Computational Network Science: Models, Algorithms and Applications

Dr. Ramasuri Narayanam
Senior Research Scientist
Adobe Research - India
Email: rnarayanam@adobe.com

19-December-2023

Outline of the Presentation

1. **Introduction to Social Networks**
2. Key Tasks in Social Network Analysis
3. Emerging Challenges
4. Summary and Conclusions

Social Networks: Introduction

Recently there is a significant interest from research community to study social networks since:

- Such networks are fundamentally different from technological networks
- Networks are powerful primitives to model several real world scenarios such as interactions among individuals/objects

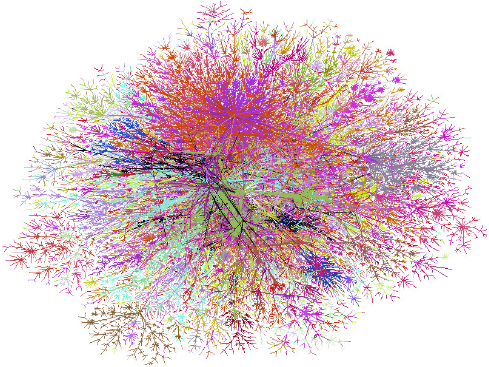
Social Networks: Introduction (Cont.)

Social networks are ubiquitous and have many applications:

- For targeted advertising
- Monetizing user activities on on-line communities
- Job finding through personal contacts
- Predicting future events
- E-commerce and e-business
- For Propagating trusts in web communities
- ...

M.S. Granovetter. The Strength of Weak Ties. American Journal of Sociology, 1973.

Example 1: Web Graph

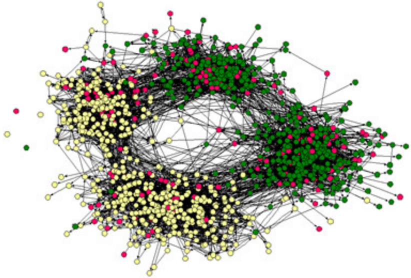


Nodes: Static web pages
Edges: Hyper-links

Reference: Prabhakar Raghavan. Graph Structure of the Web: A Survey. In Proceedings of LATIN, pages 123-125, 2000.

Example 2: Friendship Networks

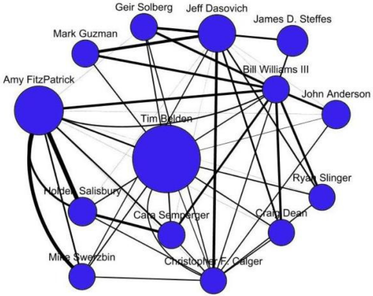
Friendship Network



Nodes: Friends
Edges: Friendship

Reference: Moody 2001

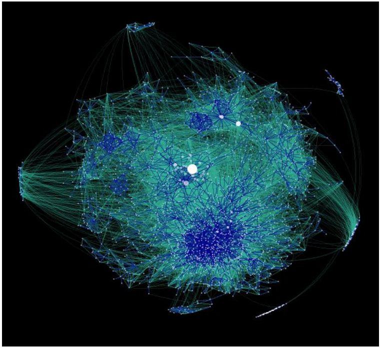
Subgraph of Email Network



Nodes: Individuals
Edges: Email Communication

Reference: Schall 2009

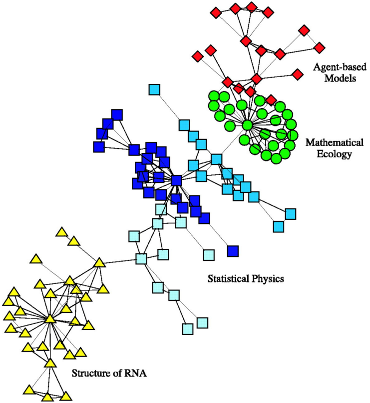
Example 3: Weblog Networks



Nodes: Blogs
Edges: Links

Reference: Hurst 2007

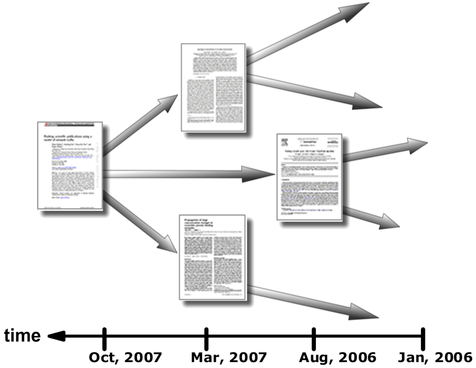
Example 4: Co-authorship Networks



Nodes: Scientists Edges: Co-authorship

Reference: M.E.J. Newman. Coauthorship networks and patterns of scientific collaboration. PNAS, 101(1):5200-5205, 2004

Example 5: Citation Networks



Nodes: Journals

Edges: Citation

Reference: http://eigenfactor.org/

Social Networks - Definition

- *Social Network:* A social system made up of individuals and interactions among these individuals

- Represented using graphs
    - Nodes - Friends, Publications, Authors, Organizations, Blogs, etc.
    - Edges - Friendship, Citation, Co-authorship, Collaboration, Links, etc.

S.Wasserman and K. Faust. Social Network Analysis. Cambridge University Press, Cambridge, 1994

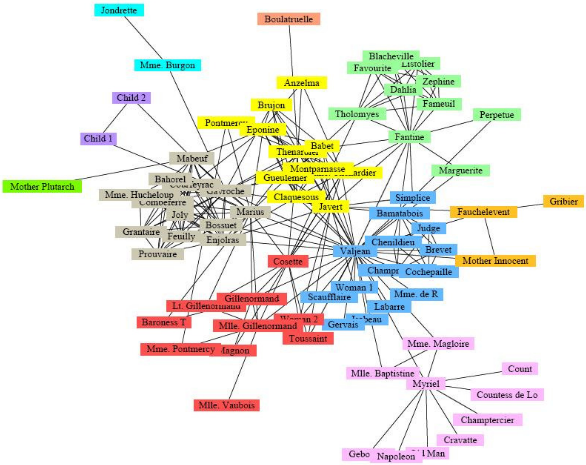
Social Networks are Different from Computer Networks

Social networks differ from technological and biological networks in two important ways:

1. non-trivial clustering or network transitivity, and
2. the phenomenon of degree correlation due to the existence of groups or components in the network

---

- M. E. J. Newman, Assortative mixing in networks. Phys. Rev. Lett. 89, 208701, 2002.
- M. E. J. Newman and Juyong Park. Why social networks are different from other types of networks. Physical Review E 68, 036122, 2003.



