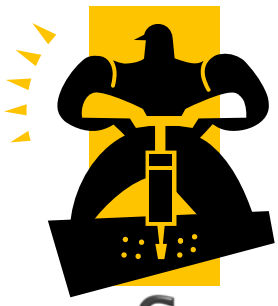


Social Network Analysis -- Social Influence

Instructor: Jen-Wei Huang

Office: 92528 in the EE building
jwhuang@mail.ncku

Outline

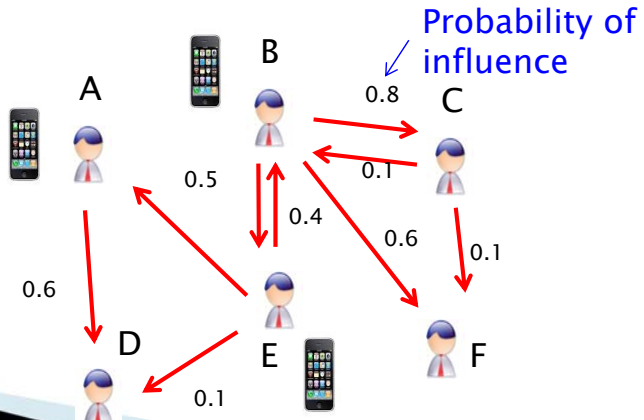
- Introduction
- Viral Marketing
- Influential model
- Influence maximization problem
- Identifying influential spreaders
- Detecting popular topics
- Six degrees of separation & weak ties

Introduction

- Social network plays a fundamental role as a medium for the spread of information among members.
  - –E.g., influence, opinion, ideas, innovation

Viral marketing

- Direct Marketing takes the "word-of-mouth" effects to significantly increase profits.
- Minimize marketing cost and more generally maximize profit.
- To get a small number of influential users to adopt a new product, and subsequently trigger a large cascade of further adoptions.



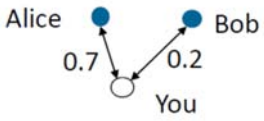
Influential Model

- Social influence problem usually adopt following models.
  1. Linear Threshold Model
  2. Independent Cascade Model
  3. Heat Diffusion Model

Linear Threshold Model

- A node v is influenced by each neighbor w according to a weight $b_{v,w}$ such that

$$\sum_{w\ neighbor\ of\ v} b_{v,w} \leq 1$$

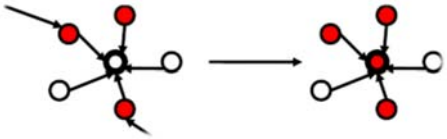


- Each node v has a loading threshold θv
  - Can be chosen uniformly at random
  - Can be proportional to the initial load

