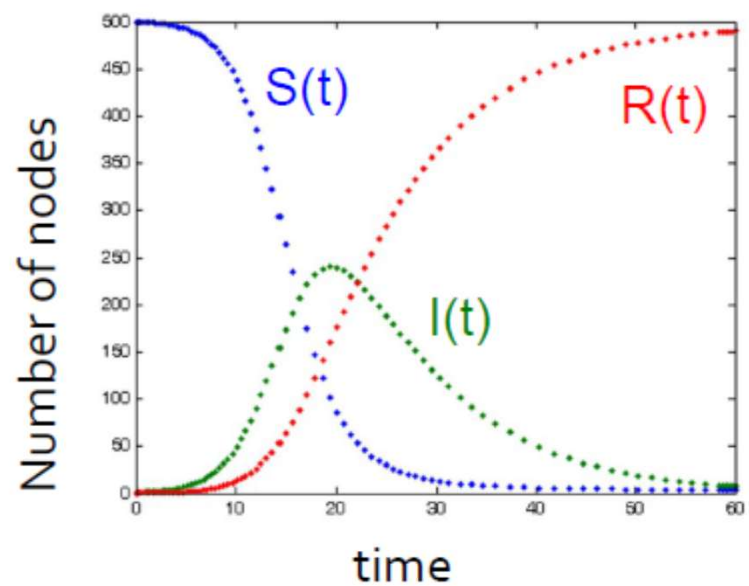
Classical Models of Disease Spreading

- Till now, nodes went from healthy→infected
- SEIR is a richer model
  - Each node goes through phases
  - Transition probabilities are governed by model parameters



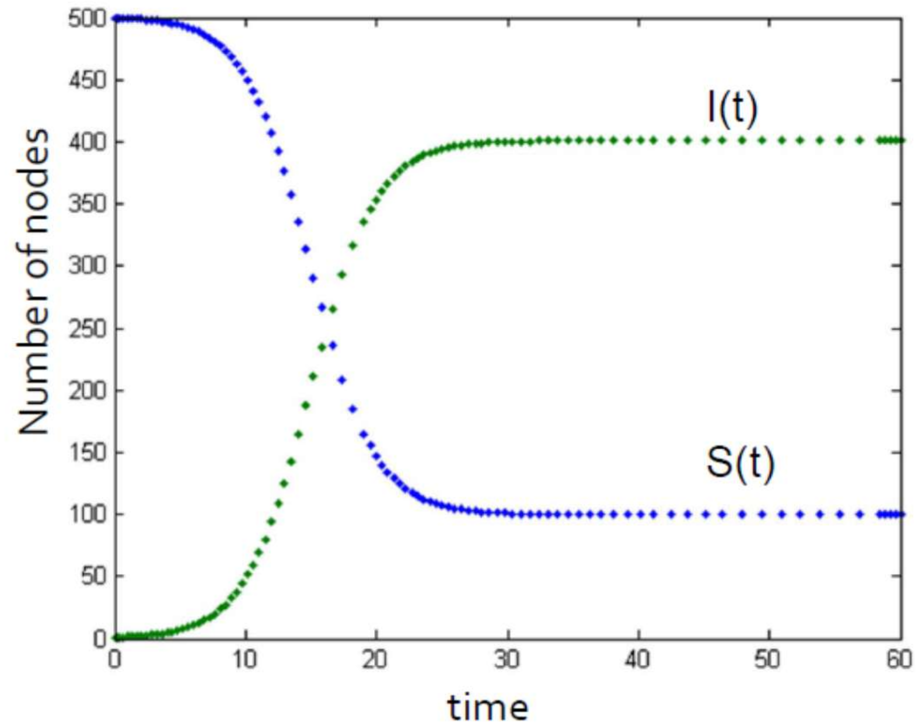
Classical Models of Disease Spreading (SIR)

- (Virus) birth rate β: probability than an infected neighbor attacks
- (Virus) death rate δ: probability that an infected node heals
- Model dynamics
  - $\frac{dS}{dt} = -\beta SI$
  - $\frac{dR}{dt} = \delta I$
  - $\frac{dI}{dt} = \beta SI - \delta I$

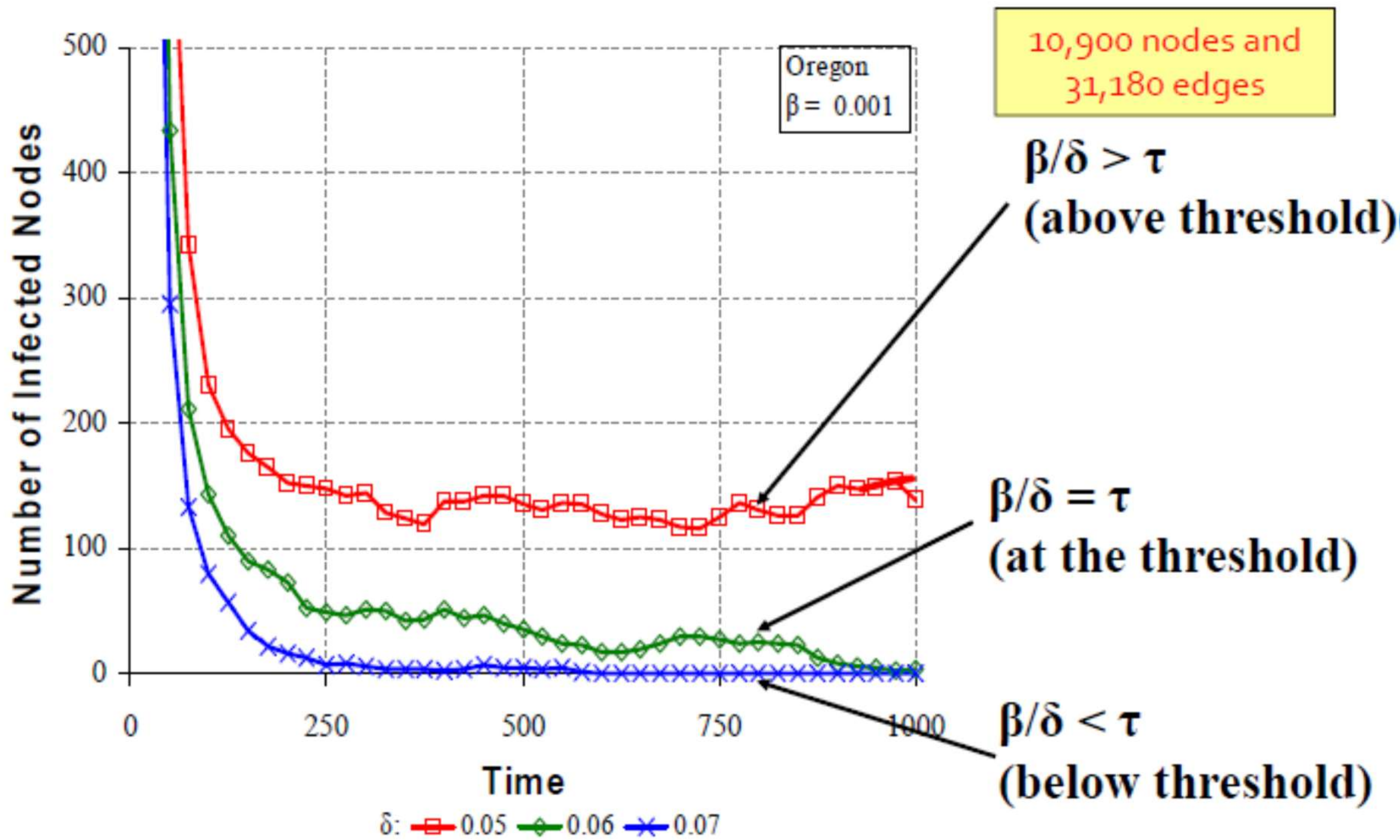


Classical Models of Disease Spreading (SIS)

- Virus "strength": s = β / δ
- Model dynamics
  - $\frac{dS}{dt} = -\beta SI + \delta I$
  - $\frac{dI}{dt} = \beta SI - \delta I$
- If virus strength<τ, epidemic cannot happen
- One can prove that $\tau = 1/\lambda$ where λ is the largest eigenvalue of the adjacency matrix of the graph