Courtesy: M. E. J. Newman and M. Girvan. *Finding and evaluating community structure in networks.* Phys. Rev. E 69, 026113, 2004.

Social Network Analysis (SNA)

- Study of structural and communication patterns
  - degree distribution, density of edges, diameter of the network

- Two principal categories:
  - **Node/Edge Centric Analysis:**
    - Centrality measures such as degree, betweeneness, stress, closeness
    - Anomaly detection
    - Link prediction, etc.
  - **Network Centric Analysis:**
    - Community detection
    - Graph visualization and summarization
    - Frequent subgraph discovery
    - Generative models, etc.

U. Brandes and T. Erlebach. Network Analysis: Methodological Foundations. Springer-Verlag Berlin Heidelberg, 2005.

Why is SNA Important?

- To understand complex connectivity and communication patterns among individuals in the network
- To determine the structure of networks
- To determine influential individuals in social networks
- To understand how social network evolve
- To determine outliers in social networks
- To design effective viral marketing campaigns for targeted advertising
- ...

Next Part of the Presentation

1. Introduction to Social Networks
2. **Key Tasks in Social Network Analysis**
3. Emerging Challenges
4. Summary and Conclusions

A Few Key SNA Tasks

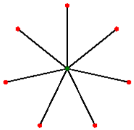
1. Measures to rank nodes (or edges)
2. Community detection
3. Link prediction problem
4. Inferring social networks from social events
5. Viral marketing
6. Graph Visualization
7. Design of incentives in networks
8. Determining implicit social hierarchy
9. Network formation
10. Sparsification of social networks (with purpose)
11. ...

Task 1: Centrality Measures

- Significant amount of attention in the analysis of social networks is devoted to understand the centrality measures
- A centrality measure essentially ranks nodes/edges in a given network based on either their positional power or their influence over the network;
- Some well known centrality measures:
  - Degree centrality
  - Closeness centrality
  - Clustering coefficient
  - Betweenness centrality
  - Eigenvector centrality, etc.

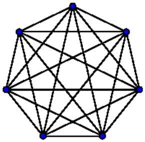
Task 1: Centrality Measures

- Significant amount of attention in the analysis of social networks is devoted to understand the centrality measures
- A centrality measure essentially ranks nodes/edges in a given network based on either their positional power or their influence over the network;
- Some well known centrality measures:
  - Degree centrality
  - Closeness centrality
  - Clustering coefficient
  - Betweenness centrality
  - Eigenvector centrality, etc.



Task 1: Centrality Measures

- Significant amount of attention in the analysis of social networks is devoted to understand the centrality measures
- A centrality measure essentially ranks nodes/edges in a given network based on either their positional power or their influence over the network;
- Some well known centrality measures:
  - Degree centrality
  - Closeness centrality
  - Clustering coefficient
  - Betweenness centrality
  - Eigenvector centrality, etc.

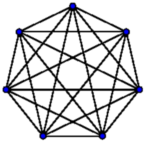


Task 1: Centrality Measures

- Significant amount of attention in the analysis of social networks is devoted to understand the centrality measures

- A centrality measure essentially ranks nodes/edges in a given network based on either their positional power or their influence over the network;

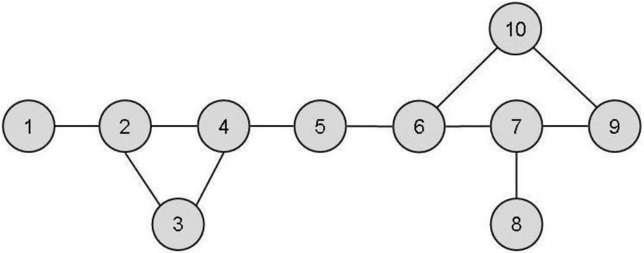
- Some well known centrality measures:
    - Degree centrality
    - Closeness centrality
    - Clustering coefficient
    - Betweenness centrality
    - Eigenvector centrality, etc.



Degree Centrality

- **Degree Centrality:** The degree of a node in a undirected and unweighted graph is the number of nodes in its immediate neighborhood.
  - Rank nodes based on the degree of the nodes in the network
  - Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. Social Networks, 1(3), 215-239
  - Degree centrality (and its variants) are used to determine influential seed sets in viral marketing through social networks

Degree Centrality (Cont.)

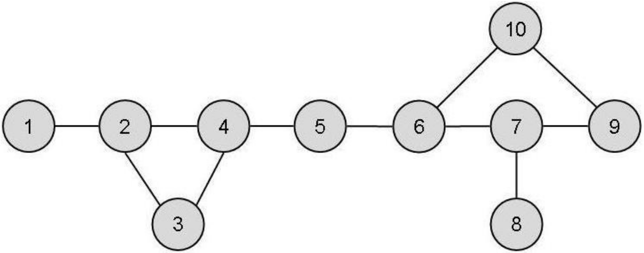


Degree Centrality										
Node	1	2	3	4	5	6	7	8	9	10
Value	1	3	2	3	2	3	3	1	2	2
Rank	9	1	5	1	5	1	1	9	5	5

Closeness Centrality

- The farness of a node is defined as the sum of its shortest distances to all other nodes;
- The closeness centrality of a node is defined as the inverse of its farness;
- The more central a node is in the network, the lower its total distance to all other nodes.

Closeness Centrality (Cont.)

