Game Theoretic Models for Social Network Analysis

Ramasuri Narayanam   Amit A. Nanavati

IBM Research, India
Email ID: {nramasuri, namit}@in.ibm.com

23-July-2012

Outline of the Presentation

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

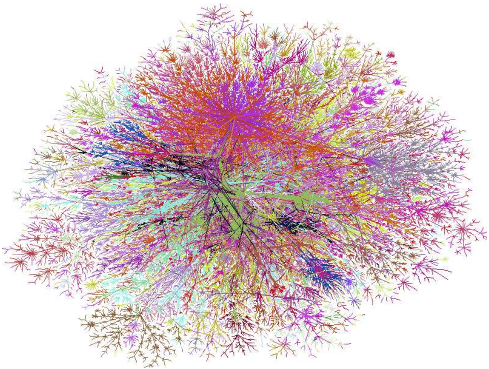
Social Networks: Introduction (Cont.)

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

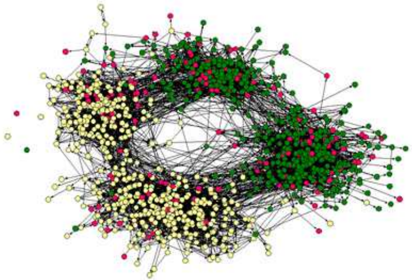


Nodes: Static web pages
Edges: Hyper-links

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

Example 2: Friendship Networks

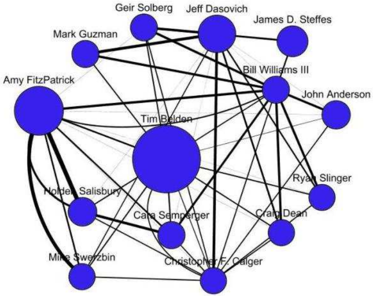
Friendship Network



Nodes: Friends
Edges: Friendship

Reference: Moody 2001

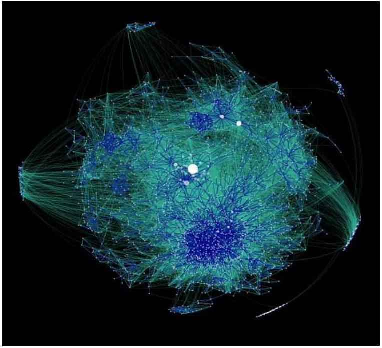
Subgraph of Email Network



Nodes: Individuals
Edges: Email Communication

Reference: Schall 2009

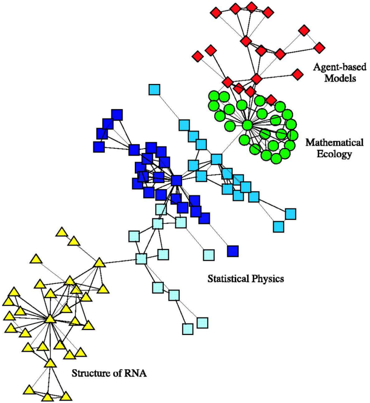
Example 3: Weblog Networks



Nodes: Blogs
Edges: Links

Reference: Hurst 2007

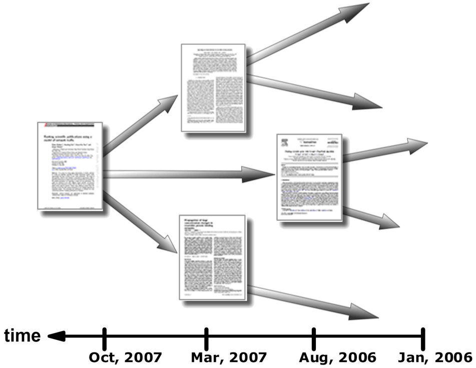
Example 4: Co-authorship Networks



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

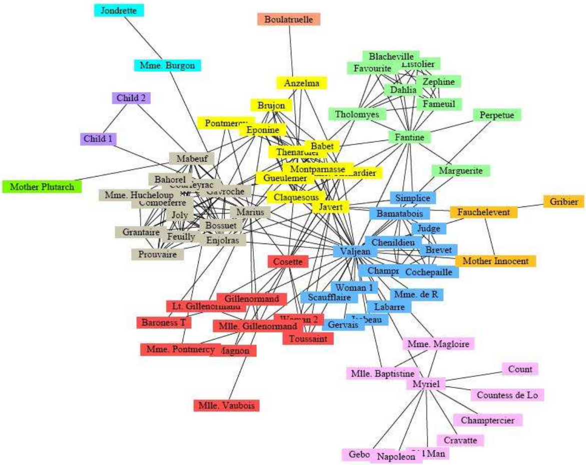
Social Networks are Different from Computer Networks

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
- ...

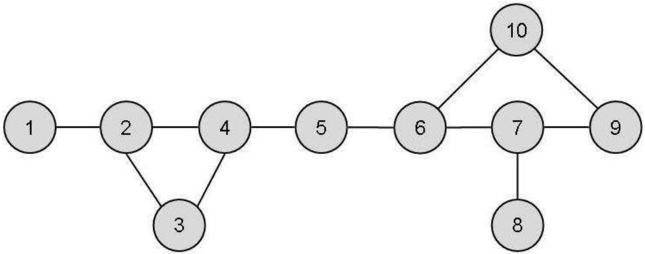
A Few Key SNA Tasks

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

Task 1: Measures to Rank Nodes

- There are various centrality measures to determine the relative importance of an actor or a relationship in a network
- **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

Task 1: Measures to Rank Nodes (Cont.)



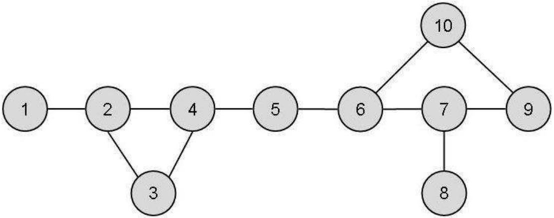
Degree Centrality										
Actor *x*	1	2	3	4	5	6	7	8	9	10
$\sigma_D(x)$	1	3	2	3	2	3	3	1	2	2
Rank	9	1	5	1	5	1	1	9	5	5

Task 1: Measures to Rank Nodes (Cont.)

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

- **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

Task 1: Measures to Rank Nodes (Cont.)



Closeness Centrality										
Actor *x*	1	2	3	4	5	6	7	8	9	10
$\sigma_C(x)$	1/34	1/26	1/27	1/21	1/19	1/19	1/23	1/31	1/29	1/25
Rank	10	6	7	3	1	1	4	9	8	5

Clustering Coefficient										
Actor *x*	1	2	3	4	5	6	7	8	9	10
$\sigma_{Cl}(x)$	0	1/3	1	1/3	0	0	0	0	0	0
Rank	3	2	1	2	3	3	3	3	3	3

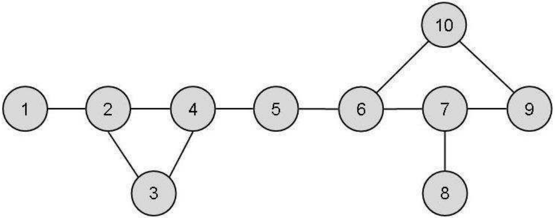
Task 1: Measures to Rank Nodes (Cont.)

- **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)).

Task 1: Measures to Rank Nodes (Cont.)



Betweenness Centrality										
Actor *x*	1	2	3	4	5	6	7	8	9	10
$\sigma_B(x)$	0	8	0	18	20	21	11	0	1	6
Rank	8	5	8	3	2	1	4	8	7	6

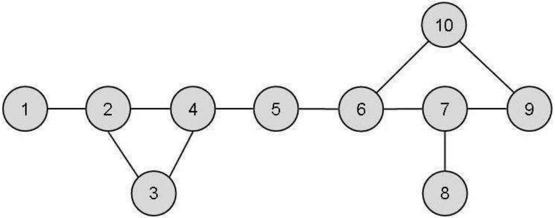
Task 1: Measures to Rank Nodes (Cont.)