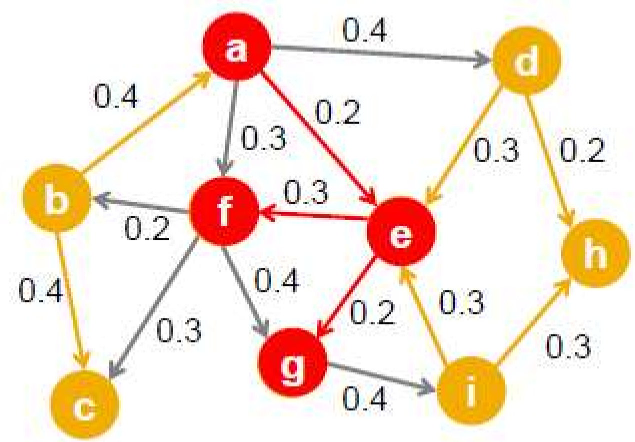


Experiments on Autonomous Systems Network



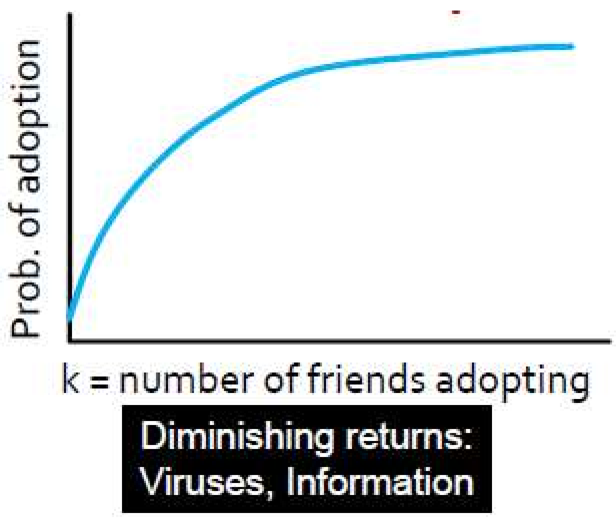
Independent Cascade Model

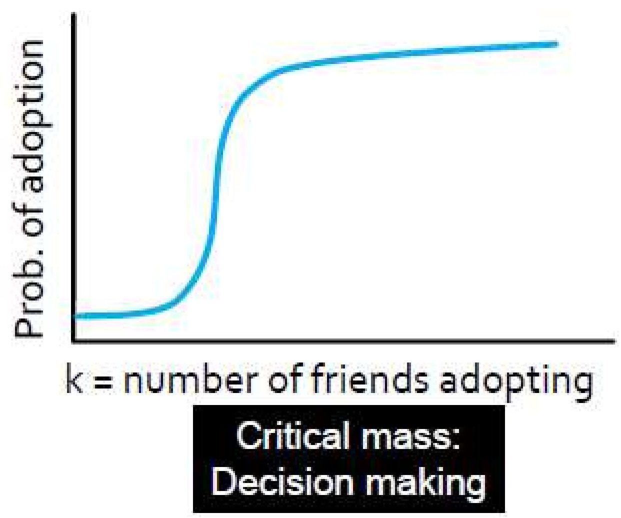
- Initially some nodes *S* are active
- Each edge *(u,v)* has probability (weight) p_{uv}
- When node v becomes active
  - It activates each out-neighbor *v* with prob. p_{uv}
- Activations spread through the network
- Independent cascade model is simple but requires many parameters!



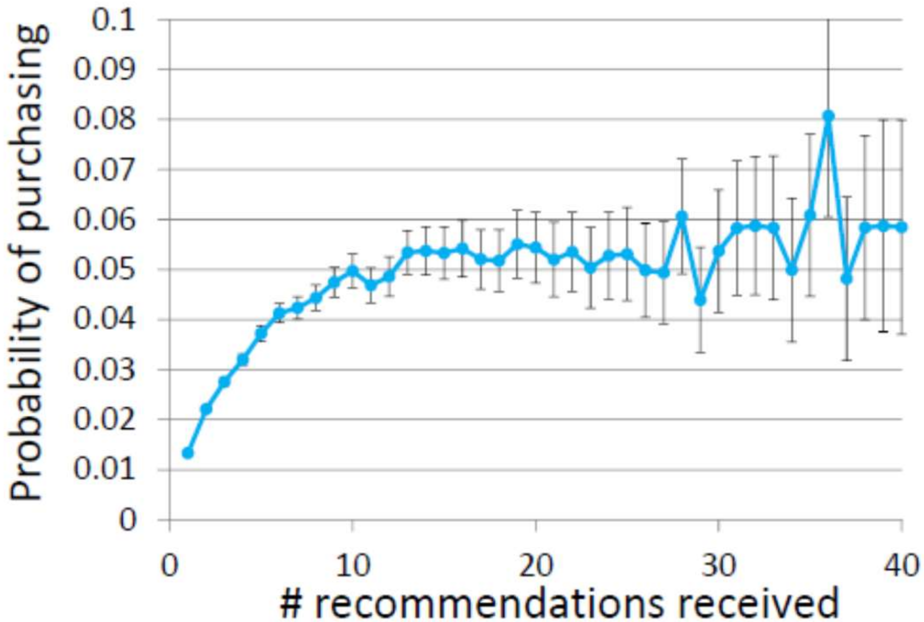
Exposure Curves

- Probability of adopting new behavior depends on the number of friends who have already adopted
- What's the dependence?

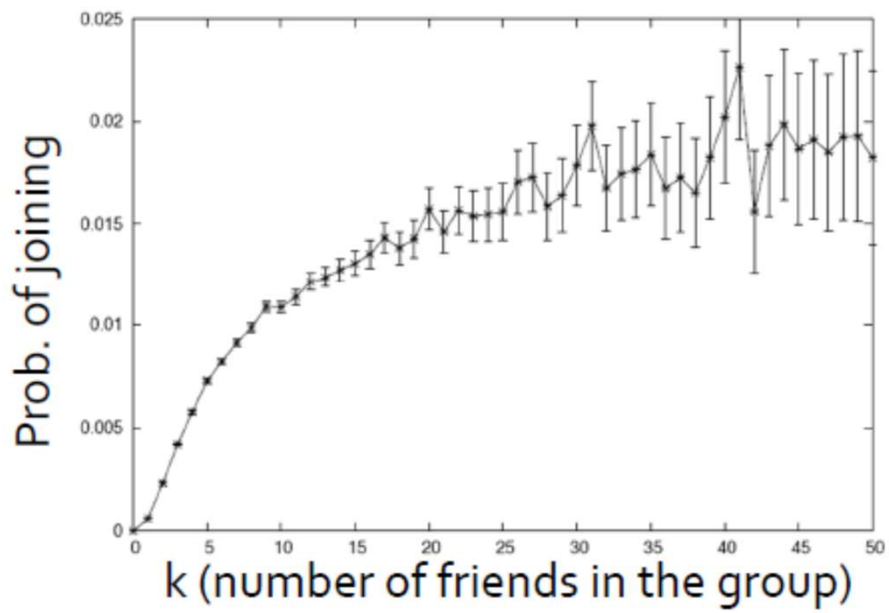




Exposure Curves for Large Real Datasets

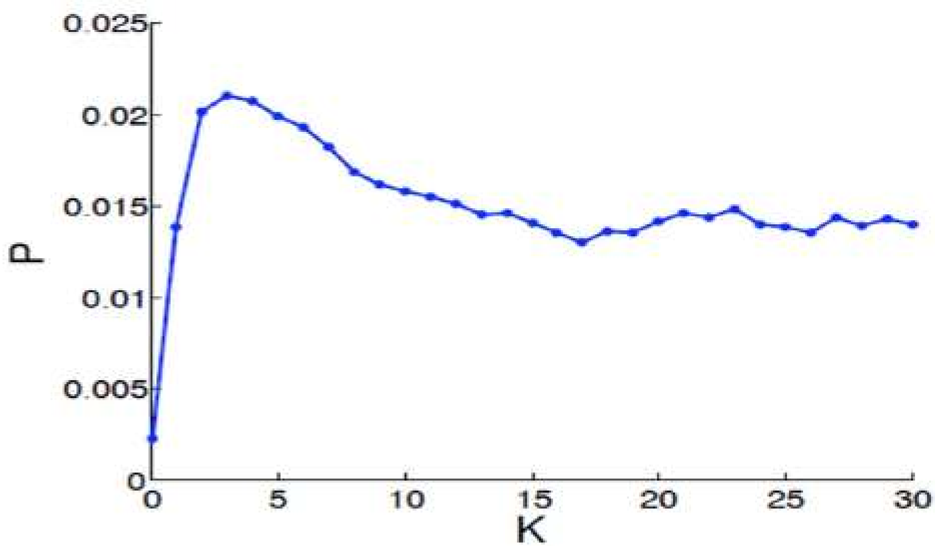


DVD Recommendations
(8.2 million observations)



