



Data Mining -- Social Network Analysis

Instructor: Jen-Wei Huang

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

Outline

- ▶ Social influence analysis
- ▶ Recommendation
- ▶ Link analysis
- ▶ Link prediction
- ▶ Community detection
- ▶ Anomaly detection
- ▶ Compressing social networks

Social Influence Analysis

- ▶ Social influence occurs when one's emotions, opinions, or behaviors are affected by others.



Social influential network

- A social network is modeled as a graph $G = \{V, E\}$, where V is the set of nodes, and E is the set of edges.
- social influence is a directional effect from node A to node B.

Social Influence Analysis

- ▶ Modeling the influence diffusion
- ▶ Influence maximization problem

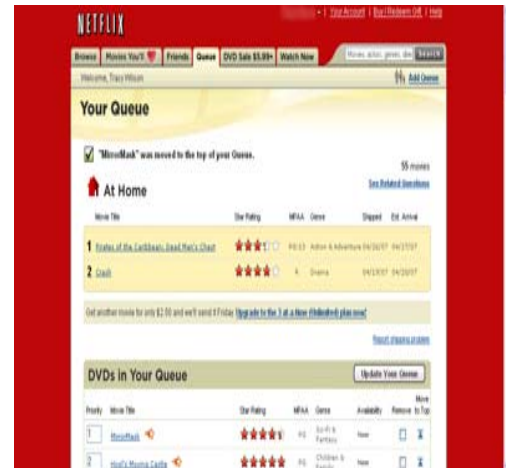
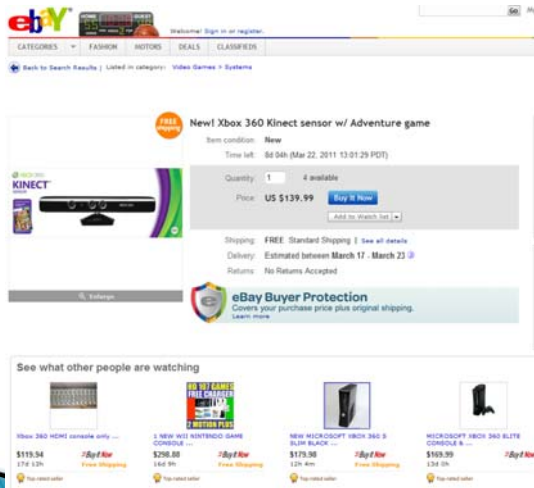


- Top influential nodes
 - Some nodes have intrinsically higher influence than others.



Social Recommendation

- ▶ Recommendation system is an information filtering system that seeks to predict the 'rating' or 'preference'.

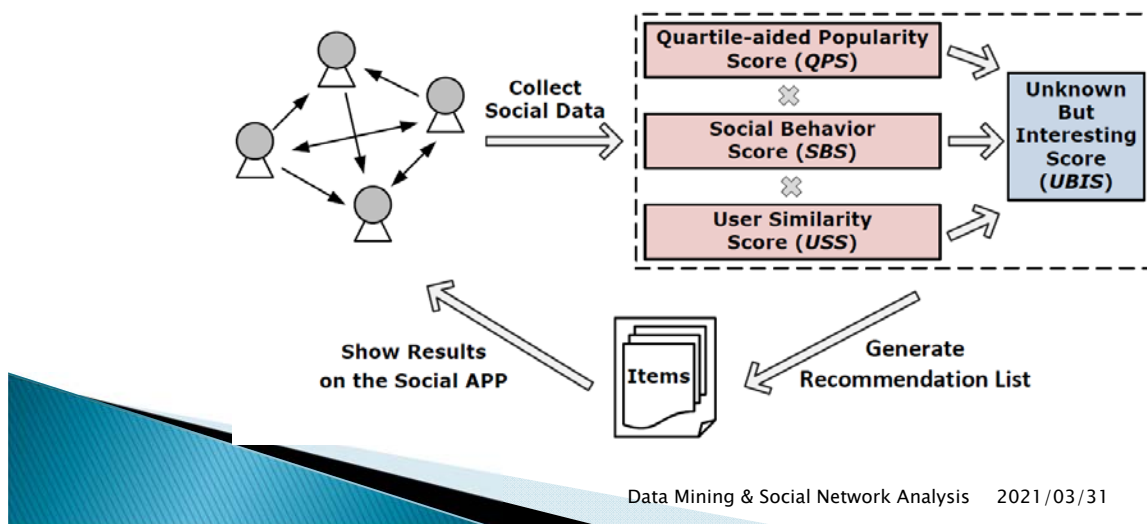


Social Recommendation

- ▶ Collaborative filtering
 - Building a model from a user's past behavior and interaction with others/friends.
 - Using that model to predict items that the user may have an interest in.
- ▶ Content-based filtering
 - Utilizing a series of discrete characteristics of an item in order to recommend additional items with similar properties.