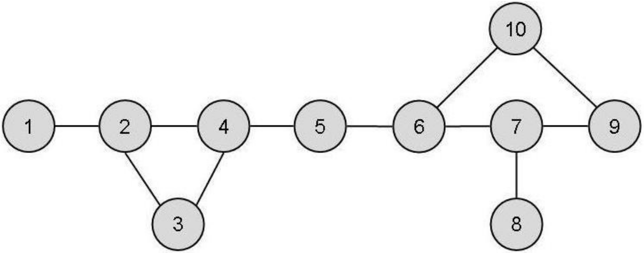


Closeness Centrality										
Node	1	2	3	4	5	6	7	8	9	10
Value	$\frac{1}{34}$	$\frac{1}{26}$	$\frac{1}{27}$	$\frac{1}{21}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{1}{23}$	$\frac{1}{31}$	$\frac{1}{29}$	$\frac{1}{25}$
Rank	10	6	7	3	1	1	4	9	8	5

Clustering Coefficient

- It measures how dense is the neighborhood of a node.
- The clustering coefficient of a node is the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them.
- D. J. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. Nature 393 (6684): 440442 , 1998.
- Clustering coefficient is used to design network formation models

Clustering Coefficient (Cont.)



Clustering Coefficient										
Node	1	2	3	4	5	6	7	8	9	10
Value	0	$\frac{1}{3}$	1	$\frac{1}{3}$	0	0	0	0	0	0
Rank	3	2	1	2	3	3	3	3	3	3

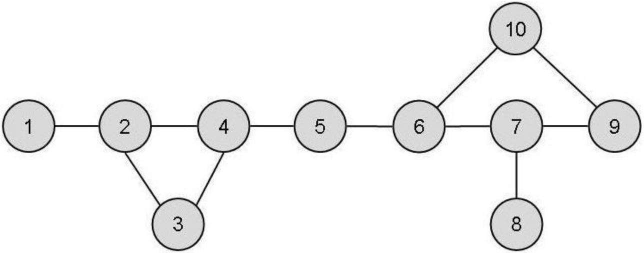
Betweeness Centrality

- **Between Centrality:** Vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen nodes have a high betweenness.
    - Formally, betweenness of a node v is given by

      $$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$

      where $\sigma_{s,t}(v)$ is the number of shortest paths from s to t that pass through v and $\sigma_{s,t}$ is the number of shortest paths from s to t.
    - L. Freeman. A set of measures of centrality based upon betweenness. Sociometry, 1977.
    - Betweenness centrality is used to determine communities in social netwoks (Reference: Girvan and Newman (2002)).

Betweenness Centrality (Cont.)



Betweenness Centrality										
Node	1	2	3	4	5	6	7	8	9	10
Value	0	8	0	18	20	21	11	0	1	6
Rank	8	5	8	3	2	1	4	8	7	6

A Simple Observation

ID	Degree Centrality	Closeness Centrality	Clustering Centrality	Betweenness Centrality	Eigenvector Centrality
1	9	10	3	8	9
2	1	6	2	5	2
3	5	7	1	8	3
4	1	3	2	3	1
5	5	1	3	2	5
6	1	1	3	1	3
7	1	4	3	4	6
8	9	9	3	8	10
9	5	8	3	7	8
10	5	5	3	6	7

Task 2: Community Detection

- *Based on Link Structure in the Social Network:*
  - Determining dense subgraphs in social graphs
  - Graph partitioning
  - Determining the best subgraph with maximum number of neighbors
  - Overlapping community detection
- Based on Activities over the Social Network
  - Determine action communities in social networks
  - Overlapping community detection
- J. Leskovec, K.J. Lang, and M.W. Mahoney. Empirical comparison of algorithms for network community detection. In WWW 2010.
- Ramasuri Narayanam and Y. Narahari. A Game Theory Inspired, Decentralized, Local Information based Algorithm for Community Detection in Social Graphs. To appear in International Conference on Pattern Recognition (ICPR), 2012.

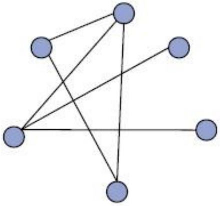
Task 3: Link Prediction Problem

- Given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future?

- D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In CIKM 2003.

Task 3: Link Prediction Problem

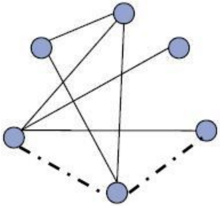
- Given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future?



- D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In CIKM 2003.

Task 3: Link Prediction Problem

- Given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future?



- D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In CIKM 2003.

Task 4: Inferring Social Networks From Social Events

- In the traditional link prediction problem, a snapshot of a social network is used as a starting point to predict (by means of graph-theoretic measures) the links that are likely to appear in the future.
- Predicting the structure of a social network when the network itself is totally missing while some other information (such as interest group membership) regarding the nodes is available.
- V. Leroy, B. Barla Cambazoglu, F. Bonchi. Cold start link prediction. In SIGKDD 2010.

Task 5: Viral Marketing

- With increasing popularity of online social networks, viral Marketing - the idea of exploiting social connectivity patterns of users to propagate awareness of products - has got significant attention
- In viral marketing, within certain budget, typically we give free samples of products (or sufficient discounts on products) to certain set of influential individuals and these individuals in turn possibly recommend the product to their friends and so on
- It is very challenging to determine a set of influential individuals, within certain budget, to maximize the volume of information cascade over the network
- P. Domingos and M. Richardson. Mining the network value of customers. In ACM SIGKDD, pages 5766, 2001.

Task 5: Viral Marketing (Cont.)

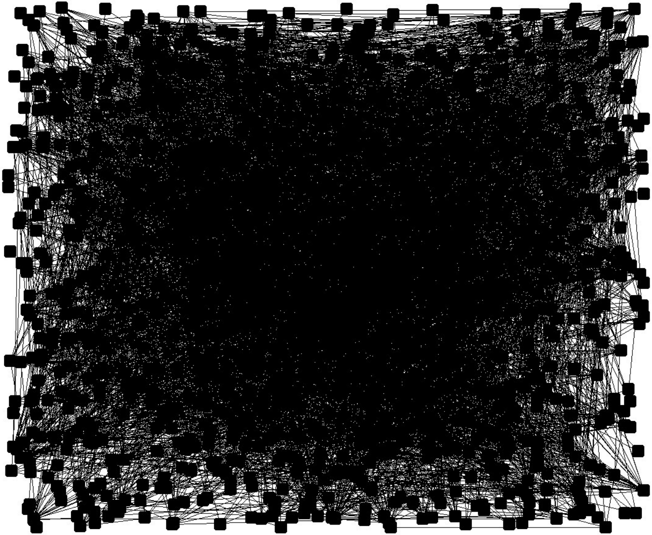
- Often not only positive opinions about the products, but also negative opinions may emerge and propagate over the social network.
- How to choose the initial seeds for viral marketing in the presence of both positive and negative opinions?
- W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In SDM 2011.
- How to choose the initial seeds for viral marketing of products in the presence of competing products already in the market?
- X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In SDM, 2012.