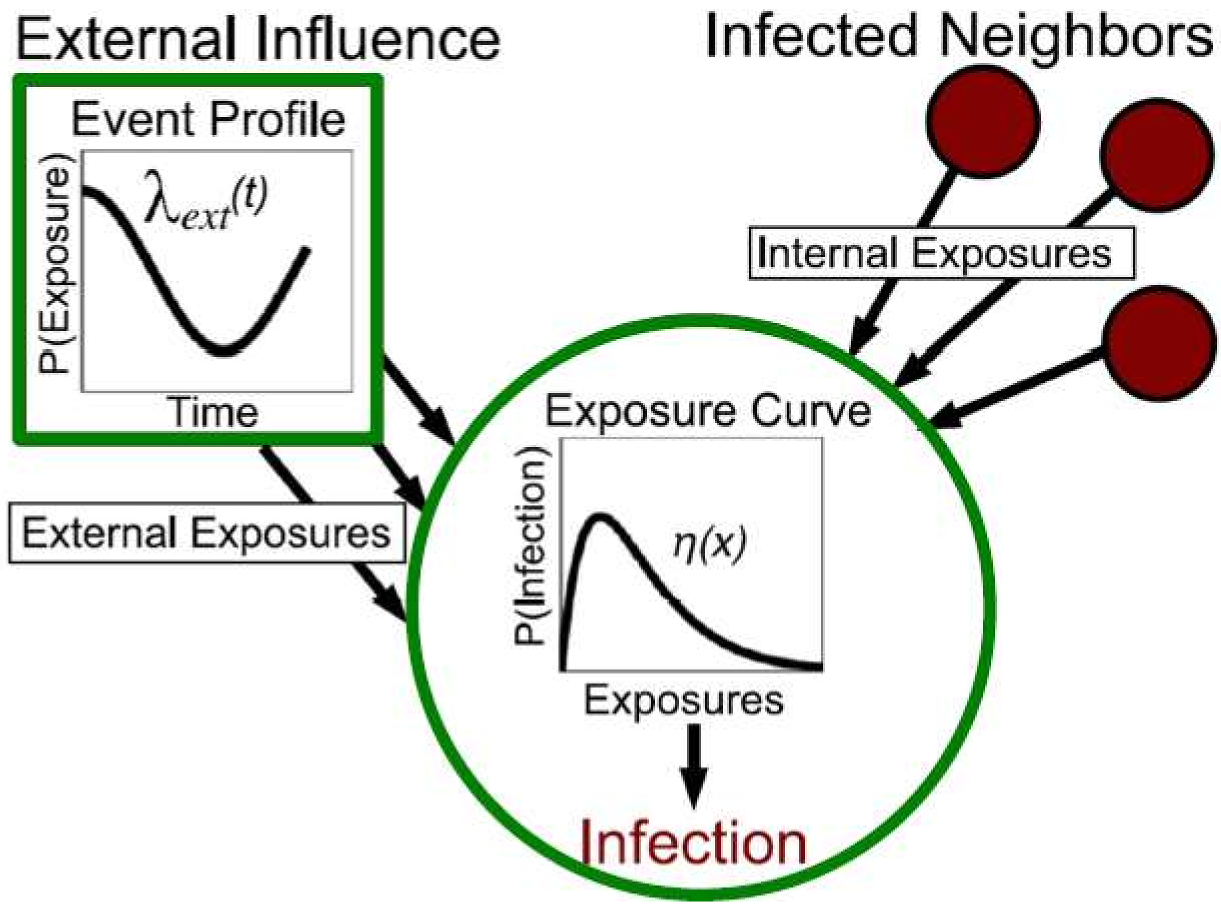
LiveJournal group
membership

Exposure Curves for Large Real Datasets (Twitter)



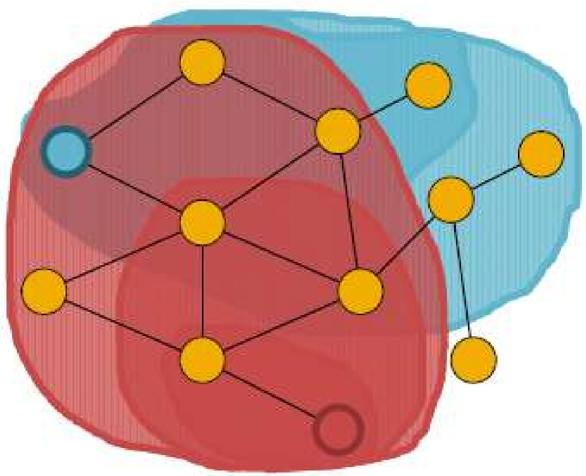
- 3B tweets, 60M users
- Curve reaches peak fast, decreases after!
- Modeling the shape of the curve
  - Persistence: Area under the curve/Area of bounding rectangle
  - Max value of the curve

Incorporating External Influences



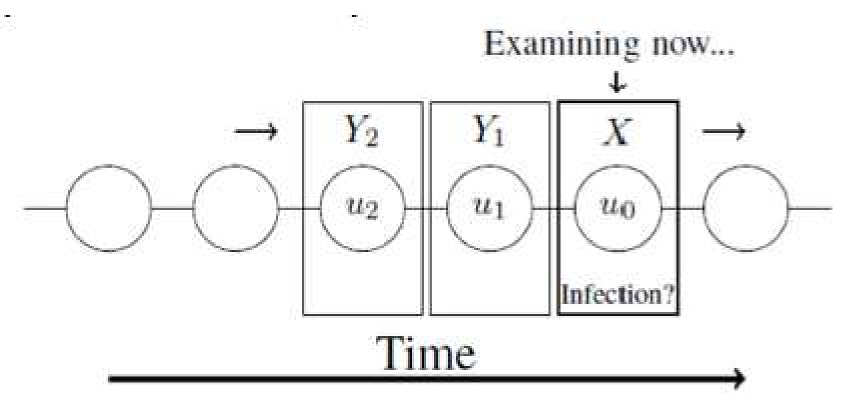
Modeling Interactions Between Contagions

- Do pieces of information interact?
  - Does being exposed to blue change the probability of talking about red?
- Goal: Model interaction between many pieces of information
  - Some pieces of information may help each other in adoption
  - Other may compete for attention



Modeling Interactions

- You are reading posts on Twitter
  - You examine posts one by one
  - Currently you are examining X
  - How does your probability of reposting X depend on what you have seen in the past?



P(post X | exposed to X, Y_1, Y_2, Y_3) = ?

Modeling Interactions

- Goal: P(post X | exp. X, Y_1, Y_2, Y_3)
- Assume contagions are independent
- P(X) is P(infection by X given just exposed to X): can easily be computed empirically by counting

$$P\left(X | \{Y_k\}_{k=1}^{K}\right) = \frac{1}{P(X)^{K-1}} \prod_{k=1}^{K} P(X|Y_k)$$

$$P(X = u_j | Y_k = u_i) \approx \underbrace{P(X = u_j)}_{\substack{\text{Prior infection} \\ \text{prob.}}} + \underbrace{\Delta_{cont.}^{(k)}(u_i, u_j)}_{\substack{\text{Interaction term} \\ \text{(still has } w^2 \text{ entries!)}}}$$

$$\Delta_{cont.}^{(k)}(u_i, u_j) = \sum_t \sum_s \mathbf{M}_{j,t} \cdot \Delta_{clust}^{(k)}(c_t, c_s) \cdot \mathbf{M}_{i,s}$$

- Each contagion u_i has a vector M_i
  - Entry M_{is} models how much u_i belongs to topic s
- $\Delta_{clust}^{(k)}(s, t)$ models the change in infection prob. given that u_i is on topic s and exposure k-steps ago was on topic t

Result of Modeling Interactions

- Maximize the data likelihood

$$\arg\max_{P(x),M,\Delta} \prod_{X\in R} P(X|X,Y_1 \ldots Y_K) \prod_{X\notin R} 1 - P(X|X,Y_1 \ldots Y_K)$$