- **Katz Centrality:** Katz centrality measures influence by taking into account the total number of walks between a pair of actors, unlike typical centrality measures which consider only the shortest path (the geodesic) between a pair of actors
  - Let A be the adjacency matrix of the network
  - Consider x-th power of A, call it A^x. This represents the number of paths of length x between any pair of nodes
  - Katz centrality of node i is given by

$$C_{Katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \alpha^k (A^k)_{ji},$$

    where α is the attenuation factor
  - L. Katz. A New Status Index Derived from Sociometric Index. Psychometrika, pages 39 – 43, 1953.

Task 1: Measures to Rank Nodes (Cont.)



Centrality Measure by Katz										
Actor *x*	1	2	3	4	5	6	7	8	9	10
$\sigma_K(x)$	1,91	4,72	3,99	5,25	4,06	4,91	4,30	1,77	3,23	3,38
Rank	9	3	6	1	5	2	4	10	8	7

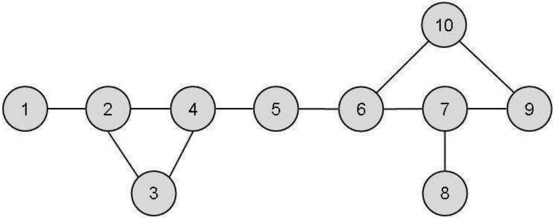
Task 1: Measures to Rank Nodes (Cont.)

- **Eigenvector Centrality:** It assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.
  - Formally, eigen-vector centrality (x_i) of a node i is given by

    $$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j$$

    where $M(i)$ is the set of nodes directly connected to node i.
  - Google Page-Rank and Kats measure are variants of the Eigenvector centrality.
  - P. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. Social Networks, 23(3):191-201, 2001.

Task 1: Measures to Rank Nodes (Cont.)



Eigenvector Centrality										
Actor *x*	1	2	3	4	5	6	7	8	9	10
$\sigma_E(x)$	0,171	0,413	0,363	0,463	0,342	0,363	0,292	0,121	0,221	0,242
Rank	9	2	3	1	5	3	6	10	8	7

Task 1: Measures to Rank Nodes (Cont.)

Task 1: Measures to Rank Nodes (Cont.)

- *Inadequacies of traditional ranking mechanisms for social networks:*
  - They are completely dependent on the structure of the underlying network. Often it is required to rank nodes/edges based on auxiliary data or the value created by the nodes in the network
  - Several empirical evidences reveal that these ranking mechanisms are not scalable to deal with large scale network data
  - They are not tailored to take into account the strategic behavior of the nodes
  - V. Satuluri, S. Parthasarathy and Y. Ruan. Local Graph Sparsification for Scalable Clustering. In SIGMOD 2011.

Task 2: Diversity among Nodes

- Nodes in the network might be having various connectivity patterns
- Some nodes might be connected to high degree nodes, some others might be connected to bridge nodes, etc.
- Determining diversity among the connectivity patterns of nodes is an interesting problem
- L. Liu, F. Zhu, C. Chen, X. Yan, J. Han, P.S. Yu, and S. Yang. Mining Diversity on Networks. In DASFAA 2010.

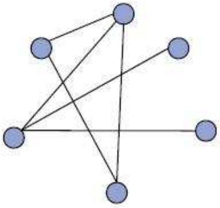
Task 3: Link Prediction Problem

- 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

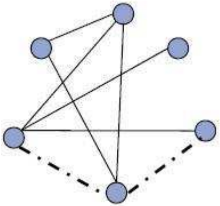
- Given a snapshot of a social network, can we infer which new interactions among its members are likely to occur in the near future?



- 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: 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 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.

Methods to Address SNA Tasks

- Traditional Approaches
  - Graph theoretic techniques
  - Spectral methods
  - Optimization techniques
  - ...

Methods to Address SNA Tasks

- Traditional Approaches
  - Graph theoretic techniques
  - Spectral methods
  - Optimization techniques
  - ...
- Recent Advances
  - Data mining and machine learning techniques
  - **Game theoretic techniques**

Why Game Theoretic Models for SNA?

- Current metrics and measures in SNA are based on
  - Graph theoretic techniques
  - Optimization techniques
  - Spectral techniques, etc.
- Generative models can produce networks with similar structural properties
- In many network settings, the behavior of the system is driven by the actions of a large number of autonomous individuals (or agents)
  - Research collaborations among both organizations and researchers
  - Online social communities such as Orkut, Facebook, LinedIn
  - Telecommunication networks (Service Providers)

Why Game Theoretic Models for SNA? (Cont.)

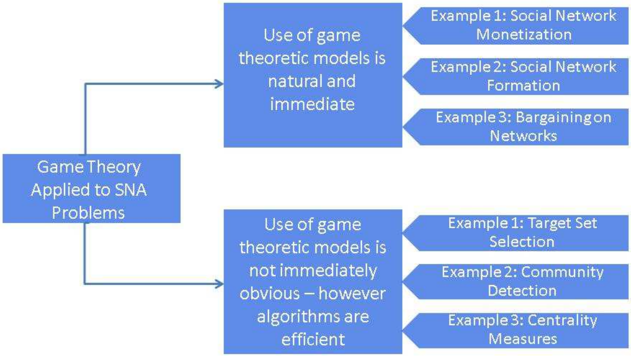
- Individuals are self-interested and optimize their respective objectives
- Inadequacies of current SNA approaches:
  - Social contacts (i.e. links) form more often by choice than by chance
  - There always exist social and economic incentives while forming links in the network
  - Do not satisfactorily capture the behavior of the individuals
  - Do not capture the dynamics of strategic interactions among the individuals in the network

Game theory helps to overcome this fundamental inadequacy

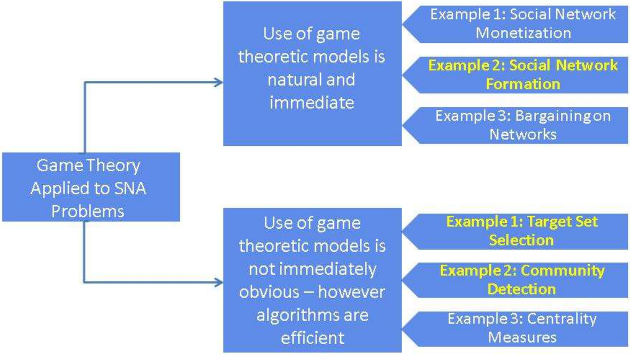
Initial Efforts in this Direction

- Siddharth Suri. The Effects of Network Topology on Strategic Behavior. PhD Thesis, Dept. of Computer and Information Science, University of Pennsylvania, USA, 2007.
- Sanjeev Goyal. Connections: An Introduction to the Economics of Networks. Princeton University Press, Princeton and Oxford, 2007.
- Eyal Even-Dar, Michael J. Kearns, Siddharth Suri. A network formation game for bipartite exchange economies. In SODA 2007.
- Jon M. Kleinberg, and Eva Tardos. Balanced outcomes in social exchange networks. In STOC, 2008.
- Jon M. Kleinberg, Siddharth Suri, Eva Tardos, and Tom Wexler. Strategic Network Formation with Structural Holes. In ACM EC, 2008.
- Matthew O. Jackson. Social and Economic Networks. Princeton University Press, Princeton and Oxford, 2008.

Game Theoretic Models for SNA: Two Viewpoints



Game Theoretic Models for SNA: Two Viewpoints



Outline of the Presentation

Next Part of the Presentation

Next Part of the Presentation

1. Social Network Analysis: Quick Primer
2. Foundational Concepts in Game Theory
3. **SNAzzy: A Social Network Analysis Suite for Business Intelligence**
4. Viral Marketing
5. Community Detection in Social Networks
6. Social Network Formation
7. Summary and To Probe Further

Next Part of the Presentation

1. Social Network Analysis: Quick Primer
2. Foundational Concepts in Game Theory
3. SNAzzy: A Social Network Analysis Suite for Business Intelligence
4. **Viral Marketing**
5. Community Detection in Social Networks
6. Social Network Formation
7. Summary and To Probe Further

Key Resources of this Topic

- D. Kempe, J.M. Kleinberg, and E. Tardos. *Maximizing the spread of influence through a social network*. In SIGKDD, 2003.
- Ramasuri Narayanam and Y. Narahari. *A Shapley Value based Approach to Discover Influential Nodes in Social Networks*. In IEEE Transactions on Automation Science and Engineering (IEEE TASE), 2011.
- Ramasuri Narayanam and Y. Narahari. *Determining Top-k Nodes in Social Networks using the Shapley Value*. In AAMAS, pages 1509-1512, Portugal, 2008.
- Ramasuri Narayanam and Amit A. Nanavati. *Viral Marketing for Product Cross-sell through Social Networks*. To appear in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 2012.