Task 5: Viral Marketing (Cont.)

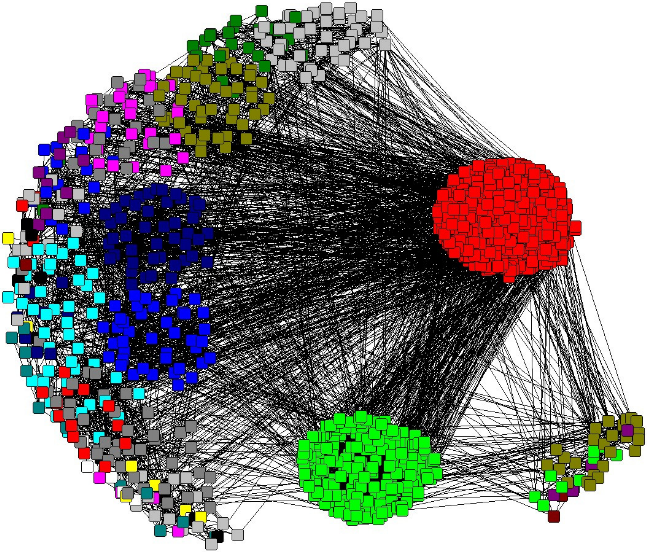
Viral Marketing with Product Dependencies

- Often cross-sell or up-sell is possible among the products
- Product specific costs for promoting the products have to be considered
- Since a company often has budget constraints, the initial seeds have to be chosen within a given budget
- Ramasuri Narayanam and Amit A. Nanavati. Viral marketing with product cross-sell through social networks. To appear in ECML-PKDD, 2012.

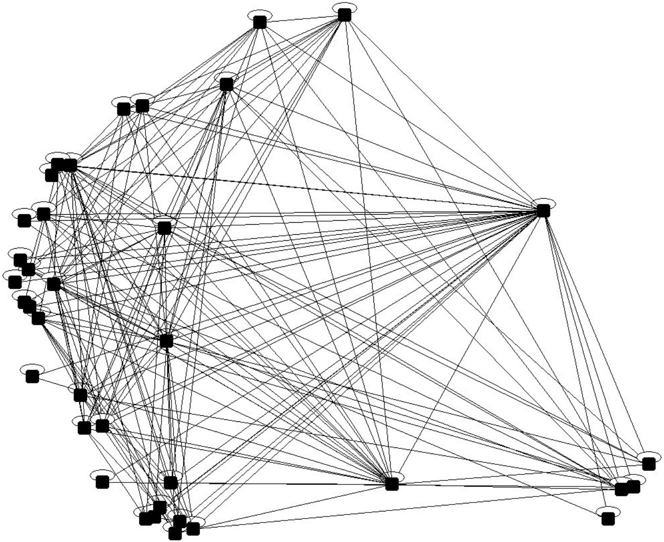
Task 6: Graph Visualization



Task 6: Graph Visualization



Task 6: Graph Visualization



Task 7: Design of Incentives in Networks

- Users pose queries to the network itself, rather than posing queries to a centralized system.
- At present, the concept of incentive based queries is used in various question-answer networks such as Yahoo! Answers, Orkuts Ask Friends, etc.
- In the above contexts, only the person who answers the query is rewarded, with no reward for the intermediaries. Since individuals are often rational and intelligent, they may not participate in answering the queries unless some kind of incentives are provided.
- It is also important to consider the quality of the answer to the query, when incentives are involved.
- J. Kleinberg and P. Raghavan. Query incentive networks. In Proceedings of 46th IEEE FOCS, 2005.

Task 8: Determining Implicit Social Hierarchy

- Social stratification refers to the hierarchical classification of individuals based on power, position, and importance
- The popularity of online social networks presents an opportunity to study social hierarchy for different types of large scale networks
- M. Gupte, P. Shankar, J. Li, S. Muthukrishnan, and L. Iftode. Finding hierarchy in directed online social networks. In the Proceedings of World Wide Web (WWW) 2011.

Task 9: Network Formation

- More often links among individuals in social networks form by choice not by chance
- These links capture the associated social and economic incentives
- How to model the formation of social networks in the presence of strategic individuals (or organizations)?
- What are the networks that will emerge due to the dynamics of network formation and what their characteristics are likely to be?
- Matthew O. Jackson. Social and Economic Networks. Princeton University Press, Princeton and Oxford, 2008
- Ramasuri Narayanam and Y. Narahari. Topologies of Strategically Formed Social Networks Based on a Generic Value Function - Allocation Rule Model. Social Networks, 33(1), 2011

Task 10: Sparsification of Social Networks

- Real world social networks are very large in the sense that they contain millions of nodes and billions of edges
- Certain applications associated with social network data need output quickly. In particular, they can compromise even on the solution quality till some extent but not on the execution time requirements
- The above leads to an interesting and challenging research problem, namely *sparsification of social networks*
- Using the sparse social graphs, we perform SNA and again map these results back to the original network if required
- V. Satuluri, S. Parthasarathy, Y. Ruan. Local graph sparsification for scalable clustering. In SIGMOD, 2011.
- M. Mathioudakis, F. Bonchi, C. Castillo, A. Gionis, A. Ukkonen. Sparsification of influence networks. In SIGKDD 2011.