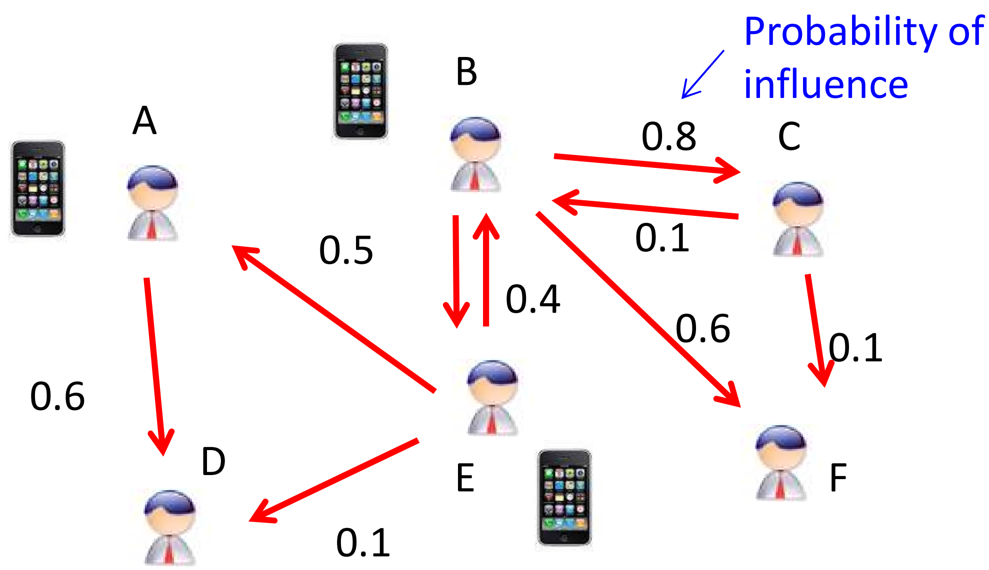
  - R … contagions X that resulted in infections
- Solve using stochastic coordinate ascent
- Compared with baselines
  - Infection Probability (IP): $P\left(X = u_i \middle| Y_k = u_j\right) = P(X = u_i)$
  - IP+Node bias: $P\left(X = u_i \middle| Y_k = u_j\right) = P(X = u_i) + \gamma_n$
  - Exposure Curve (EC): $P\left(X = u_i \middle| Y_k = u_j\right) = P(X|\#times\ exposed\ to\ X)$
- Found that the model with interactions works best
  - 71% of the adoption probability comes from the topic interactions
  - 69% exposures on Twitter come from the network and 29% from external sources

Today's Agenda

- 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

Influence Maximization

- Influence maximization
  - Minimize marketing cost and more generally to maximize profit.
  - E.g., to get a small number of influential users to adopt a new product, and subsequently trigger a large cascade of further adoptions.



[1] 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.

Problem Abstraction

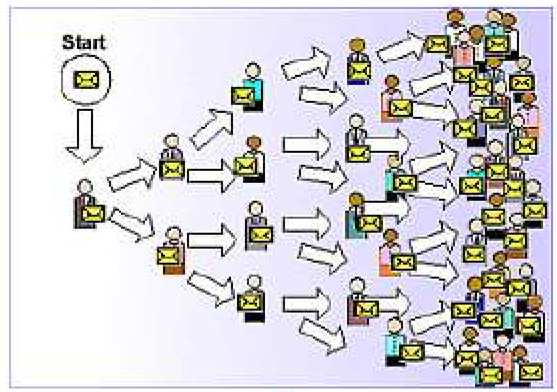
- We associate each user with a status:
  - Active or Inactive
  - The status of the chosen set of users (seed nodes) to market is viewed as active
  - Other users are viewed as inactive
- Influence maximization
  - Initially all users are considered inactive
  - Then the chosen users are activated, who may further influence their friends to be active as well

Why Maximize Influence?

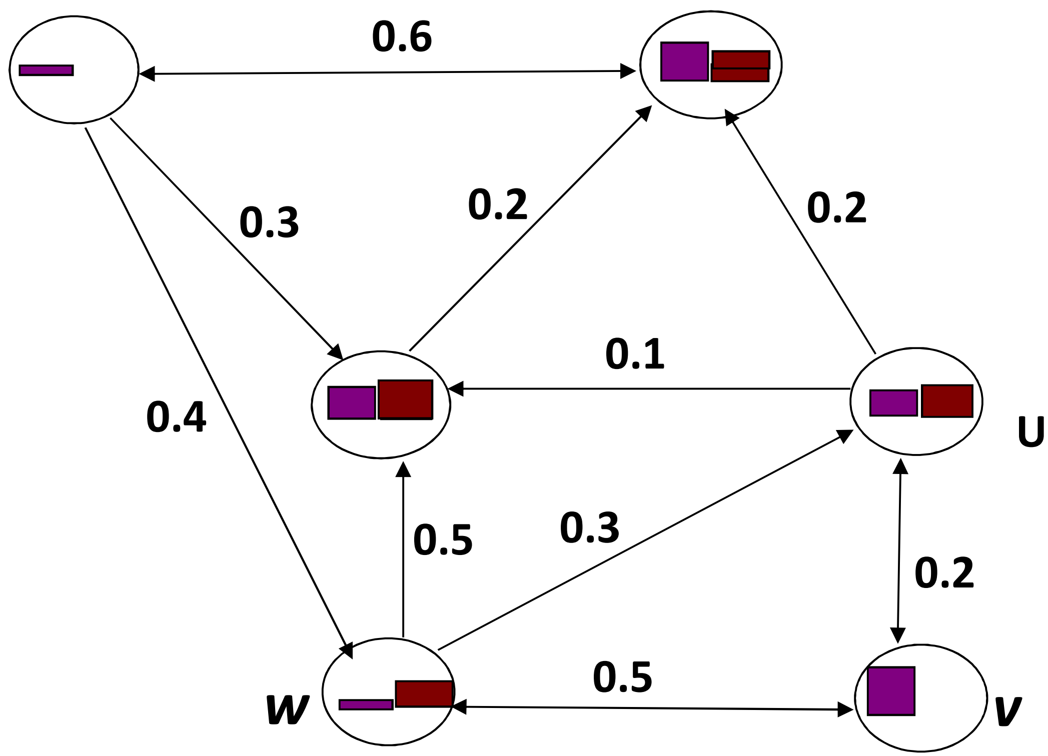
Identify influential customers

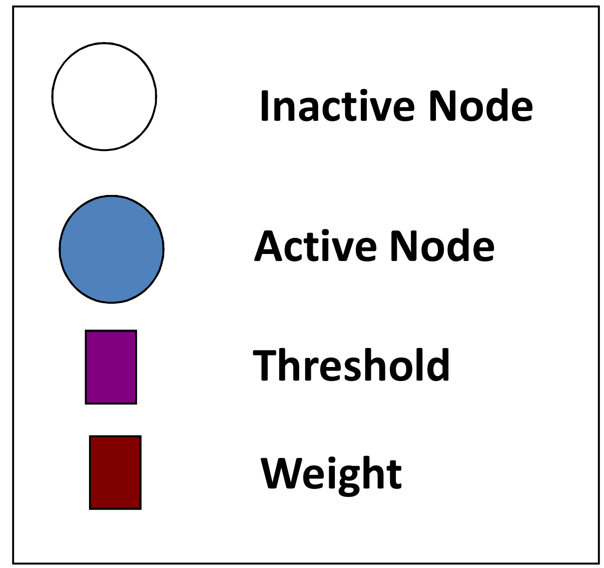
Convince them to adopt the product – Offer discount/free samples

These customers endorse the product among their friends



Independent Cascade Model





Stop!

Source: David Kempe's slides

Influence Maximization Problem

- Define the influence of a set of nodes *A*, denoted by $\sigma(A)$, to be the expected number of active nodes at the end of the process.
- Problem Definition:
  - Given a parameter *k*, find a *k*-node set *A* to maximize $\sigma(A)$.
- Hardness of this problem
  - It is NP-hard to determine the optimum for influence maximization for both independent cascade model and linear threshold model.

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 S
  - Choose an element that provides the largest marginal increase in the function value $\sigma(A)$.
  - Until $|S| = k$

Algorithms

- General Greedy
- Low-distance Heuristic
- High-degree heuristic
- Degree Discount Heuristic

General Greedy

- General idea: 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.

Algorithm 1 GeneralGreedy(G, k)

initialize $S = \emptyset$ and $R = 20000$
for $i = 1$ to k **do**
    for each vertex $v \in V \setminus S$ **do**
        $s_v = 0.$
        for $i = 1$ to R **do**
            $s_v \mathrel{+}= \boxed{|RanCas(S \cup \{v\})|}$
        end for
        $s_v = s_v / R$
    end for
    $S = S \cup \{\arg\max_{v \in V \setminus S} \{s_v\}\}$
end for
output S.