Contents to be Covered

- We first present a simple game theoretic model that brings out several aspects of viral marketing.
    - Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. *A Note on Competitive Diffusion Through Social Networks.* Information Processing Letters, 110:221-225, 2010.

- We then bring out the challenges involved in viral marketing

- We discuss two standard models for propagation of influence in social networks and introduce the influence maximization problem

- Finally discuss a cooperative game theoretic model for influence maximization problem and some other extensions of the influence maximization problem

Diffusion Game

- A game $\Gamma = (G, N)$ is induced by an undirected graph $G = (V, E)$, representing the underlying social network, and the set of agents N.

- The strategy space of each agent is the set of vertices V in the graph, that is, each agent i selects a single node and that node is colored in color i at time 1. We call them *initial trend setters*.

- Note that if two or more agents select the same vertex at time 1 then that vertex becomes gray.

- *Diffusion Process*: At time $t + 1$, each white vertex that has neighbors colored in color i, but does not have neighbors colored in color j for any $j \in N$, is colored in color i.

Diffusion Game (Cont.)

- A white vertex that has two neighbors colored by two distinct colors $i, j \in N$ is colored gray. That is, we assume that if two agents compete for a user at the same time, they cancel out and the user is removed from the game.

- The process continues until it reaches a fixed point, that is, all the remaining white vertices are unreachable due to gray vertices.

- A strategy profile is a vector $x = (x_1, x_2, \ldots, x_n)$, where $x_i \in V$ is the initial vertex selected by agent i. We also denote $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$

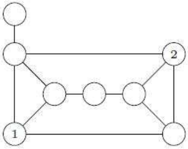
Diffusion Game (Cont.)

- Given a strategy profile $x \in V^n$, the utility of agent $i \in N$, denoted $U_i(x)$, is the number of nodes that are colored in color i when the diffusion process terminates.

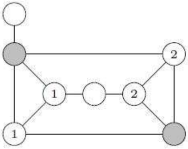
- **Nash Equilibrium:** A strategy profile x is a (pure strategy) Nash equilibrium of the game Γ if an agent cannot benefit from unilaterally deviating to a diffusion strategy. That is, for every $i \in N$, and $x_i^{'} \in V$, it holds that

$$U_i(x_i^{'}, x_{-i}) \leq U_i(x)$$

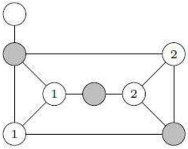
Diffusion Game - Example



(a) Time 1.



(b) Time 2.



(c) Time 3, the process terminates.

Diffusion Game - Nash Equilibrium

- If we can find a Nash equilibrium then we can often predict the behavior of the agents and the outcome of this competitive diffusion process.
- **Theorem:** Every game $\Gamma = (G, N)$, where $D(G) \leq 2$ admits a Nash equilibrium. Furthermore, an equilibrium can be found in polynomial time.
- **Theorem:** Let $N = \{1, 2\}$. There exists a graph G with $D(G) = 3$ such that the game $\Gamma = (G, N)$ does not admit a Nash equilibrium.

Challenges in Viral Marketing

- Propagation of influence is a stochastic process, but not a deterministic process
- The number of individuals in the social network that are getting influenced by the initial trend setters is an *expected quantity*
- Viral marketing for single or multiple products
- There can possibly exists certain types of dependencies among these products

Motivating Example 1: Viral Marketing

- Social networks play a key role for the spread of an innovation or technology
- We would like to market a new product that we hope will be adopted by a large fraction of the network
- Which set of the individuals should we target for?
- Idea is to initially target a few influential individuals in the network who will recommend the product to other friends, and so on
- A natural question is to find a target set of desired cardinality consisting of influential nodes to maximize the volume of the information cascade

Motivating Example 2: Weblogs

- In the domain of weblogs, bloggers publish posts and use hyper-links to refer to other posts and content on the web
- Possible to observe the spread of information in the blogosphere, as each post is time stamped
- In this setting, our goal is to select a small set of blogs (to read) which link to most of the stories that propagate over the blogosphere

Models for Diffusion of Information

- Linear Thresholds Model
- Independent cascade model,
- Decreasing cascade model, etc.

Models for Diffusion of Information

- **Linear threshold model**
- Independent cascade model,
- Decreasing cascade model, etc.

Linear Thresholds Model

- Call a node active if it has adopted the information
- Initially every node is inactive
- Let us consider a node i and represent its neighbors by the set $N(i)$
- Node i is influenced by a neighbor node j according to a weight w_{ij}. These weights are normalized in such a way that

$$\sum_{j \in N(i)} w_{ij} \leq 1.$$

- Further each node i chooses a threshold, say θ_i, uniformly at random from the interval [0,1]
- This threshold represents the weighted fraction of node i's neighbors that must become active in order for node i to become active

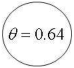
Given a random choice of thresholds and an initial set (call it S) of active nodes, the diffusion process propagates as follows:

- in time step t, all nodes that were active in step $(t-1)$ remain active
- we activate every node i for which the total weight of its active neighbors is at least θ_i
- if $A(i)$ is assumed to be the set of active neighbors of node i, then i gets activated if

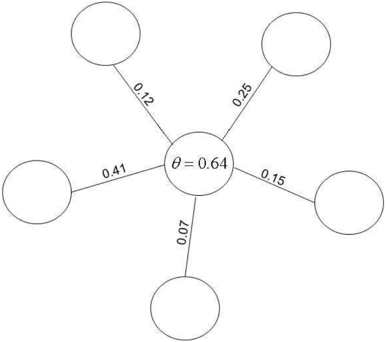
$$\sum_{j \in A(i)} w_{ij} \ \geq \ \theta_i .$$

- This process stops when there is no new active node in a particular time interval

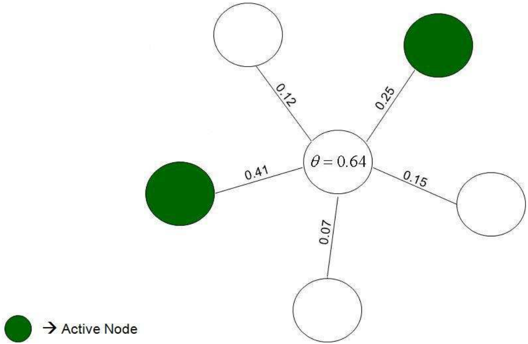
Linear Thresholds Model: An Example



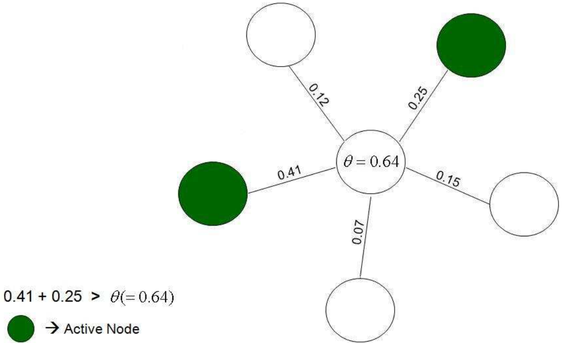
Linear Thresholds Model: An Example



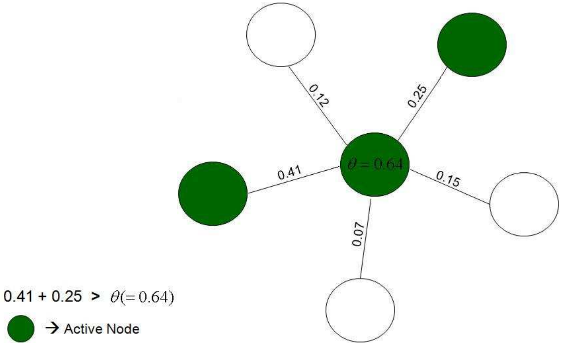
Linear Thresholds Model: An Example



Linear Thresholds Model: An Example



Linear Thresholds Model: An Example



Top-k Nodes Problem

- *Objective function ($\sigma(.)$):* Expected number of active nodes at the end of the diffusion process
- If S is the initial set of target nodes, then $\sigma(S)$ is the expected number of active nodes at the end of the diffusion process
- For economic reasons, we want to limit the size of S
- For a given constant k, the top-k nodes problem seeks to find a subset of nodes S of cardinality k that maximizes the expected value of $\sigma(S)$