Social Recommendation

- ▶ Social recommendation
 - Considering the user's social information in the social network such as relationships between users or social behavior of users to recommend items



Data Mining & Social Network Analysis 2021/03/31

7

Link Analysis

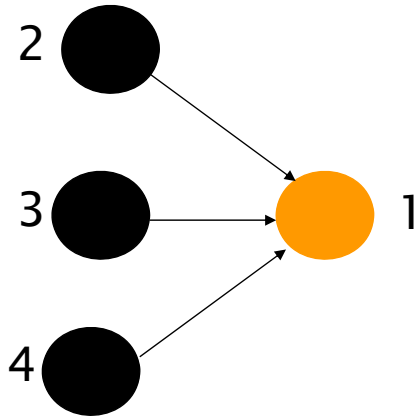
- ▶ Link analysis is to evaluate relationships (connections) between nodes.
- ▶ Relationships may be identified among various types of nodes, including organizations, people and transactions
- ▶ Applications of the link analysis
 - Web search
 - HITS and Page Rank

Data Mining & Social Network Analysis 2021/03/31

8

Link Analysis

- ▶ Hubs and Authorities:
 - HITS – Hypertext Induced Topic Selection

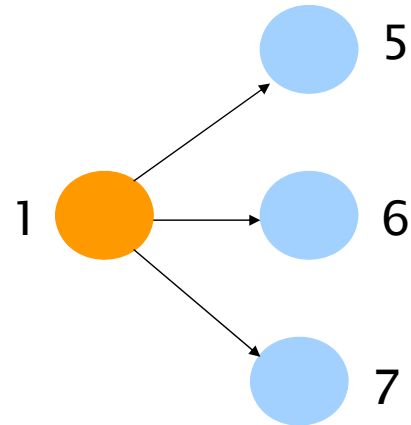


$$a(1) = h(2) + h(3) + h(4)$$

- Recursive dependency:

$$a(v) \leftarrow \sum h(w)$$

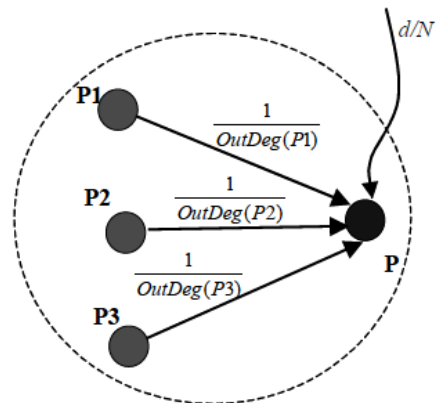
$$h(v) \leftarrow \sum a(w)$$



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

Link analysis

- ▶ PageRank
 - PageRank is a metric for ranking hypertext documents that determines their quality

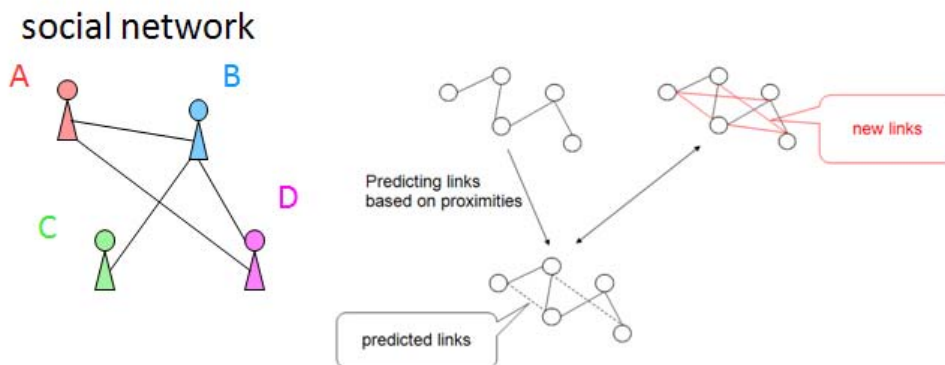


- N is the number of nodes in the graph
- OutDeg (q) is the number of hyperlinks on page q
- d is the probability that a random surfer chooses a URL directly

$$PR(P) = \frac{d}{N} + (1 - d) \left(\frac{PR(P1)}{OutDeg(P1)} + \frac{PR(P2)}{OutDeg(P2)} + \frac{PR(P3)}{OutDeg(P3)} \right)$$

Link Prediction

- ▶ Given a snapshot of a social network at time t , we seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time t' .