Any random diffusion process

Low-distance Heuristic

- Consider the nodes with the shortest paths to other nodes as seed nodes
- Intuition
  - Individuals are more likely to be influenced by those who are closely related to them.

High-degree heuristic

- Choose the seed nodes according to their degree.
- Intuition
  - The nodes with more neighbors would arguably tend to impose more influence upon its direct neighbors.
  - Known as “degree centrality”

Degree Discount Heuristic[1]

- General idea: 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*
- Specifically, for a node *v* with d_v neighbors of which t_v are selected as seeds, we should discount *v*'s degree by

$$2t_v + (d_v - t_v)\, t_v p$$

where p=0.1.

Algorithm 4 DegreeDiscountIC(G, k)

initialize $S = \emptyset$
for each vertex v **do**
  compute its degree d_v
  $dd_v = d_v$
  initialize t_v to 0
end for
for $i = 1$ to k **do**
  select $u = \arg\max_v \{dd_v \mid v \in V \setminus S\}$
  $S = S \cup \{u\}$
  for each neighbor v of u and $v \in V \setminus S$ **do**
    $t_v = t_v + 1$
    $dd_v = d_v - 2t_v - (d_v - t_v)t_v p$
  end for
end for
output S

[1] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In KDD'09, pages 199-207, 2009.

Other Interesting Problems

- Tracing the flow of sentiments over cascades
- Detecting contamination outbreaks
  - Given a real city water distribution network and data on how contaminants spread in the network, detect the contaminant as quickly as possible
  - Which blogs should one read to detect cascades as effectively as possible?
  - Given a dynamic process spreading over a network, we want to select a set of nodes to detect the process effectively
- Social Influence Analysis for Heterogeneous Networks

Today's Agenda

- 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**

Application: Social Advertising[1]

- Conducted two very large field experiments that identify the effect of social cues on consumer responses to ads on Facebook
- Exp. 1: measure how responses increase as a function of the number of cues.
- Exp. 2: examines the effect of augmenting traditional ad units with a minimal social cue
- Result: Social influence causes significant increases in ad performance

[1] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn. Social influence in social advertising: evidence from field experiments. In EC'12, pages 146-161, 2012.

Application: Opinion Leader Finding[1]

- Propose viral marketing through frequent pattern mining.
- Assumption
  - Users can see their friends actions.
- Basic formation of the problem
  - Actions take place in different time steps, and the actions which come up later could be influenced by the earlier taken actions.
- Approach
  - Define leaders as people who can influence a sufficient number of people in the network with their actions for a long enough period of time.
  - Finding leaders in a social network makes use of action logs.

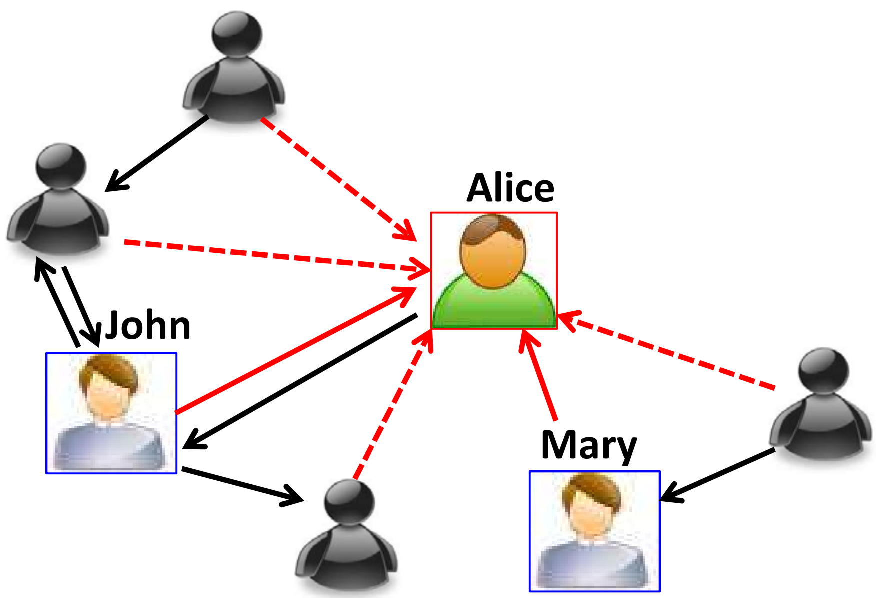
[1] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In CIKM'08, pages 499–508, 2008.

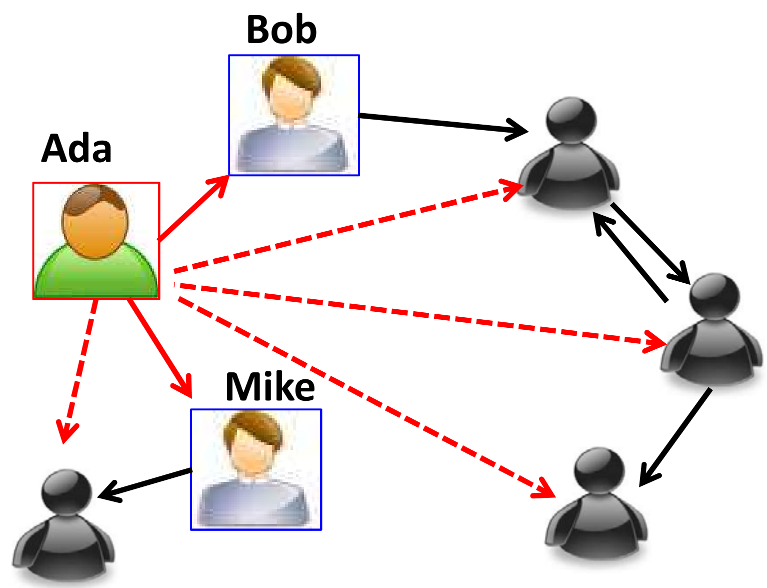
Application: Influential Blog Discovery[1]

- Influential Blog Discovery
  - In the web 2.0 era, people spend a significant amount of time on user-generated content web sites, like blog sites.
  - Opinion leaders bring in new information, ideas, and opinions, and disseminate them down to the masses.
- Four properties for each bloggers
  - **Recognition**: A lot of inlinks to the article.
  - **Activity generation**: A large number of comments indicates that the blog is influential.
  - **Novelty**: with less outgoing links.
  - **Eloquence**: Longer articles tend to be more eloquent, and can thus be more influential.

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In WSDM’08, pages 207–217, 2008.

Applications: Influence Maximization and Friend Recommendation

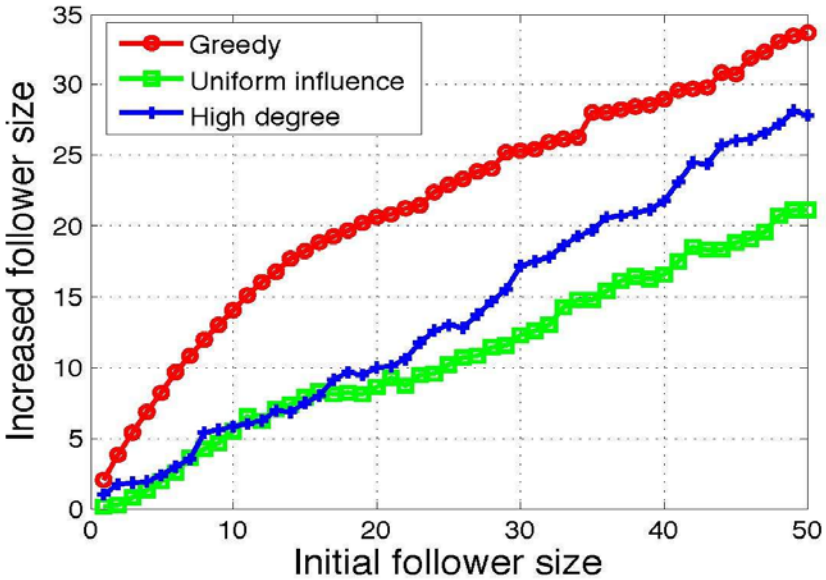




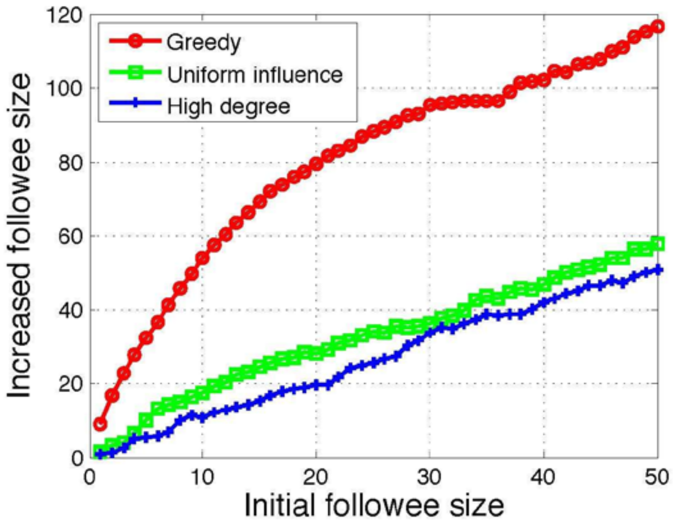
Find a set *S* of *k* initial followers to follow user *v* such that the number of newly activated users to follow *v* is maximized.