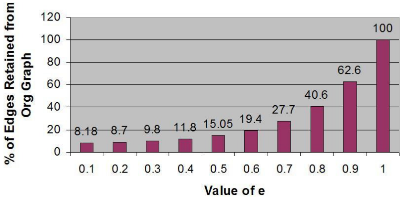
Next Part of the Talk

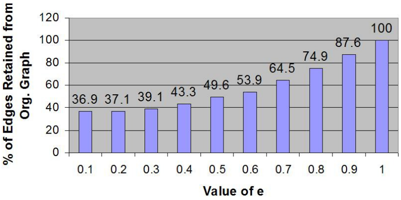
Emerging Challenges in SNA

- Availability of Auxiliary Data
  - Recent applications witness data related to not only who is connected to whom, but also the activities performed by the users
- Availability of Large Data Sets
  - Technological advancements made it easy to collect network data sets with very large sizes
- Dynamic Nature of the Network Data Sets
  - The structure of the network changes over time due to user activity
- Strategic Behavior of Users
  - More often the nodes in the social network are individuals or organizations
  - Such entities more often exhibit strategic behavior
  - Game theory and mechanism design can naturally model such scenarios
- Nature of the Recent Applications
- Privacy Related Issues

Challenge 1: Availability of Auxiliary Data

Data Set	**Nodes**	*Edges*	**Auxiliary Data**
Digg	68,633	1,383,941	Comments Data
FilmTip	39,581	176,436	Ratings Data





Challenge 1: Availability of Auxiliary Data (Cont.)

e	Digg		FilmTip	
2-3 4-5	CNM	Metis	CNM	Metis
0.1	0.2076	**0.0751**	**0.7784**	**0.6477**
0.2	**0.2653**	**0.0743**	**0.782**	**0.6449**
0.3	**0.2817**	0.0738	**0.7913**	**0.6454**
0.4	**0.2563**	**0.0743**	**0.7991**	**0.6429**
0.5	**0.27**	**0.0744**	**0.8003**	**0.6344**
0.6	**0.2682**	0.0739	**0.7985**	**0.6282**
0.7	**0.2958**	0.0726	**0.7942**	**0.6215**
0.8	**0.3029**	0.0704	**0.7886**	0.6160
0.9	0.2553	0.0700	**0.7718**	**0.6122**
1	0.2561	0.0741	0.771	0.6188

Table: Performance Evaluation: Modularity scores

Challenge 2: Availability of Large Network Data

- Size of data poses three challenging problems: Space, Running Time, and Quality
- Ulrik Brandes, A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology 25(2):163-177, 2001
  - Standard algorithm for betweenness centrality requires $\Theta(n^3)$ time and $\Theta(n^2)$ space
  - Faster algorithm requires $\Theta(nm)$ time and $\Theta(n+m)$ space

Challenge 2: Availability of Large Network Data (Cont.)

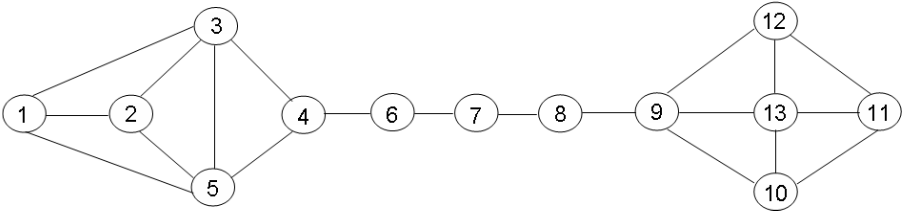
Dataset	Nodes	Edges
Amazon	334,863	925,872
DBLP	317,080	1,049,866
FilmTip	*39,581*	*88,218*
Flixster	786,936	7,058,819

Dataset	CN	CS	CS-AL	JC	RA	SI	Original
Amazon	0.9914	0.9944	**0.9981**	0.9947	0.9927	0.9932	0.8805
DBLP	0.9024	0.9453	**0.9878**	0.9546	0.9058	0.9331	0.8735
FilmTip	0.9134	0.9742	**0.9947**	0.9671	0.9477	0.9595	0.8447
Flixster	0.7399	0.7997	**0.9567**	0.8143	0.718	0.7886	0.6556

Table: Performance Evaluation: Modularity scores

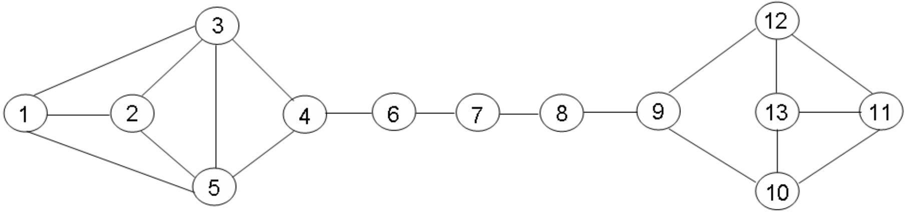
Hemank Lamba and Ramasuri Narayanam. *Community Detection in Social Networks using Proximity Scores*. To appear in 14th International Conference on Web Information System Engineering (WISE), 2013.

Challenge 3: Dynamic Nature of the Network Data



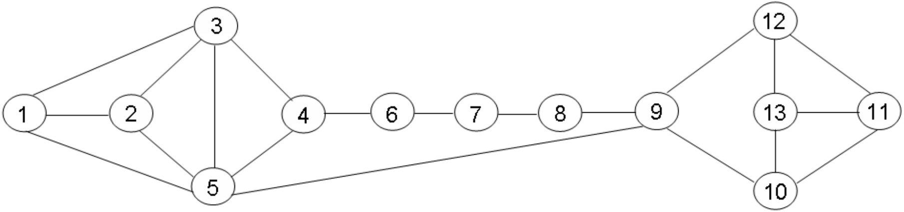
- Design of efficient algorithms for dynamic networks
- Min-Joong Lee, Jungmin Lee, Jaimie Yejean Park, Ryan Hyun Choi, Chin-Wan Chung: QUBE: a quick algorithm for updating betweenness centrality. WWW 2012: 351-360

Challenge 3: Dynamic Nature of the Network Data



- Design of efficient algorithms for dynamic networks
- Min-Joong Lee, Jungmin Lee, Jaimie Yejean Park, Ryan Hyun Choi, Chin-Wan Chung: QUBE: a quick algorithm for updating betweenness centrality. WWW 2012: 351-360