Linear Threshold Model

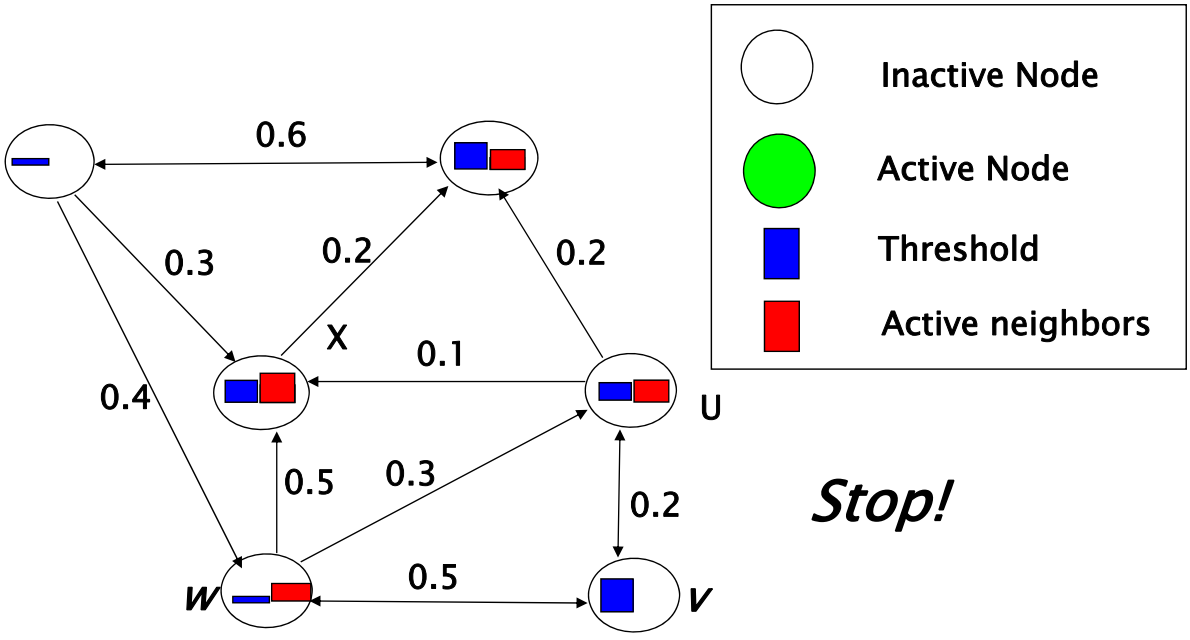
- A node v becomes active/overheat if

$$\sum_{w \text{ neighbor of } v} b_{v,w} \geq \theta_v$$



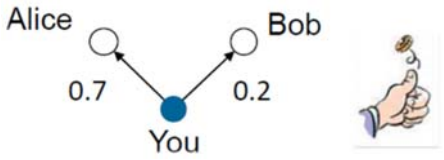
- Continue until no more activations are possible

Linear Threshold Model



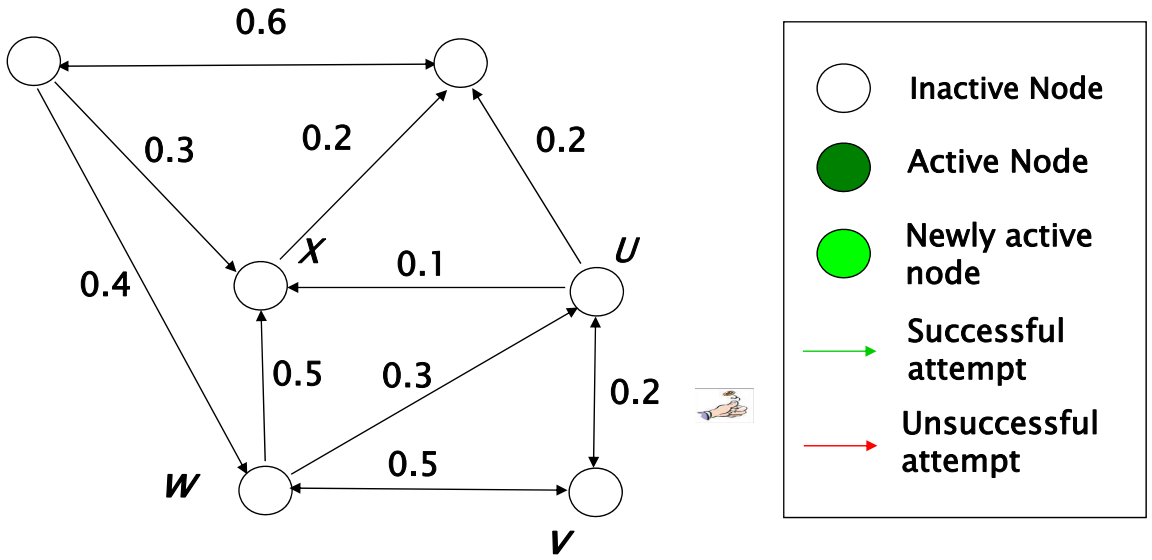
Independent Cascade Model

- When node v becomes active, it has a single chance of activating each currently inactive neighbor w.
- The activation attempt succeeds with probability $p_{v,w}$



- Run until no more activations are possible

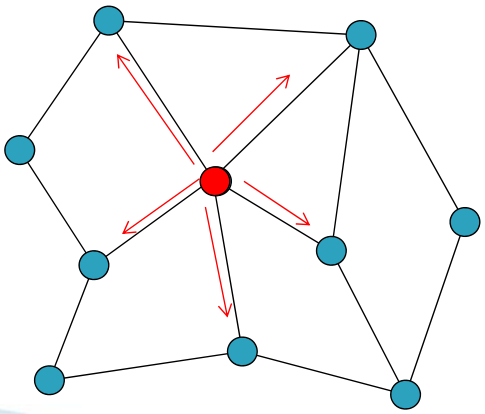
Independent Cascade Model



Stop!