Find a set *S* of *k* initial followees for user *v* such that the total number of new followees accepted by *v* is maximized

Effect of Influence in Social Recommendations



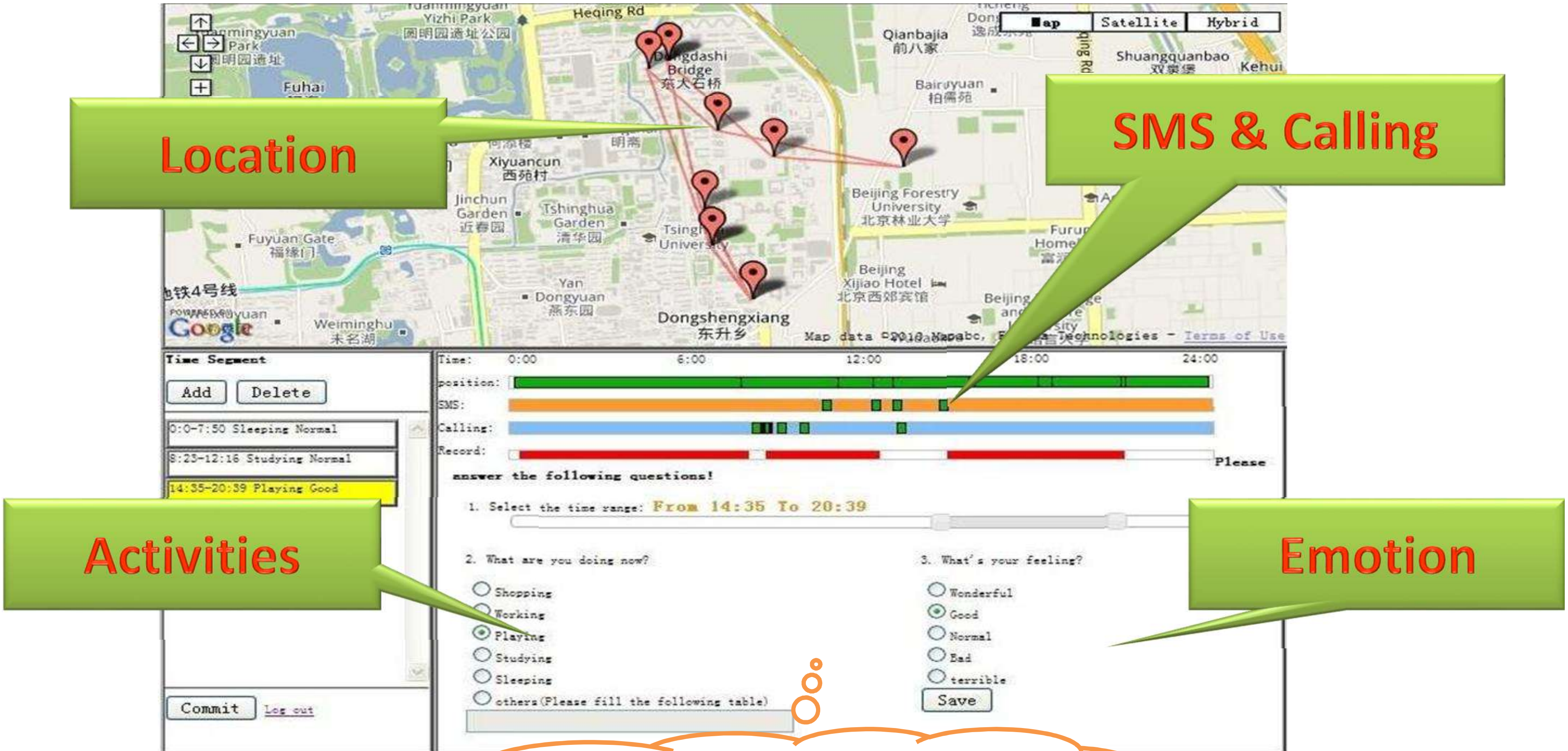
Influence Maximization



Recommendation

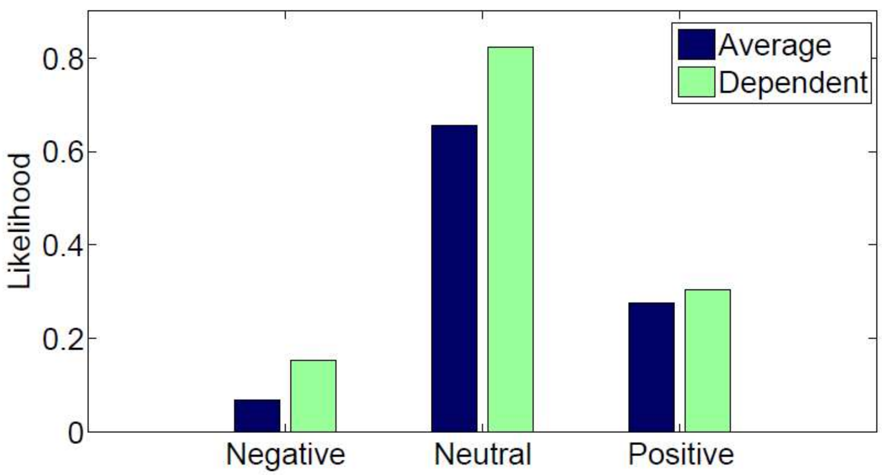
- High degree
  - May select the users that do not have large influence on following behaviors.
- Uniform configured influence
  - Can not accurately reflect the correlations between following behaviors.
- Greedy algorithm based on the influence probabilities
  - Captures the entire features of three users in a triad (i.e., triad structures and triad statuses)

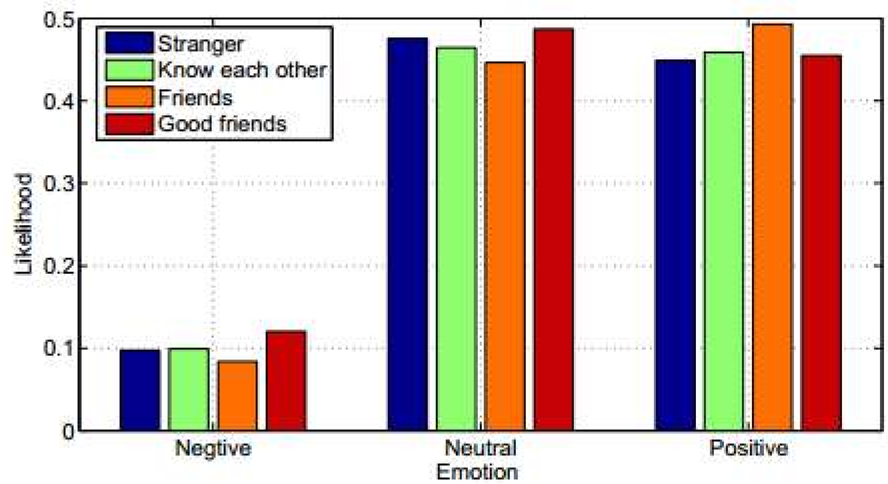
Emotional Influence Applications



Can we predict users' emotion?