Challenge 3: Dynamic Nature of the Network Data

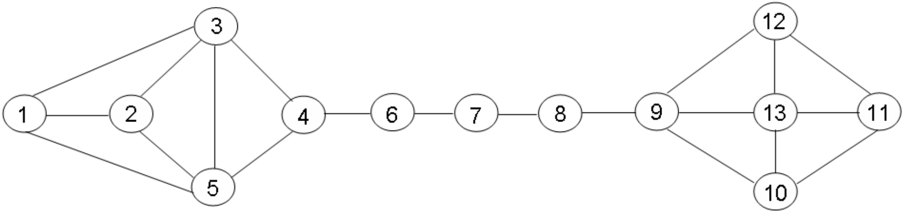


- Design of efficient algorithms for dynamic networks
- Min-Joong Lee, Jungmin Lee, Jaimie Yejean Park, Ryan Hyun Choi, Chin-Wan Chung: QUBE: a quick algorithm for updating betweenness centrality. WWW 2012: 351-360

Challenge 4: Nature of Recent Applications

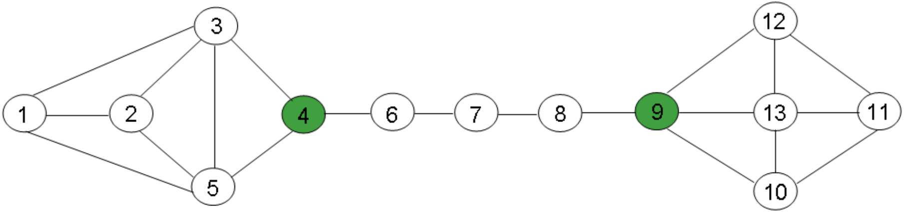
- We consider a network of individuals (such as social network of the buyers) or a network of objects (such as intranet of a company);
- Assume that certain unwanted process may attack a node uniformly at random and then starts spreading over the network effecting the function of all reachable nodes/individuals;
- We have some limited budget to reach out at most k nodes;
- The problem is which k nodes that we should target to minimize the expected number of the nodes that receive the misinformation.

A Stylized Example



Centrality Measure	Rank 1	Rank 2
Degree	9	3,5,10,11,12,13
Closeness	7	6,8
Betweenness	7	6,8
Clustering Coefficient	10,11,12,13	9
EigenVector	1,2,10,11,12,13	3,5
PageRank	9	3,5

A Stylized Example



Centrality Measure	Rank 1	Rank 2
Degree	9	3,5,10,11,12,13
Closeness	7	6,8
Betweenness	7	6,8
Clustering Coefficient	10,11,12,13	9
EigenVector	1,2,10,11,12,13	3,5
PageRank	9	3,5

Challenge 5: Privacy Related Issues

- Computing resources available with PCs and Cluster of CPUs is inadequate for Huge Volumes of network data
- Cloud computing paradigm is alternative for the above limitation
- This poses the problem of sharing the network data with the remote location
- Privacy preserving computation is on demand!

Next Part of the Talk

1. Introduction to Social Networks
2. Key Tasks in Social Network Analysis
3. Emerging Challenges
4. **Summary and Conclusions**

Summary of the Presentation

- First gave a brief introduction to social networks and social network analysis
- Then we saw a number of classical key tasks in social network analysis
- Next we explored a set of important challenges associated with social network analysis

Some Important Text Books

- D. Easley and J. Kleinberg. Networks, Crowds, and Markets. Cambridge University Press, 2010.
- Y. Narahari, Dinesh Garg, Ramasuri Narayanam, Hastagiri Prakash. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*. Springer Series in Advanced Information and Knowledge Processing (AIKP), London, UK, 2009.
- M.E.J. Newman. Networks: An Introduction. Oxford University Press, 2010.
- M.O. Jackson. Social and Economic Networks. Princeton University Press, 2008.
- U. Brandes and T. Erlebach. Network Analysis: Methodological Foundations. Springer-Verlag Berlin Heidelberg, 2005.

Some Leading Researchers

- Jon M. Kleinberg from Cornell University
- Christos Faloutsos from CMU
- Matthew O. Jackson from Stanford University
- Sanjeev Goyal from University of Cambridge
- Eva Tardos from Cornell University
- Jure Leskovec from Stanford University
- Nicole Immorlica from Northwestern University
- David Kempe from University of Southern California
- ...

Network Dataset Repositories

- Jure Leskovec: http://snap.stanford.edu/data/index.html
- MEJ Newman: http://www-personal.umich.edu/m̃ejn/netdata
- Albert L. Barabasi: http://www.nd.edu/ñetworks/resources.htm
- NIST Data Sets: http://math.nist.gov/R̃Pozo/complex_datasets.html
- MPI Data Sets: http://socialnetworks.mpi-sws.org/
- ...

Software Tools for Network Analysis

- Gephi (Graph exploration and manipulation software)
- Pajek (Analysis and Visualization of Large Scale Networks)
- UCINET (Social Network Analysis tool)
- CFinder (Finding and visualizing communities)
- GraphStream (Dynamic graph library)
- Graphviz (Graph vizualisation software)
- Refer to Wikipedia for more information (http://en.wikipedia.org/wiki/Social_network_analysis_software)

A List of Important Conferences

- ACM Conference on Electronic Commerce (ACM EC)
- Workshop on Internet and Network Economics (WINE)
- ACM SIGKDD
- WSDM
- ACM Internet Measurement Conference (ACM IMC)
- CIKM
- ACM SIGCOMM
- Innovations in Computer Science (ICS)
- AAMAS
- AAAI
- IJCAI
- ...

A List of Important Journals

- American Journal of Sociology
- Social Networks
- Physical Review E
- Data Mining and Knowledge Discovery
- ACM Transactions on Internet Technology
- IEEE Transactions on Knowledge and Data Engineering
- Games and Economic Behavior
- ...