Heat Diffusion Model

- The influence will spread as similar as heat and lead to heat balance.
- There are no active and inactive states in this model but the model take effect of time into account.



Heat Diffusion Model

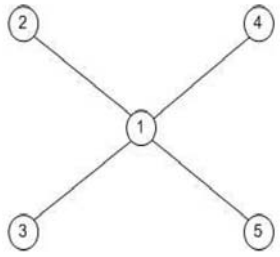
$$H_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \text{ or } (v_j, v_i) \in E \\ -d(v_i), & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{f}(t) = e^{\alpha t \mathbf{H}} \mathbf{f}(0)$$

α	Thermal conductivity
$d(v_i)$	Degree of node i
$f_i(t)$	Heat value of node i at time t
$\mathbf{f}(0)$	Vector of the initial heat distribution
$\mathbf{f}(t)$	Vector of the heat distribution at time t

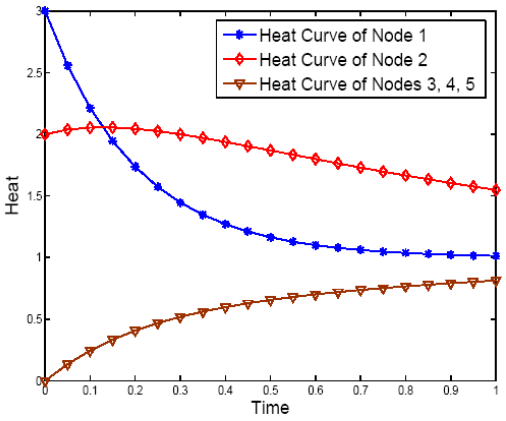
Heat Diffusion Model

At time 0, suppose node 1 is given 3 units of heat, and node 2 is given 2 units of heat.



$$\mathbf{f}(0) = [3, 2, 0, 0, 0]^{\mathbf{T}}$$

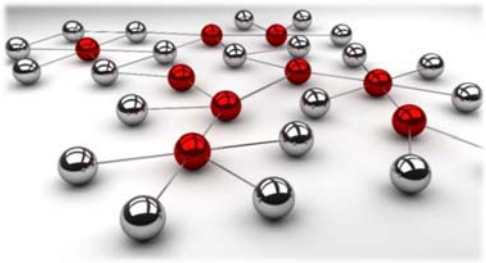
$$\mathbf{H} = \begin{pmatrix} -4 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix}$$



Curve of heat change with time

Influence Maximization Problem

- Influence of a set of nodes S: f(S)
  Given a set of initial nodes S, the expected number of infected nodes at the end.



- Problem
  Given a constant k, find a k-node set S to *maximize f(S)*

Approaches

- General Greedy
  - General idea: In each round, adding one vertex into the selected set S to maximize the influence spread.

- 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 (Know as “degree centrality”)
  - 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.

Identifying Influential Spreaders

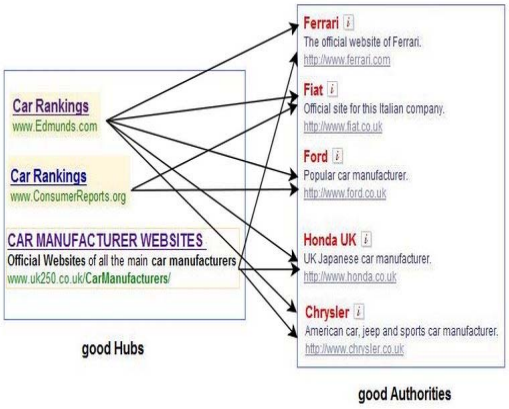
- Identifying the most influential spreaders in a network is critical for ensuring efficient diffusion of information.
- A social media campaign can be optimized by targeting influential individuals who can trigger large cascades of further adoption.