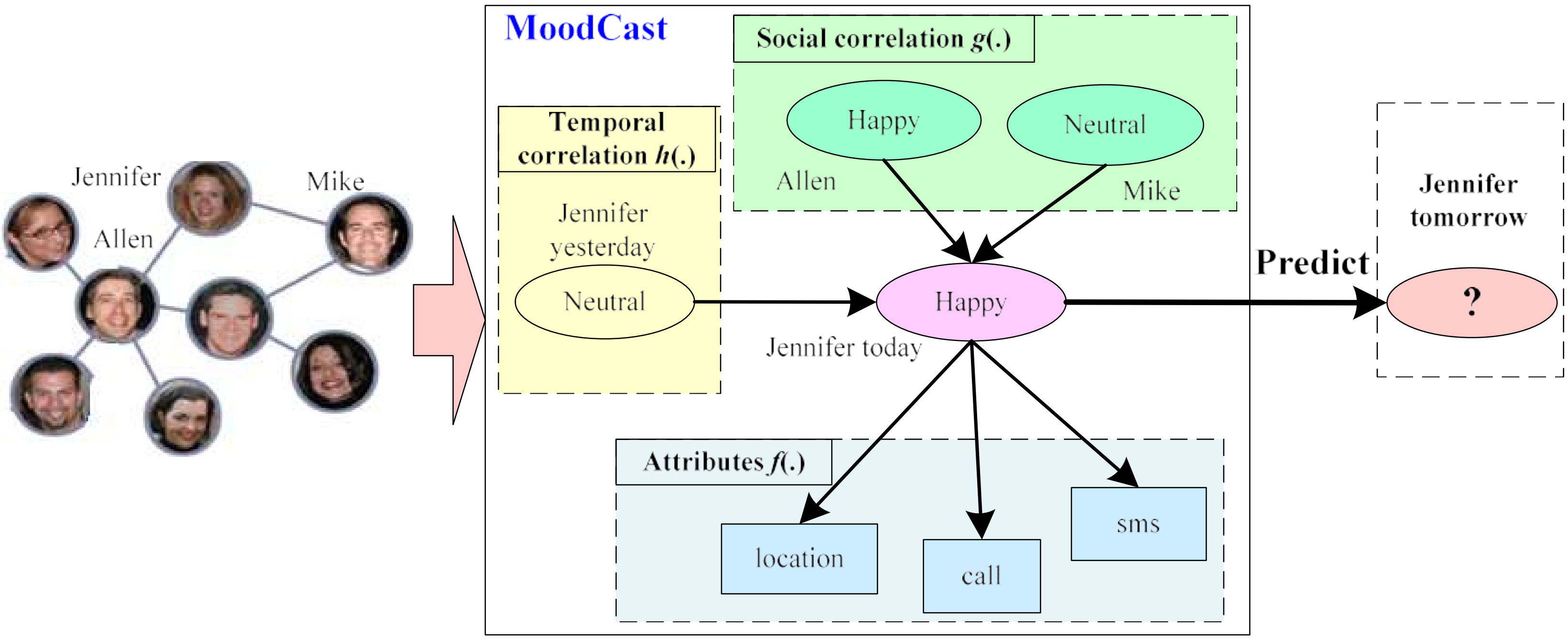
The Social Factor in Emotional Influence