Thank You

Game Theoretic Centrality Design

Limitations of Existing Approaches

- The common phenomenon of these standard centrality measures is that they assess the importance of each node by focusing only on the role played by that node itself.
- Such an approach is inadequate to capture the synergies that may occur if the functioning of nodes as groups is considered.
- Here I discuss a new approach to measure social capital which builds upon cooperative game theory. It is:
  - not only a natural tool to model social capital,
  - but also quantifies various aspects that are not accounted for by other approaches such as *synergies that may occur if the functioning of nodes as groups is considered*

Motivating Scenario 1 (Cont.)

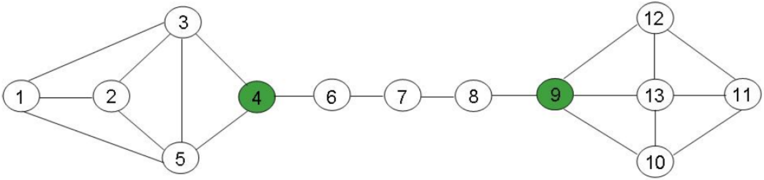
- Any team of employees to perform project *should be connected in the graph*
- Now we derive the values of the following coalitions:
  - $v(\{e_1\}) = 1$ (as employee e_1 can perform only P_1),
  - $v(\{e_2\}) = 1$ (as employee e_2 can perform only P_2),
  - $v(\{e_3\}) = 0$ (as employee e_3 cannot perform any project alone),
  - $v(\{e_1, e_2\}) = 3$ (as employees together can perform P_1, P_2, P_3),
  - $V(\{e_1, e_3\}) = 0$ (as it is a disconnected team).

- **Observation:** *The value of a group of employees is NOT the same as the sum of values of the individual employees in that group.*

Motivating Scenario 2: Limiting Misinformation Spread

- More recently, companies often rely on viral marketing or word-of-marketing of products to maximize their revenue;
- At times, not only positive opinions, but also negative opinions may emerge and spread over the network of potential buyers;
- The company which owns this product wants to minimize the loss incurred due to the negative opinions;
- The question is which individuals the company should target (for convincing) in order to minimize the expected number of individuals that receive the negative opinion.

The Problem Scenario: Example when k = 2



Centrality Measure	Rank 1	Rank 2
Degree	9	3,5,10,11,12,13
Closeness	7	6,8
Betweenness	7	6,8
Clustering Coefficient	10,11,12,13	9
EigenVector	1,2,10,11,12,13	3,5
PageRank	9	3,5

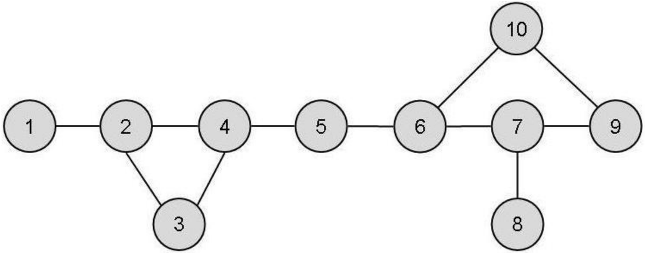
Motivating Scenario 3: Betweeness Centrality

- **Between Centrality:** Vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen nodes have a high betweenness.
    - Formally, betweenness of a node v is given by

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$

      where $\sigma_{s,t}(v)$ is the number of shortest paths from s to t that pass through v and $\sigma_{s,t}$ is the number of shortest paths from s to t.
    - L. Freeman. A set of measures of centrality based upon betweenness. Sociometry, 1977.
    - Betweenness centrality is used to determine communities in social netwoks (Reference: Girvan and Newman (2002)).

Motivating Scenario 3: Betweenness Centrality (Cont.)

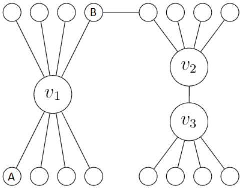


Betweenness Centrality										
Node	1	2	3	4	5	6	7	8	9	10
Value	0	8	0	18	20	21	11	0	1	6
Rank	8	5	8	3	2	1	4	8	7	6

Motivating Scenario 3: Betweenness Centrality (Cont.)

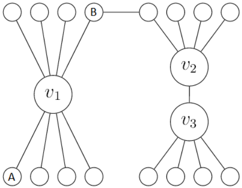
- Betweenness centrality of nodes v_1 and v_2 is the same as both of these nodes control same number of shortest paths

- Consider the roles that nodes play within groups:
  - A significant percentage of the shortest paths controlled by v_2 are also controlled by v_3,
  - The number of shortest paths controlled by the group $\{v_2, v_3\}$ is much smaller than the group $\{v_1, v_3\}$.



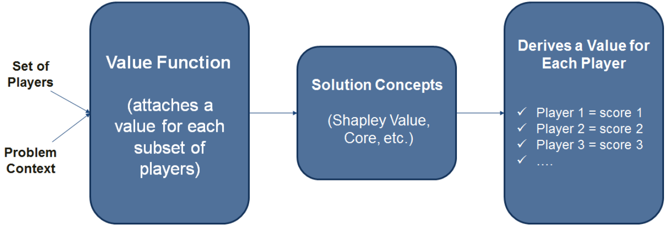
Motivating Scenario 3: Betweenness Centrality (Cont.)

- Betweenness centrality of nodes v_1 and v_2 is the same as both of these nodes control same number of shortest paths
- Consider the roles that nodes play within groups:
  - A significant percentage of the shortest paths controlled by v_2 are also controlled by v_3,
  - The number of shortest paths controlled by the group $\{v_2, v_3\}$ is much smaller than the group $\{v_1, v_3\}$.



Game Theory based Centrality Design

- The classical approach to centrality in networks tries to come up with a value for a node by just focusing on that node itself
- Cooperative game theory based approach to design centrality measures on networks takes a sophisticated approach by taking into account the synergies that are possible with groups of nodes



- Computation of scores / values using game theoretic approach is challenging in general

Cooperative Game Theory

- **Definition:** A cooperative game is defined as the pair (N, v) where $N = \{1, 2, \ldots, n\}$ is a set of players and $v : 2^N \rightarrow \mathbb{R}$ is a characteristic function (or value function), with $v(.) = 0$.
- **Example:** There is a seller s and two buyers b_1 and b_2. The seller has a single unit to sell and his willingness to sell the item is 10. Similarly, the valuations for b_1 and b_2 are 15 and 20 respectively. The corresponding cooperative game is:
  - $N = \{s, b_1, b_2\}$
  - $v(\{s\}) = 0 \quad , v(\{b_1\}) = 0 \quad , v(\{b_2\}) = 0 \quad , v(\{b_1, b_2\}) = 0$
    $v(\{s, b_1\}) = 5 \quad , v(\{s, b_2\}) = 10 \quad , v(\{s, b_1, b_2\}) = 10$

Cooperative Game Theory (Cont.)