IP Algorithm

- IP (Influence-passivity) algorithm adopts the concept of HITS algorithm
  - Influence score depends on :
    - how many people you influenced as well as passivity.
    - how dedicated the people you influenced are.
  - Passivity score depends on :
    - how many people she's exposed to but is not influenced by.
    - how much she rejects other user's influence.

HITS Algorithm

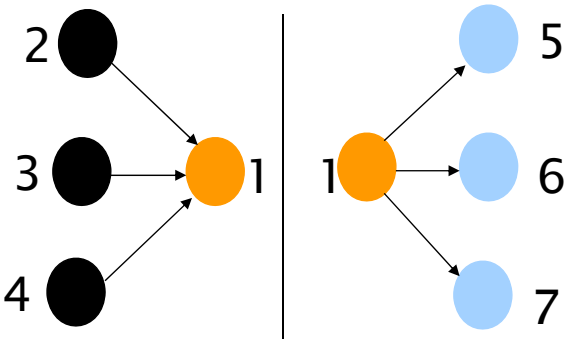
- HITS (Hyperlink-Induced Topic Search) is a link analysis algorithm that rates web pages.
- Two categories of web pages:
  - Authorities : Pages contain valuable information about the subject you queried.
  - Hubs : Hubs contain useful links toward authoritative pages.



query: top car makers

HITS Algorithm

- Each web pages contain two values:
  - (authority and hub).
- Recursive dependency:
  - auth(v) ← Σ hub(w)
  - hub(v) ← Σ auth(w)



a(1) = h(2) + h(3) + h(4)

h(1) = a(5) + a(6) + a(7)

IP Algorithm

$$\text{hub}(p) = \sum_{i=1}^{n} \text{auth}(i)$$

$$\text{auth}(p) = \sum_{i=1}^{n} \text{hub}(i)$$

Acceptance rate

$$I_i \leftarrow \sum_{j:(i,j)\in E} u_{ij} P_j$$

$$P_i \leftarrow \sum_{j:(j,i)\in E} v_{ji} I_j$$

Rejection rate

Acceptance rate $u_{ij} = \dfrac{w_{i,j}}{\sum_{k:(k,j)\in E} w_{kj}}$

Rejection rate $v_{ji} = \dfrac{1 - w_{ji}}{\sum_{k:(j,k)\in E} (1 - w_{jk})}$

Detecting Popular Topics

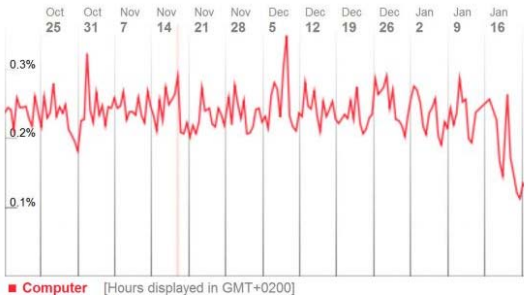
- Providing a global view of topics that are popular over time or will become popular in the future.

Definition of Topics

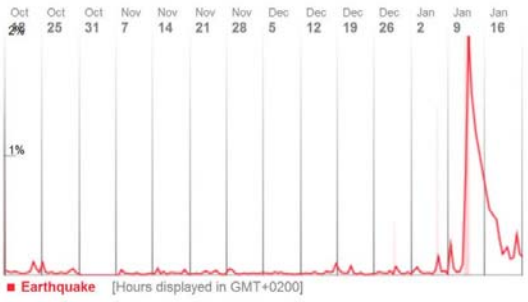
- The definition of a topic is :
  "A seminal event or activity, along with all directly related events and activities."
- The followings are some types of topic:
  - Elections
  - Legal/Criminal cases
  - Natural Disasters
  - New Law
  - Sports
  - Finance
  - War
  - Accidents(e.g. plane, car)

How to do?