Applications

- Databases
- Water Distribution Networks
- Blogspace
- Newsgroups
- Virus propagation networks

- R. Akbarinia, F.E. Pacitti, and F.P. Valduriez. Best Position Algorithms for Top-k Queries. In VLDB, 2007.
- J. Leskovec, A. Krause, and C. Guestrin. Cost-effective outbreak detection in networks. In ACM KDD, 2007.
- N. Agarwal, H. Liu, L. Tang, and P.S. Yu. Identifying influential bloggers in a community. In WSDM, 2008.

A Glimpse of State-of-the-Art

- P. Domingos and M. Richardson. Mining the network value of customers. In ACM SIGKDD, 2001.
  - Introduced this problem as an algorithmic problem
  - A model using Markov Ramom Fields
  - Show that selecting the right set of users for a marketing campaign can make a substantial difference.
- D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In SIGKDD, 2003.
  - Show that the optimization problem of selecting most influential nodes is NP-hard problem.
  - Show that this objective function is a sub-modular function.
  - Propose a greedy algorithm that achieves an approximation guarantee of $(1 - \frac{1}{e})$.

A Glimpse of State-of-the-Art (Cont.)

Greedy Algorithm - KKT (2003)

1. Set $A \leftarrow \phi$.
2. **for** $i = 1$ to k **do**
3. Choose a node $n_i \in N \setminus A$ maximizing $\sigma(A \cup \{n_i\})$
4. Set $A \leftarrow A \cup \{n_i\}$.
5. **end for**

- Running time of Greedy Algorithm: $O(knRm)$.

A Glimpse of State-of-the-Art (Cont.)

- J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In ACM SIGKDD, 2007.
  - Develop an efficient algorithm that is reportedly 700 times faster than the greedy algorithm (KKT (2003)).
  - There are two aspects to this speed up:
    1. Speeding up function evaluations using the sparsity of the underlying problem, and
    2. Reducing the number of function evaluations using the submodularity of the influence functions.

A Glimpse of State-of-the-Art (Cont.)

- W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In ACM SIGKDD, 2009.
- N. Chen. on the approximability of influence in social networks. In ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 1029-1037, 2008.
- M. Kimura and K. Saito. Tractable models for information diffusion in social networks. In PKDD, 2006.

Research Gaps

- All the existing approximation algorithms are sensitive to the number of initial trend setters (i.e. k)
- All the approximation algorithms crucially depend on the submodularity of the objective function. It is quite possible that the objective function can be non-submodular

Our Proposed Approach

- We present a cooperative game theoretic framework for the top-k nodes problem.
- We measure the influential capabilities of the nodes as provided by the Shapley value.
- ShaPley value based discovery of Influential Nodes (SPIN):
  1. Ranking the nodes,
  2. Choosing the top-k nodes from the ranking order.
- Advantages of SPIN:
  1. Quality of solution is same as that of popular benchmark approximation algorithms,
  2. Works well for both sub-modular and non-submodular objective functions,
  3. Running time is independent of the value of k.

Ranklist Construction

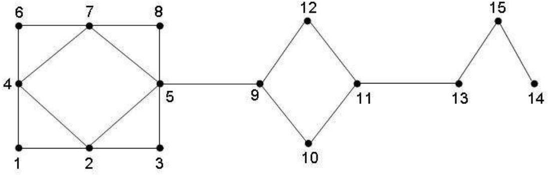
1. Let π_j be the j-th permutation in $\hat{\Omega}$ and R be repetitions.
2. Set $MC[i] \leftarrow 0$, for $i = 1, 2, \ldots, n$.
3. **for** $j = 1$ to t **do**
4. Set $temp[i] \leftarrow 0$, for $i = 1, 2, \ldots, n$.
5. **for** $r = 1$ to R, **do**
6. assign random thresholds to nodes;
7. **for** $i = 1$ to n, **do**
8. $temp[i] \leftarrow temp[i] + v(S_i(\pi_j) \cup \{i\}) - v(S_i(\pi_j))$
9. **for** $i = 1$ to n, **do**
10. $MC[i] \leftarrow temp[i]/R$;
11. **for** $i = 1$ to n, **do**
12. compute $\Phi[i] \leftarrow \frac{MC[i]}{t}$
13. Sort nodes based on the average marginal contributions of the nodes

Efficient Computation of Rank List

- Initially all nodes are inactive.
- Randomly assign a threshold to each node.
- Fix a permutation π and activate $\pi(1)$ to determine its influence.
- Next consider $\pi(2)$. If $\pi(2)$ is already activated, then the influence of $\pi(2)$ is 0. Otherwise, activate $\pi(2)$ to determine its influence.
- Continue up to $\pi(n)$.
- Repeat the above process R times (for example 10000 times) using the same π.
- Repeat the above process $\forall \pi \in \hat{\Omega}$.

Choosing Top-k Nodes

1. Naive approach is to choose the first k in the RankList[] as the top-k nodes.
2. *Drawback:* Nodes may be clustered.
3. RankList[]=$\{$5,4,2,7,11,15,9,13,12,10,6,14,3,1,8$\}$.
4. Top 4 nodes, namely $\{5, 4, 2, 7\}$, are clustered.
5. Choose nodes:
    - rank order of the nodes
    - spread over the network



k value	*Greedy Algorithm*	*Shapley Value Algorithm*	*MDH based Algorithm*	*HCH*
1	4	4	4	2
2	8	7	7	4
3	10	10	8	6
4	12	12	8	7
5	13	13	10	8
6	14	14	13	8
7	15	15	13	8
8	15	15	13	8
9	15	15	13	10
10	15	15	13	11
11	15	15	13	13
12	15	15	13	13
13	15	15	14	14
14	15	15	15	15
15	15	15	15	15

Running Time of SPIN

- Overall running time of SPIN is $O(t(n+m)R + n\, log(n) + kn + kRm)$ where t is a polynomial in n.
- For all practical graphs (or real world graphs), it is reasonable to assume that $n < m$. With this, the overall running time of the SPIN is $O(tmR)$ where t is a polynomial in n.

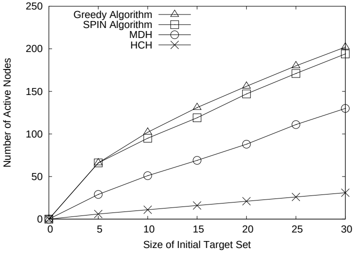
Experimental Results: Data Sets

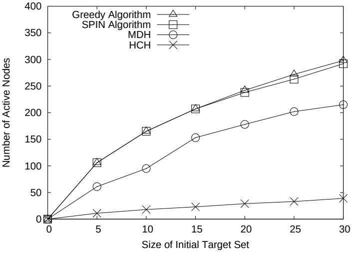
- Random Graphs
  - Sparse Random Graphs
  - Scale-free Networks (Preferential Attachment Model)
- Real World Graphs
  - Co-authorship networks,
  - Networks about co-purchasing patterns,
  - Friendship networks, etc.