Social correlation

MoodCast: Dynamic Continuous Factor Graph Model



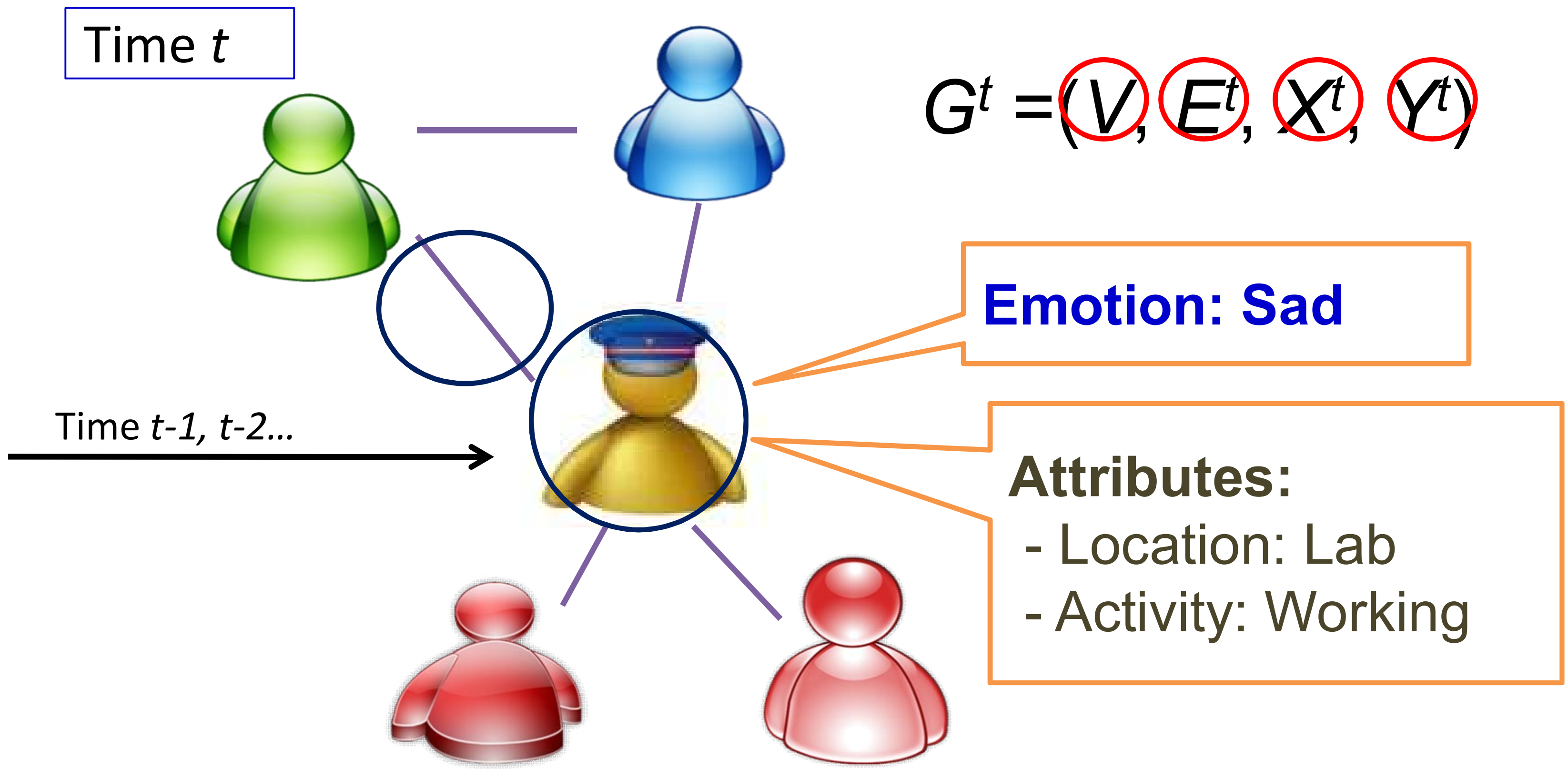
Our solution

1. We directly define continuous feature function;

2. Use Metropolis-Hasting algorithm to learn the factor graph model.

Problem Formulation

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

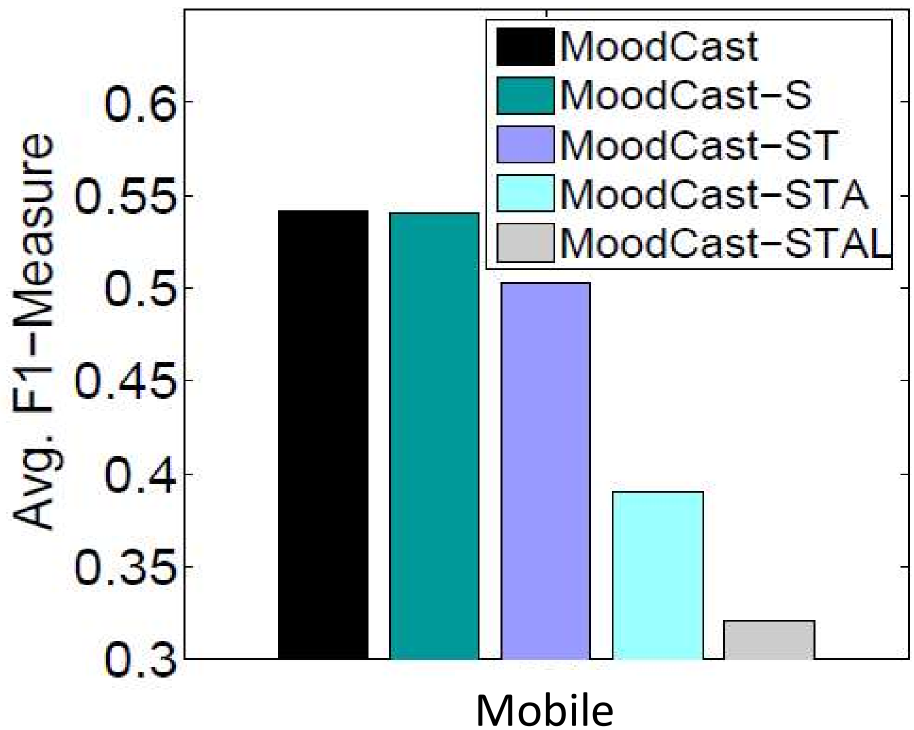


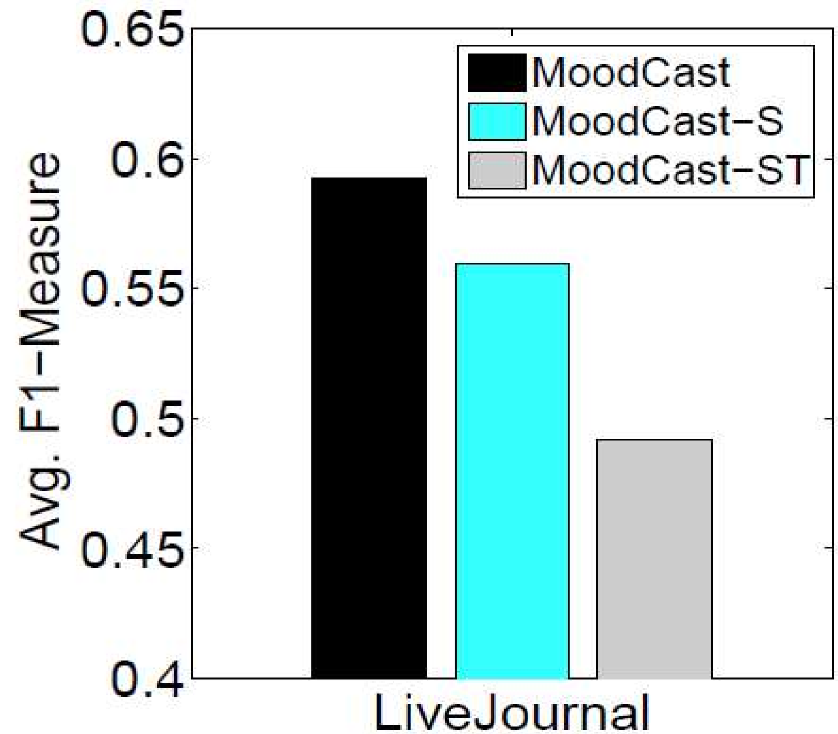
Learning Task:

$$f(V, E^{(t+1)}, X^{(t+1)} | G^t) \rightarrow Y^{(t+1)}$$

Factor Contributions

- All factors are important for predicting user emotions





MoodCastF stands for ignoring friend influence factor. MoodCastFP stands ignoring both the friend and time-dependent factor. MoodCast-FPA stands for further ignoring activity attribute and MoodCast-FPAL for further ignoring location attribute.

Take-away Messages

- 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
- Studying Social Influence Analysis is important for
  - Social advertising
  - Opinion leader finding
  - Social recommendation
  - Emotion analysis

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/

Preview of Lecture 13: Analysis of Microblogs (Part 1)

- Event Detection on Twitter

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!

References: Decision Based Methods

- S. Morris. Contagion. Review of Economic Studies 67, 57-78, 2000.
- N. Immorlica, J. Kleinberg, M. Mahdian, T. Wexler. The Role of Compatibility in the Diffusion of Technologies Through Social Networks. In Proc. ACM EC, 2007.
- E. Berger. Dynamic Monopolies of Constant Size. Journal of Combinatorial Theory Series B 83, 191-200, 2001.
- D. Centola, M. Macy. Complex Contagions and the Weakness of Long Ties. American Journal of Sociology, 2007.
- M. Jackson, L. Yariv. Diffusion of Behavior and Equilibrium Properties in Network Games. American Economic Review , Vol 97, No. 2, 2007.
- S. Bikhchandani, D. Hirshleifer, I. Welch. A theory of fads, fashion, custom and cultural change as information cascades. Journal of Political Economy. Vol. 100, pp. 992-1026, 1992.
- T. Schelling. Micromotives and Macrobehavior. Norton, 1978.
- D. Strang, S. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. Annual Review of Sociology, 24:265--290, 1998.
- D. Watts. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci., vol. 99 no. 9, 5766-5771, 2002.
- H. P. Young. The Diffusion of Innovations in Social Networks. Santa Fe Institute Working Paper 02-04-018.
- P. Dodds and D. J. Watts. Universal Behavior in a Generalized Model of Contagion. Phyical Review Letters, 2004.
- E. Lieberman, C. Hauert, M. A. Nowak. Evolutionary Dynamics on Graphs. Nature 433: 312-316, 2005.
- D. Centola, M. Macy, V. Eguiluz. Cascade Dynamics of Multiplex Propagation. Physica A 374, 449-456, 2007.
- D. Centola. The Spread of Behavior in an Online Social Network Experiment. Science, 2010.
- M. Granovetter. Threshold models of collective behavior. American Journal of Sociology 83(6):1420-1443, 1978.
- A. V. Banerjee. A Simple Model of Herd Behavior. The Quarterly Journal of Economics, Vol. 107, No. 3, pp. 797-817, 1992.

References: Probabilistic Methods

- S. Myers, C. Zhu, J. Leskovec. Information Diffusion and External In?uence in Networks. In Proc. KDD, 2012.
- A. Goyal, F. Bonchi, L.V.S. Lakshmanan. Learning influence probabilities in social networks. In Proc. WSDM, 2010.
- D. Cosley, D. Huttenlocher, J. Kleinberg, X. Lan, S. Suri.Sequential Influence Models in Social Networks. In Proc. ICWSM, 2010.
- J. Leskovec, A. Singh, J. Kleinberg. Patterns of Influence in a Recommendation Network In Proc PAKDD, 2006.
- L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In Proc. KDD, 2006.
- E. Adar, L. Adamic. Tracking information epidemics in blogspace. In Proc. Wed intelligence, 2005.
- J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, M. Hurst. Cascading Behavior in Large Blog Graphs. In Proc. SIAM International Conference on Data Mining, 2007.
- D. Gruhl, R. Guha, D. Liben-Nowell, A. Tomkins. Information Diffusion through Blogspace. In Proc. International WWW Conference, 2004.
- M. Goetz, J. Leskovec, M. Mcglohon, C. Faloutsos. Modeling blog dynamics, In AAAI Conference on Weblogs and Social Media (ICWSM), 2009.
- M. Miller, C. Sathi, D. Wiesenthal, J. Leskovec, C. Potts. Sentiment Flow Through Hyperlink Networks. In Proc. ICWSM, 2011.
- J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg. Structural Diversity in Social Contagion. In Proc. National Academy of Sciences, 2012.

References: Influence Maximization

- M. Richardson, P. Domingos. Mining Knowledge-Sharing Sites for Viral Marketing. In Proc. KDD, 2002.
- M. Richardson, P. Domingos. Mining the Network Value of Customers. In Proc. KDD, 2001.
- S. Hill, F. Provost, C. Volinsky. Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. Statistical Sciece, 2006.
- J. Leskovec, L. Adamic, B. Huberman. The Dynamics of Viral Marketing. TWEB, 2007.
- E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, Everyone's an in?uencer: quantifying in?uence on twitter. In Proc. WSDM, 2011.
- S. Aral, L. Muchnik, A. Sundararajan. Distinguishing in?uence-based contagion from homophily-driven diffusion in dynamic networks. In Proc. National Academy of Sciences, 2009.
- A. Goyal, W. Lu, L. S.V. Lakshmanan. SIMPATH: An Ef?cient Algorithm for In?uence Maximization under the Linear Threshold Model. In Proc. ICDM, 2011.
- W. Chen, Y. Wang, S. Yang. Efficient Influence Maximization in Social Networks. In Proc. KDD, 2009.
- W. Chen, Y. Yuan, L. Zhang. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. In Proc. ICDM, 2010.
- C. Lerman, R. Ghosh. Information Contagion: Empirical Study of the Spread of News on Digg and Twitter Social Networks. In Proc. ICWSM, 2010.
- Y. Singer. How to win friends and influence people, truthfully: Influence maximization mechanisms for social networks. In Proc. WSDM, 2012.