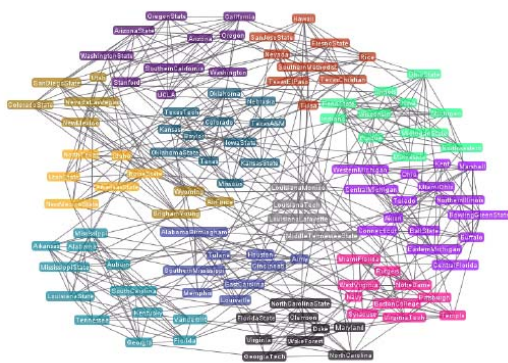
Community Detection

- ▶ Community structure
 - Has strong internal node–node connections
 - Weaker external connections

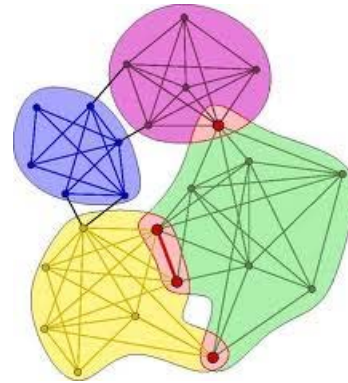


Community Detection

- ▶ **Community detection** in a network is the gathering of network vertices into groups in such a way that nodes in each group are **densely connected inside** and **sparser outside**.
- ▶ Overlapped community detection
 - One node may belong to several different communities



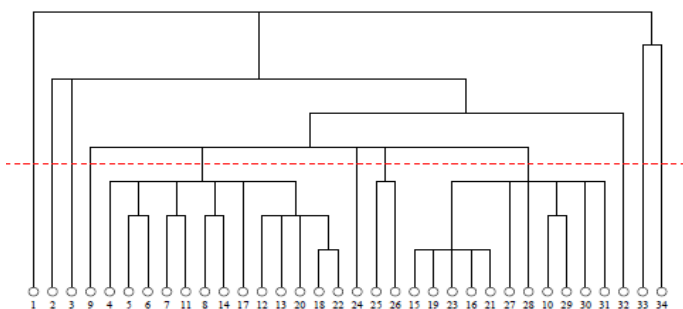
Non-overlapped Communities



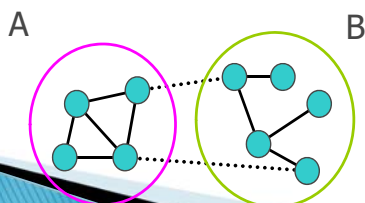
overlapped communities

Community Detection

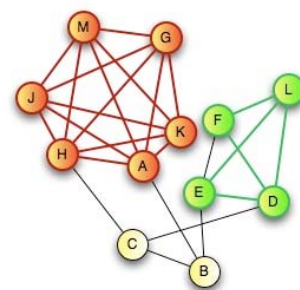
- Hierarchical Clustering techniques



- GN- Edge Betweenness



- Clique-based



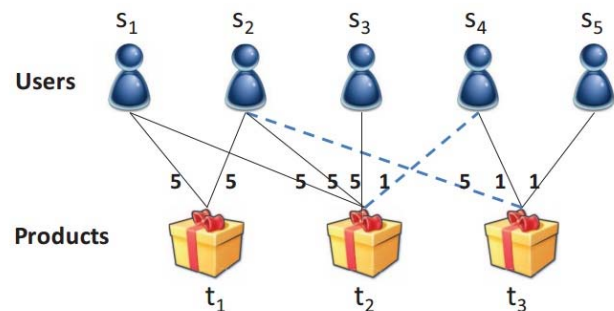
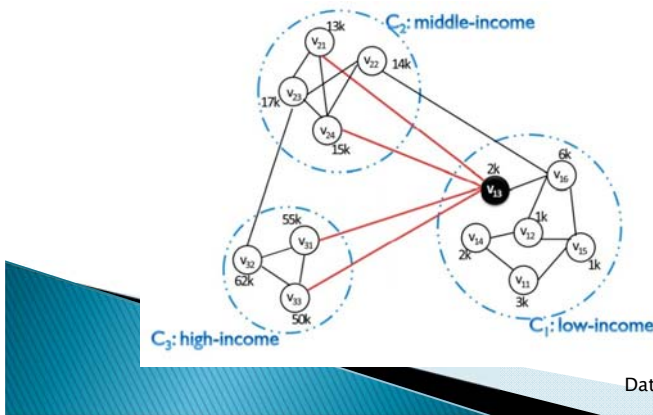
- Newman Fast Algorithm