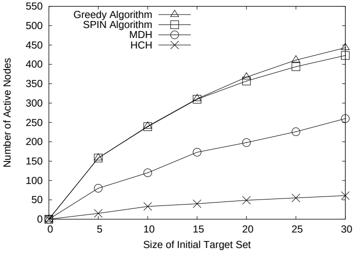
Experimental Results: Data Sets

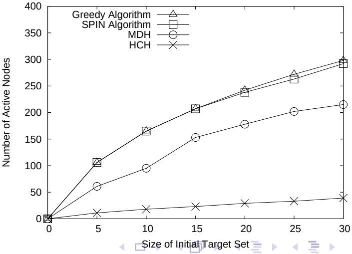
Dataset	Number of Nodes	Number of Edges
Sparse Random Graph	500	5000 (approx.)
Scale-free Graph	500	1250 (approx.)
Political Books	105	441
Jazz	198	2742
Celegans	306	2345
NIPS	1061	4160
Netscience	1589	2742
HEP	10748	52992

Experiments: Random Graphs

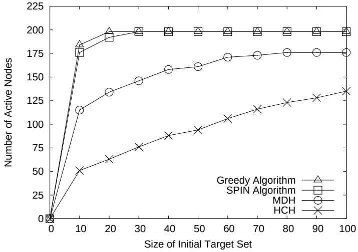


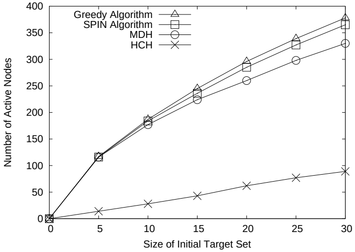


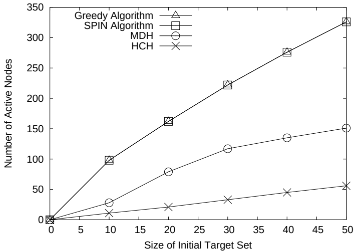


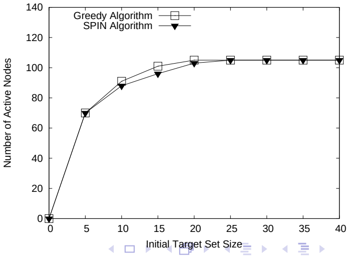


Experiments: Real World Graphs

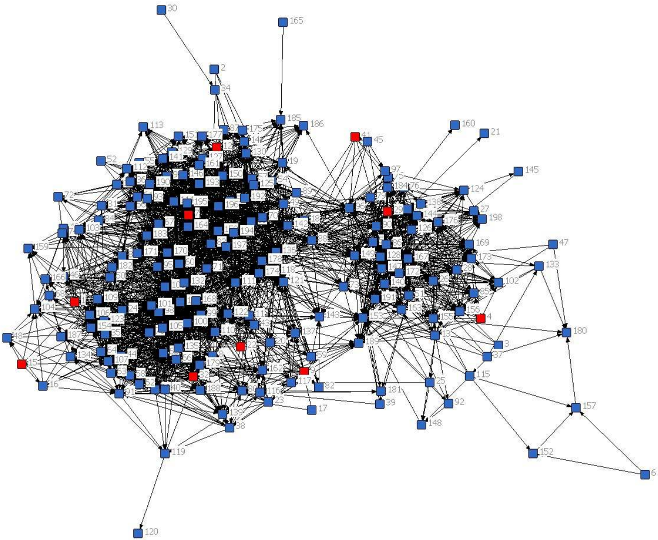




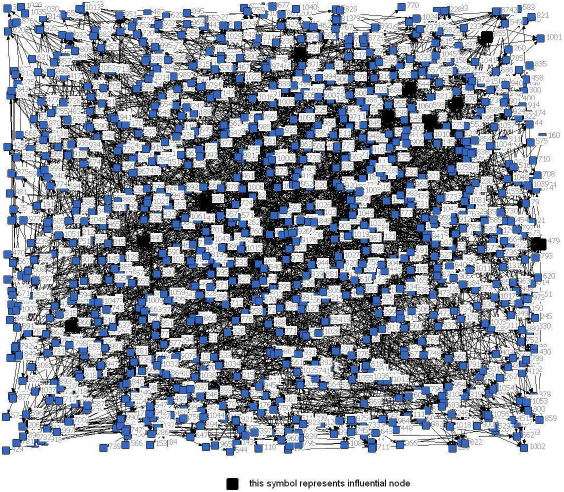




Top-10 Nodes in Jazz Dataset



Top-10 Nodes in NIPS Co-Authorship Data Set



Running Times: SPIN vs KKT

Dataset	Nodes	SPIN (MIN)	KKT (MIN)	Speed-up
Random graph ($p = 0.005$)	500	13.9	824.93	59
Random graph ($p = 0.01$)	500	14.8	1123.16	75
Random graph ($p = 0.02$)	500	16.3	1302.46	79
Political Books	105	0.89	44.64	50
Jazz	198	1.1	366	332
Celegans	306	14.02	901	64
NIPS	1062	15.2	7201.54	473
Network-Science	1589	28.25	8539.48	302

Table: Speedup of the SPIN algorithm to find Top-30 nodes on various datasets compared to that of KKT algorithm

Running Times: SPIN vs KKT

Top-k Nodes	Running Time (in MINUTES) SPIN Algorithm	KKT Algorithm	LKG Algorithm	Speed-up of SPIN over KKT
$k = 10$	28.04	1341.29	77.07	47
$k = 20$	28.09	4297.02	79.75	152
$k = 30$	28.13	8539.48	85.04	302
$k = 40$	28.18	13949.9	90.33	493
$k = 50$	28.25	20411.1	99.03	722

Table: Running times of the SPIN, KKT, and LKG algorithms on the Netscience data set ($n = 1589$) to determine top-k nodes where $k = 10, 20, 30, 40, 50$ and the speed up of the SPIN algorithm over the KKT algorithm

Viral Marketing: Extensions

- Viral Marketing with both Positive and Negative Opinions
- Viral Marketing with Competing Companies
- Viral Marketing with Multiple Independent Products
- Viral Marketing with Cross-sell of Products
- Many more . . .

Viral Marketing with Negative Opinions

- 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.

Viral Marketing: Competing Companies

- A company may introduce some product into the market when certain other company already has introduced a competing product into the market
- How to choose the initial seeds for viral marketing of products in the presence of competing products already present 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.
- T. Carnes, C. Nagarajan, S. Wild, A. van Zuylen. Maximizing influence in a competitive social network: a followers perspective. In the Proceedings of the 9th International Conference on Electronic Commerce (ICEC), pages 351360, 2007.

Viral Marketing with Multiple Independent Products

- Often companies introduce *multiple independent products* into the market
- Need to satisfy certain domain specific constraints while choosing the initial seeds for each of these products (Eg. no individual should be part of the set of initial seeds for more than certain number of products
- The procedure for selecting the initial seeds for each of the independent products is very different from that of viral marketing with single product
- S. Datta, A. Majumder, N. Shrivastava. Viral marketing for multiple products. In: Proceedings of IEEE ICDM, pages 118127, 2010

Viral Marketing with Cross-sell of Products

Viral Marketing with Product Dependencies

- Often cross-sell among products is possible
- 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 the proceedings of ECML-PKDD, Springer, 2012.

Main Results

- Proposed a novel influence propagation model to capture both the cross-sell phenomenon and product specific costs and benefits
- We refer to this problem as the B-IMCP problem
- Show that it is a NP-hard computationally
- Present a simple greedy approximation algorithm and then derive the approximation guarantee of this greedy algorithm by drawing upon the results from the theory of matroids
- Also outline two efficient heuristics based on well known concepts in the sociology literature
- Finally, several experimental results are presented based on proposed information propagation model and algorithms

Framework of the Model

- Let P_1 and P_2 be two sets of independent products consisting of t_1 and t_2 products respectively
- We propose to work with two scenarios for costs and benefits of the products, namely
- Assume that product specific costs and benefits for the products in P_1 and P_2 respectively; and
- Assume that unit cost and unit benefit for each product in P_1 and P_2 respectively

Framework of the Model (Cont.)

- We assume that cross-sell is possible from the products in P_1 to the products in P_2.

- We propose to work with two variants of the cross-sell, namely
  - First Variant of Cross-sell (Eg. Computers and Printers), and
  - Second Variant of Cross-sell (Eg. Machine learning books and Statistics books)

Framework of the Model (Cont.)