- **Solution Concept:** This distributes the value of grand coalition in (fair) manner to all the players in the game.
- Some popular examples for solution concepts include: Imputation, Core, Shapley Value, etc.

- A payoff allocation $x = (x_1, x_2, \ldots, x_n)$ is any vector in $\mathbb{R}^n$ such that $\sum_i x_i = v(N)$ and x_i indicates the utility payoff to player i.
- Any payoff allocation $x = (x_1, x_2, \ldots, x_n)$ is said to be *feasible* for a coalition C if and only if

$$\sum_{i \in C} x_i \leq v(C)$$

The Shapley Value

- Shapley value is a solution concept which is motivated by the need to have a theory that would predict a unique expected payoff allocation for every given coalitional game
- The Shapley value concept was proposed by Shapley in 1953, following an axiomatic approach. This was part of his doctoral dissertation at the Princeton University. Given a cooperative game (N, v), the Shapley value is denoted by $\phi(v)$:

$$\phi(v) = \{\phi_i(v), \phi_2(v), \ldots, \phi_n(v)\}$$

  where $\phi_i(v)$ is the expected payoff to player i
- Shapley proposed three axioms: Symmetry, Linearity, and Carrier

The Shapley Value (Cont.)

- Let (N, v) be a coalitional game and π be a permutation of the players in N. Let $(N, \pi v)$ be a coalitional game such that

$$\pi v(\{\pi(i) : i \in C\}) = v(C), \forall C \subseteq N$$

  That is, the role of any player i in (N, v) is same as the role of player $\pi(i)$ in $(N, \pi v)$
- *Symmetry:* For any $v \in \mathbb{R}^{2^n-1}$, any permutation π on N, and any player $i \in N$,

$$\phi_{\pi(i)}(\pi v) = \phi_i(v)$$

- *Linearity:* Let (N, v) and (N, w) be any two coalitional games. Suppose $p \in [0, 1]$. Define the game $(N, pv + (1 - p)w)$ as follows:

$$(pv + (1 - p)w)(C) = pv(C) + (1 - p)w(C), \ \forall C \subseteq N$$

  Then the axiom of linearity says that

$$\phi_i(pv + (1 - p)w) = p\phi_i(v) + (1 - p)\phi_i(w)$$

The Shapley Value (Cont.)

- *Carrier:* A coalition D is said to ba a carrier of a coalitional game (N, v) if

$$v(C) = v(C \cap D), \ \forall C \subseteq N$$

The carrier axiom states that, for any (N, v) and any carrier D,

$$\sum_{i \in D} \phi_i(v) = v(D) = v(N)$$

The Shapley's Theorem

- **Theorem:** There is exactly one mapping $\phi : \mathbb{R}^{2^N-1} \rightarrow \mathbb{R}^N$ that satisfies Symmetry, Linearity, and Carrier axioms. This function satisfies: $\forall i \in N$, $\forall v \in \mathbb{R}^{2^N-1}$,

$$\phi_i(v) = \sum_{C \subseteq N \setminus \{i\}} \frac{|C|!(n-|C|-1)!}{n!} \{v(C \cup \{i\}) - v(C)\}$$

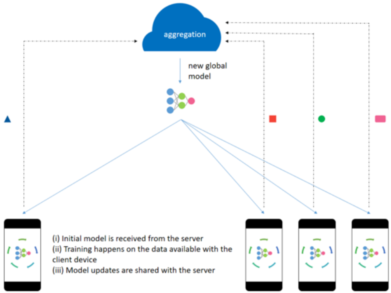
- **Example:** Consider the following cooperative game: $N = \{1, 2, 3\}$, $v(1) = v(2) = v(3) = v(23) = 0$, $v(12) = v(13) = v(123) = 300$. Then we have that

$$\phi_1(v) = \frac{2}{6}v(1) + \frac{1}{6}(v(12) - v(2)) + \frac{1}{6}(v(13) - v(3)) + \frac{2}{6}(v(123) - v(23))$$

It can be easily computed that $\phi_1(v) = 200$, $\phi_2(v) = 50$, $\phi_3(v) = 50$

Applications of This New Approach

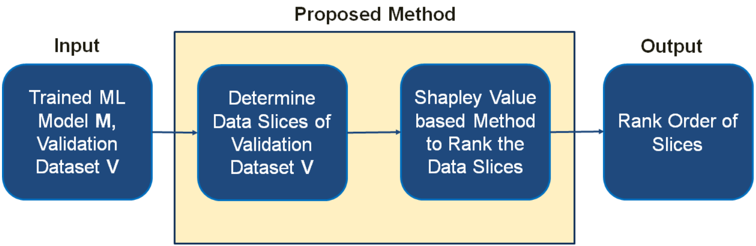
- **Relevant Client Selection in Federated Learning:**
  - Unlike in traditional setttings, all data is not available at one server in Federated Learning setting.
  - In Federated learning, there would several participating clients owning private data (in distributed manner)
  - To select Relevant Clients in Federated Learning, we proposed Approximation methods and Developed Shapley-Fed-Avg Algorithm.



Usage of This New Approach (Cont.)

ML Model Validation and Development:

- To help design Test Cases in *Black-box AI Testing* (Polynomial time, Developed Slice-Ranking-Algorithm)



Usage of This New Approach (Cont.)

- **Influence Maximization through Social Networks:** To select Influential Nodes for Viral Marketing in Social Networks (Approximation methods, Developed SPIN Algorithm)
- **Limiting Mis-information Spread thorugh Social Networks:** To Limit Misinformation Spread through Social Networks (Approximation methods, Developed algorithm to compute Gatekeeper Centrality of nodes in a graph/network)

THANK YOU