- Term frequency



Usage of the term "computer" in Twitter from October 2009 to January 2010.

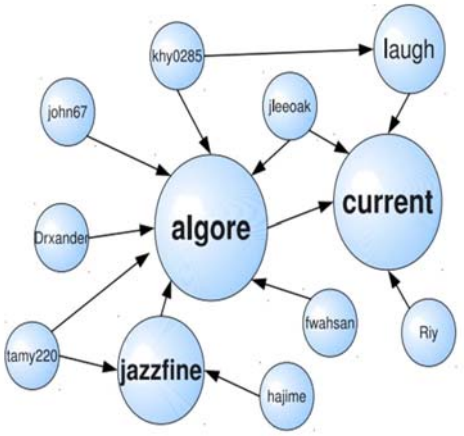


Usage of the term "earthquake" in Twitter from October 2009 to January 2010.

- If a topic has been extensively discussed but rarely before and after, it can be called emergent.

How to do?

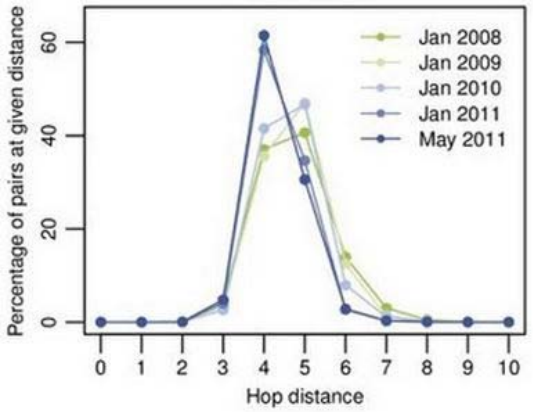
- Social relationship



- Follower on Twitter

  The size of the nodes highlights their importance in the considered community.

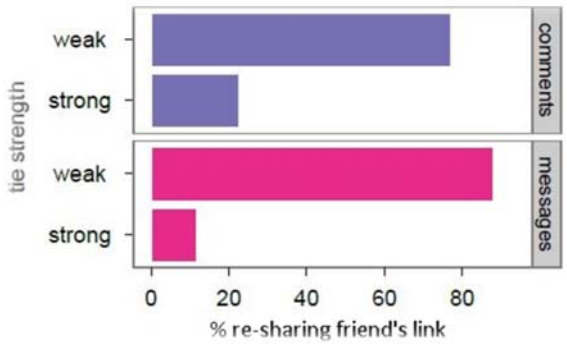
Six Degrees of Separation



- Six degrees of separation: Everyone is six or fewer steps away from any other person in the world.
- Because of the appearance of Facebook, the world become smaller.
- The research shows 99.6% just need five steps, moreover, 92% just need four!

- M. S. Granovetter, “The Strength of Weak Ties.”

→Finding a job through contacts:
Frequently = at least twice a week (16.7%)
Occasionally = more than once a year (55.6%)
Rarely = once a year or less (27.8%)



- Weak ties are argued to have access to more diverse information because they are expected to have fewer mutual contacts.

References

- Slides from Prof. S.-D. Lin, NTU
- P. Domingos and M. Richardson. “Mining the network value of customers.” In Proceedings of the seventh ACM SIGKDD (KDD’01), pages 57-66, 2001.
- A. Guille, H. Hacid, C. Favre and D. A. Zighed. “Information diffusion in online social networks: A survey.” SIGMOD, 2013.
- M. Cataldi, L. D. Caro and C. Schifanella, “Emerging topic detection on Twitter based on temporal and social terms evaluation.” MDMKDD, 2010.
- M. S. Granovetter, “The Strength of Weak Ties,” AJS, 1973.
- E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, “The role of social networks in information diffusion”, WWW, 2012.
- D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, “Influence and passivity in social media.” ECML/PKDD, 2011.
- J.Allan, editor. “Topic detection and tracking: event-based information organization.” Kluwer Academic Publishers, Norwell, MA, USA, 2002.