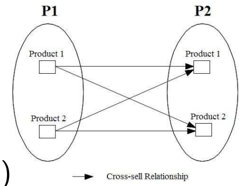
Based on the above setting, we have the following four model frameworks:

- First variant of cross-sell with variable costs and variable benefits
- First variant of cross-sell with unit costs and unit benefits
- Second variant of cross-sell with variable costs and variable benefits
- Second variant of cross-sell with unit costs and unit benefits

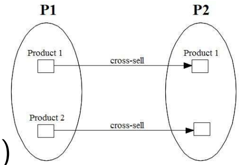
Experimental Results: Data Sets

Data Set	Number of Nodes	Number of Edges
WikiVote	7115	103689
HEP	10748	52992
Epinions	75879	508837
Telco Call Data	354700	368175

Experimental Results: Cross-sell Graphs



(i)



(ii)

Experimental Results (Cont.)

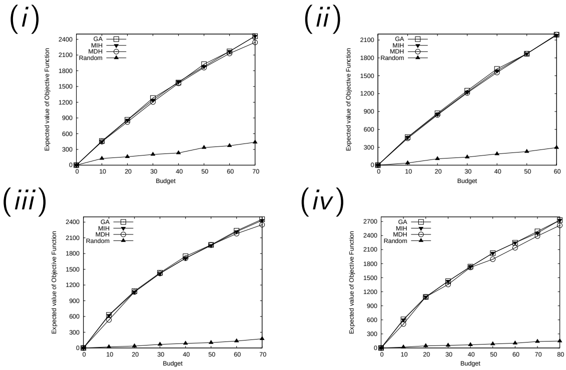


Figure: Given the first variant of cross-sell and complete bi-partite graph of the cross-sell relationships. Performance comparison of GA, MIH, MDH, and Random when (i) Data set is HEP and cross-sell threshold is $[0, 0.5]$, (ii) Data set is HEP and cross-sell threshold is $[0, 0.2]$, (iii) Data set is WikiVote and cross-sell threshold is $[0, 0.5]$, and (vi) Data set is WikiVote and cross-sell threshold is $[0, 0.2]$

Experimental Results (Cont.)

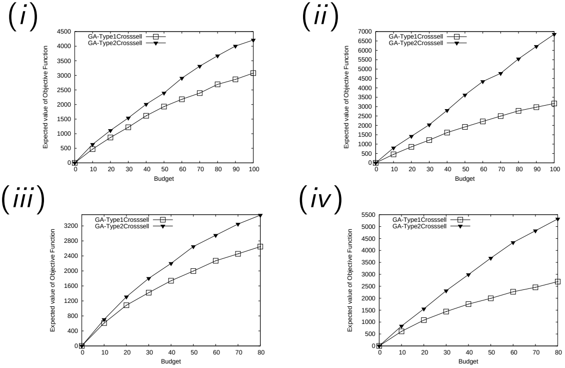


Figure: Performance comparison of GA using the first variant of cross-sell and the second variant of cross-sell when (i) Data set is HEP and cross-sell threshold is $[0, 0.5]$, (ii) Data set is HEP and cross-sell threshold is $[0, 0.2]$, (iii) Data set is WikiVote and cross-sell threshold is $[0, 0.5]$, and (vi) Data set is WikiVote and cross-sell threshold is $[0, 0.2]$

Outline of the Presentation

1. Social Network Analysis: Quick Primer
2. Foundational Concepts in Game Theory
3. Viral Marketing
4. **Community Detection in Social Networks**
5. Social Network Formation
6. Summary and To Probe Further

Key Resource of this Topic

- Ramasuri Narayanam and Y. Narahari. A Game Theory Inspired, Decentralized, Local Information based Algorithm for Community Detection in Social Graphs. To appear in 21st International Conference on Pattern Recognition (ICPR), 2012.

Community Detection in Social Networks

- Significant interest in the research community to detect communities in real-world graphs
- Community detection in networks is the division of network nodes into groups within which the network connections are dense, but between which the connections are sparser
- We use the terms *graph clustering* and *community detection in graphs* interchangeably

Community Detection in Social Networks (Cont.)

Several variants of community detection are possible:

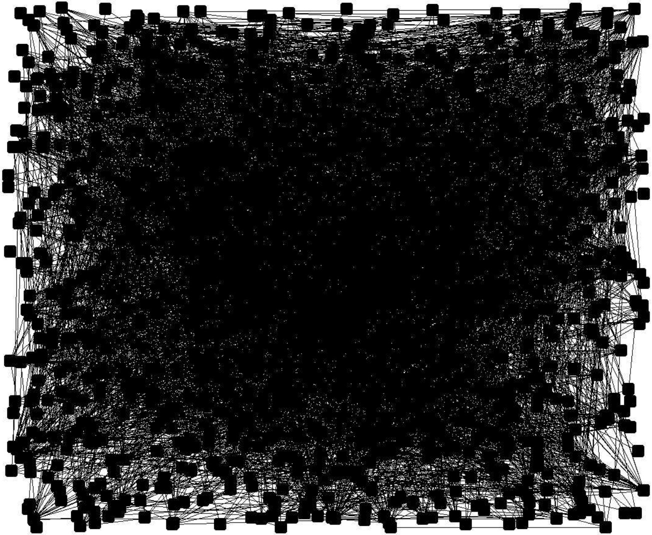
- Graph partition
- Dense sub-graph identification
- Over lapping or non-over lapping communities identification

Community Detection in Social Networks (Cont.)

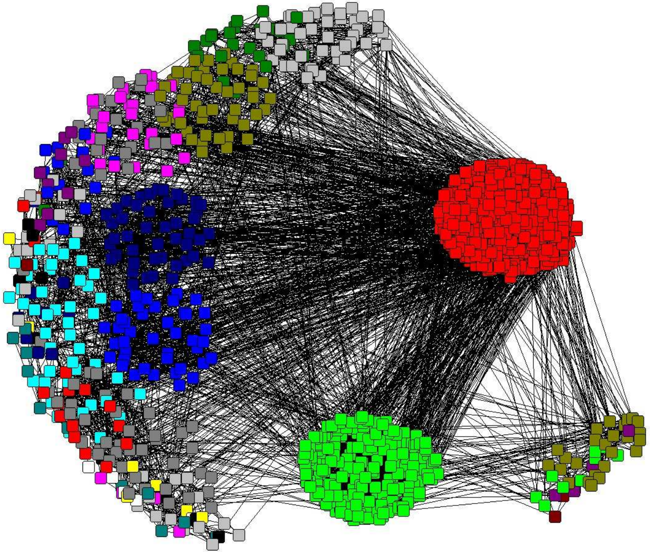
1. VLSI circuit design
2. Parallel computing
3. Social network analysis
4. Graph visualization and summarization

B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. The Bell System Technical Journal, 1970.

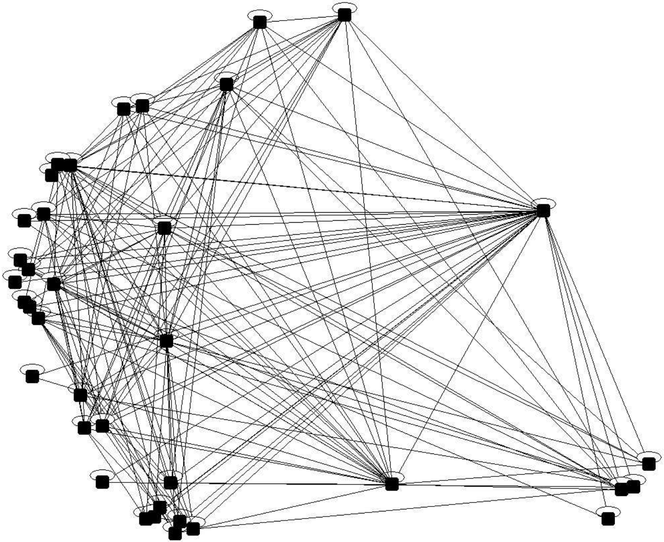
Graph Visualization and Summarization



Graph Visualization and Summarization



Graph Visualization and Summarization



Approaches for Community Detection

- Spectral Approach
- Geometric Approach
- Multi-level Approach
- Social Network Analysis Approach
    - Algorithms based on Centrality Measures
    - Algorithms based on Random Walks
    - Algorithms based on Optimization

1. G. Karypis and V. Kumar. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. SIAM Journal on Scientific Computing, 1997.
2. S. Arora, S. Rao, U.V. Vazirani. Geometry, flows, and graph-partitioning algorithms. Communications of ACM, 51(10):96-105, 2008.