Optimizing Modularity

$$Q = \sum_i (e_{ii} - a_i^2)$$

Anomaly Detection

- ▶ Detecting nodes in a social network that do not conform to an established normal behavior.
- ▶ There are several definitions of anomaly nodes
 - The nodes have a higher density of external links, compared to its internal links.
 - In a bipartite graph, anomalous nodes are the minority and are inconsistent with the rest of the nodes in the same partite



Data Mining & Social Network Analysis 2021/03/31

15

Compressing Social Network

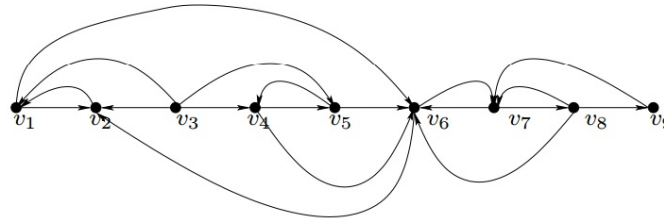
- ▶ Studying the compressibility of a social network is akin to studying the degree of “randomness” in the social network.
 - Web graph (Web pages are nodes, hyperlinks are directed edges) is highly compressible.
- ▶ Motivations
 - Web properties require high-speed indexes for serving adjacencies in the social network.
 - There is a wealth of evidence that social networks are not random graphs in the usual sense.

Data Mining & Social Network Analysis 2021/03/31

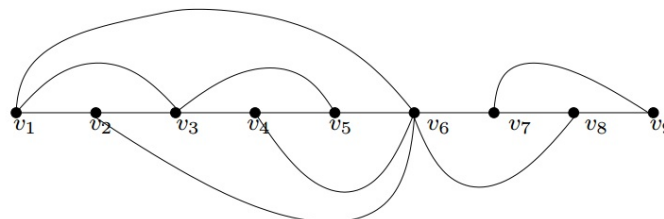
16

Compressing Social Network

- Compressing social network for neighbor query effectively.



a) Graph G_2



Compressing example

Reference

- ▶ David Easley and Jon Kleinberg, "Networks, Crowds, and Markets: Reasoning About a Highly Connected World", Cambridge University Press, pp. 397–435, 2010.
- ▶ Page, Larry, "PageRank: Bringing Order to the Web", Stanford Digital Library Project, talk, 1997.
- ▶ Jie Tang, Jimeng Sun, Chi Wang and Zi Yang, "Social Influence Analysis in Large-scale Networks," Proceedings of the 15th ACM SIGKDD, pp807–816, 2009.
- ▶ M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proceedings of National Academy of Sciences USA 99 , pp. 7821–7826, 2002.
- ▶ V.D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, 2008.
- ▶ M. E. J. Newman, "Detecting community structure in networks," Eur. Phys. J. B 38, pp. 321–330, 2004.

Reference

- ▶ S. Fortunato, "Community detection in graphs," Phys. Rep. 486, pp. 75–174, 2010.
- ▶ Juang-Lin Duan, Shashi Prasad, Jen-Wei Huang, "Discovering Unknown But Interesting Items on Personal Social Network," In Proceedings of PAKDD, pp. 145–156, 2012.
- ▶ Hanbo Dai, Feida Zhu, Ee-Peng Lim, HweeHwa Pang, "Detecting Anomalies in Bipartite Graphs with Mutual Dependency Principles," In Proceedings of ICDM, pp. 171–180, 2012.
- ▶ Tengfei Ji, Jun Gao, and Dongqing Yang, "A Scalable Algorithm for Detecting Community Outliers in Social Networks," In Proceedings of WAIM, pp. 434–445, 2012.
- ▶ Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, Michael Mitzenmacher, Alessandro Panconesi, and Prabhakar Raghavan, "On compressing social networks," In Proceedings of the 15th ACM SIGKDD, pp. 219–228, 2009.
- ▶ Hossein Maserrat and Jian Pei, "Neighbor query friendly compression of social networks," In Proceedings of the 16th ACM SIGKDD, pp. 533–542, 2010.