Research Gap and Motivation

- First, many of these algorithms work with a global objective function (such as modularity, conductance) for detecting communities in networks
- Communities in social networks emerge due to the actions of autonomous individuals in the network, without regard to a central authority enforcing certain global objective function
- Thus, algorithms that work with global objective functions do not satisfactorily explain the emergence of communities in social networks
- Second, many algorithms for community detection require the number of communities to be detected as input

Community Detection Problem

- **Problem Statement:** Given an undirected and unweighted graph, we want to design a decentralized algorithm (due to autonomous actions of individuals in the network) that determines a partitioning of the graph with appropriate number of groups (or communities) such that each group in that partition is *dense*

- We propose a game theoretic approach to address the problem

Challenges Involved (Also Wish List)

- Algorithm should use only local information
- Existence of an equilibrium to be guaranteed
- Communities should be dense at equilibrium
- Algorithm itself should determine the number of communities

Community Detection Problem: Our Contributions

- Game theoretic framework based on Nash stable partition
- Propose a utility function based on only local information - it guarantees the existence of a Nash stable partition
- Lower bound on the coverage of any Nash stable partition
- Equivalence of NSP with another popular notion of equilibrium partition under the proposed utility function
- Leads to an efficient algorithm
  - NASHCoDe (Nash Stable Community Detection Algorithm)
- NASHCoDe does not require the number of communities as input

Graph Partitioning: Basic Definitions

- Let $G = (N, E)$ be an undirected and unweighted graph where $N = \{1, 2, \ldots, n\}$ is the finite set of nodes and E is the set of edges
- N_i is the set of all neighbors of i including i
- $A_i = \{H \mid H \textit{ is a subgraph of } G \textit{ and } i \in H\}$
- (N, u_i) is the graph partitioning game where N is the set of nodes in G and $u_i : A_i \rightarrow \mathbb{R}$ is the utility of node i
- Ψ is the set of all partitions
- $\Pi(i) = \{S \mid S \in \Pi \textit{ and } S \cap N_i \neq 0\}$

Graph Partitioning: Basic Definitions

Nash Stable Partition (NSP): A partition $\Pi \in \Psi$ is called a Nash stable partition of the given graph if $\forall i \in N$,

$$u_i(S_\Pi(i), G) \geq u_i(S_k \cup \{i\}, G),\ \forall S_k \in \Pi(i).$$

A. Bogomolnaia and M.O. Jackson. The stability of hedonic coalition structures. Games and Economic Behavior, 38:201-230, 2002.

Proposed Utility Function

- Utility function should use only local information
- It should ensure the existence of NSP
- It should produce dense communities at NSP
- For each $i \in N$, the utility of node i is defined to be the number of neighbors of that node in its community plus a function of the fraction of neighbors in its community that are connected themselves

Proposed Utility Function (Cont.)

- For each $i \in N$ and $\forall S \in A_i$,

$$u_i(S) = d_i(S) + \frac{T_i(S)}{\binom{d_i(S)}{2}} f(d_i(S))$$

  where (i) $d_i(S)$ is the number of neighbors of node i in community S, and (ii) $T_i(S)$ is number of pairs of neighbors of node i in S that are connected themselves, and (iii) $f(d_i(S))$ is a weight function

- To keep matters simple, we consider linear weight function $f(.)$ such as:

$$f(d_i(S)) = \alpha d_i(S), \ \forall i \in N, \ \forall S \in A_i$$

  where $\alpha > 0$ is a constant

- To keep matters further simple, we work with $\alpha = 1$ and $\alpha = 2$

Proposed Utility Function (Cont.)

- When $\alpha = 1$, for each $i \in N$ and $\forall S \in A_i$,

$$u_i(S) = d_i(S) + \frac{2T_i(S)}{d_i(S) - 1}$$

- When $\alpha = 2$, for each $i \in N$ and $\forall S \in A_i$,

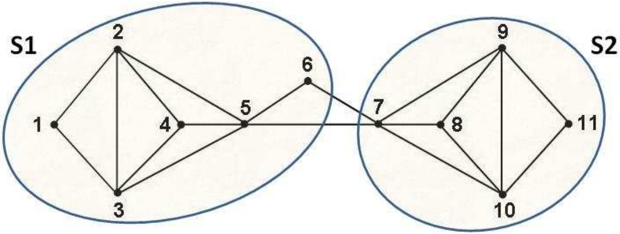
$$u_i(S) = d_i(S) + \frac{4T_i(S)}{d_i(S) - 1}$$

Proposed Utility Function (Cont.)

- **Lemma:** When $\alpha = 1$, it holds that $u_i(S) \leq u_i(\bar{S})$ for each $S \subseteq \bar{S}$.
  - This can be proved using simple algebra
  - The utility of node i does not decrease when new nodes join its community
  - Suitable for certain practical applications such as collaborative editing in Wikipedia

- **Lemma:** When $\alpha = 2$, for each $S \subseteq \bar{S}$, then $u_i(S) \leq u_i(\bar{S})$ holds only if $T_i(S) \leq \frac{d_i(S)(d_i(S)-1)}{4}$, $\forall i \in S$
  - The utility of node i need not necessarily increase when new nodes join its community
  - Suitable for certain practical settings such as collaboration networks

Nash Stable Partition: An Example



$\Pi = \{S1, S2\}$ where $S1 = \{1, 2, 3, 4, 5, 6\}$, and $S2 = \{7, 8, 9, 10, 11\}$

$u_1(S1) = 4, \quad u_1(S2) = 0;$
$u_2(S1) = 6.668, \quad u_2(S2) = 0;$
$u_3(S1) = 6.66, \quad u_3(S2) = 0;$
$u_4(S1) = 6, \quad u_4(S2) = 0;$
$u_5(S1) = 6, \quad u_5(S2) = 1;$
$u_6(S1) = 1, \quad u_6(S2) = 1;$

$u_7(S2) = 6, \quad u_7(S1) = 1;$
$u_8(S2) = 6, \quad u_8(S1) = 0;$
$u_9(S2) = 6.66, \quad u_9(S1) = 0;$
$u_{10}(S2) = 6.66, \quad u_{10}(S1) = 0;$
$u_{11}(S2) = 4, \quad u_{11}(S1) = 0;$

Important Properties of NSPs

- For each $i \in N$, let $\Psi(\Pi, i)$ be the set of all partitions where each partition is derived from Π by moving node i from $S_\Pi(i)$ to some $X \in \Pi(i)$. That is, $|\Psi(\Pi, i)| = |\Pi(i)| - 1$, (because $S_\Pi(i) \in \Pi(i)$).
- Let $INTRA(\Pi) = \{(i,j) \in E \mid \exists S \in \Pi \ni \ i,j \in S\}$ be the set of edges within communities in a partition Π. We also define $INTER(\Pi) = E \setminus INTRA(\Pi)$ to be the set of edges across communities in partition Π.
- Consider $i, j, k \in N$. If $(i,j) \in E$, $(j,k) \in E$, and $(k,i) \in E$, then we say that nodes i, j, k form a *triangle* in the graph.
- Finally, let $E(S)$ be the set of all edges among nodes in S only. That is, $E(S) = \{(i,j) \in E \mid i,j \in S\}$.

Important Properties of NSPs (Cont.)

$$\boxed{coverage(\Pi) = \frac{INTRA(\Pi)}{INTRA(\Pi)+INTER(\Pi)}}$$

Lemma: Assume that $\alpha = 1$. Given an undirected and unweighted graph $G = (N, E)$ and any Nash stable partition $\Pi = \{S_1, S_2\}$ with two communities of G, then $coverage(\Pi) \geq \frac{1}{3}$.

Proof Sketch:

- For each $i \in S_1$, we have that

$$d_i(S_2) + \frac{2T_i(S_2)}{d_i(S_2) - 1} \leq d_i(S_1) + \frac{2T_i(S_1)}{d_i(S_1) - 1} \quad (1)$$

- Now using fact that $|INTER(\Pi)| = \sum_{i \in S_1} d_i(S_2)$, we get that

$$|INTER(\Pi)| + \sum_{i \in S_1} \frac{2T_i(S_2)}{d_i(S_2) - 1} \leq \sum_{i \in S_1} d_i(S_1) + \sum_{i \in S_1} \frac{2T_i(S_1)}{d_i(S_1) - 1}. \quad (2)$$

Important Properties of NSPs (Cont.)

- On similar lines, repeating the above argument for S_2, we get that

$$|INTER(\Pi)| + \sum_{i \in S_2} \frac{2T_i(S_1)}{d_i(S_1) - 1} \leq \sum_{i \in S_2} d_i(S_2) + \sum_{i \in S_2} \frac{2T_i(S_2)}{d_i(S_2) - 1} \quad (3)$$

- After summing the expressions (2) and (3), we use the fact that $\sum_{i \in S_1} d_i(S_1) = 2|E(S_1)|$, $\sum_{i \in S_2} d_i(S_2) = 2|E(S_2)|$, and rearranging terms, we get that

$$\sum_{i \in S_1} \frac{T_i(S_2)}{d_i(S_2) - 1} + \sum_{i \in S_2} \frac{T_i(S_1)}{d_i(S_1) - 1} + |INTER(\Pi)| \quad \leq$$

$$|E(S_1)| + |E(S_2)| + \sum_{i \in S_1} \frac{T_i(S_1)}{d_i(S_1) - 1} + \sum_{i \in S_2} \frac{T_i(S_2)}{d_i(S_2) - 1}$$

Important Properties of NSPs (Cont.)

- Now substituting $INTER(\Pi) = E \setminus INTRA(\Pi)$ and $E(S_1) \cup E(S_2) = INTRA(\Pi)$ in the above expression and readjusting the terms, we get that

$$\Rightarrow \ |INTRA(\Pi)| \geq \frac{1}{2}|E| - \frac{1}{2}\left[\sum_{i \in S_1} \frac{T_i(S_1)}{d_i(S_1) - 1} + \sum_{i \in S_2} \frac{T_i(S_2)}{d_i(S_2) - 1}\right]$$

- Since the maximum value for $T_i(X)$ is $\frac{d_i(X)(d_i(X)-1)}{2}$ for each $X \in \{S_1, S_2\}$; and bounding the above expression using this fact, we get that

$$|INTRA(\Pi)| \geq \frac{1}{2}|E| - \frac{1}{2}\left[\sum_{i \in S_1} \frac{d_i(S_1)}{2} + \sum_{i \in S_2} \frac{d_i(S_2)}{2}\right]$$

$$\Rightarrow \ |INTRA(\Pi)| \geq \frac{1}{2}|E| - \frac{1}{2}|INTRA(\Pi)| \ \Rightarrow coverage(\Pi) \geq \frac{1}{3}.$$

Important Properties of NSPs (Cont.)

- **Lemma:** Assume that $\alpha = 2$. Given an undirected and unweighted graph $G = (N, E)$ and a Nash stable partition $\Pi = \{S_1, S_2\}$ with 2 communities of G, then $coverage(\Pi) \geq \frac{1}{4}$.

An Algorithm for Graph Partitioning

- NASHCoDe: Nash Stability based Community Detection
  - Initial configuration
  - Order of nodes
  - Stopping criterion
- Running time : $O(n\ log(n) + nmd^2_{max})$
- The resultant community structure with NASHCoDe is obviously a Nash stable partition

NASHCoDe

Let Π be the initial partition of the graph G.
Let *visit_order*[] contains nodes in non-decreasing order of the degree.
while *true*, *do*
  for $i := 1$ to n
    flag $\leftarrow 0$;
    $j \leftarrow visit\_order[i]$
    if $u_i(S_\Pi(i)) < u_i(S_k)$ for some $S_k \in \Pi(i)$, *then*
      move node i from $S_\Pi(i)$ to S_k;
      flag $\leftarrow 1$;
    end if
  end for
  if flag $= 0$, *then*
    break;
  end if
end while

Existence of NSP

Lemma: When $\alpha = 1$, Algorithm 1 always guarantees convergence to a Nash stable partition, given any undirected and unweighted graph.

- Given that $\alpha = 1$. We first define a function $\Phi : \Psi \to \mathbb{N}$ such that, for each $\Pi \in \Psi$, $\Phi(\Pi)$ represents the sum of the number of intra community edges in Π and the number of triangles within the communities in Π.
- For each $\Pi \in \Psi$, we call $\Phi(\Pi)$ as the *capacity* of the partition Π. More formally, $\forall \Pi \in \Psi$,

$$\Phi(\Pi) = \sum_{S \in \Pi} \sum_{j \in S} \frac{d_j(S)}{2} + \sum_{S \in \Pi} \sum_{i \in S} \sum_{p,q \in S \cap N_i} \frac{I(p,q)}{3} \qquad (4)$$

where $I(p,q)$ is an indicator function that takes value 1 if nodes p and q are adjacent in G and 0 otherwise.

Existence of NSP (Cont.)

- Define a partition $\Pi_x \in \Psi$ to be *maximal* if there *does not exist* any $\Pi_y \in \Psi$ such that (i) $\Pi_y \in \Psi(\Pi_x, i)$ for some $i \in N$, and (ii) the capacity of Π_y is strictly greater than the capacity of Π_x.
- Now consider a partition $\Pi_1 \in \Psi$. Let $\Pi_2 \in \Psi(\Pi_1, i)$ be a partition that is obtained from Π_1, when node i jumps from $S_{\Pi_1}(i)$ to some $X \in \Pi_1(i)$.
- Note that node i jumps from $S_{\Pi_1}(i)$ to some $X \in \Pi_1(i)$ to improve its utility, then

$$d_i(S_{\Pi_1}(i)) + \frac{2T_i(S_{\Pi_1}(i))}{d_i(S_{\Pi_1}(i)) - 1} < d_i(X) + \frac{2T_i(X)}{d_i(X) - 1}. \tag{5}$$

Existence of NSP (Cont.)

- This is possible when either $d_i(X) > d_i(S_{\Pi_1}(i))$ or $T_i(X) > T_i(S_{\Pi_1}(i))$ holds.
- A simple algebra based on this fact implies that $\Phi(\Pi_2) > \Phi(\Pi_1)$.
- This further implies that, whenever a node moves from one group to the other group, the capacity of the new partition strictly improves upon the capacity of the old partition.
- Since the number of partitions is finite, a maximal partition certainly exists.

Improved Version of NASHCoDe

- We further refine the Nash stable partition produced through NASHCoDe to improve its modularity
- *Greedy Strategy:* In each step, we determine a pair of communities to merge so that the modularity of the resultant community structure is maximized. We repeat this step until we do not find any pair of communities to merge to improve modularity
- Running time of the improved version of NASHCoDe is $O(n\ log(n) + nmd_{max}^2 + k^3 n)$

Benchmark Algorithms for Comparison

1. Edge Betweenness Algorithm (Girvan and Newman (2002)),
2. Fast Greedy Algorithm (Newman (2004)),
3. Spectral Algorithm (Newman (2006)).
4. A Randomized and Game Theoretic Algorithm (Chen et. al. (2010))

Performance Metric

Modularity of a partition Π of given undirected and unweighted graph $G = (N, E)$ is:

$$Q(\Pi, G) = \frac{1}{2m} \sum_{i,j \in N} \left(a_{i,j} - \frac{n_i n_j}{2m} \right) \delta(S_\Pi(i), S_\Pi(j))$$

where (i) m is number of edges in G, (ii) n_x is degree of node $x \in N$, (iii) $S_\Pi(x)$ represents ID of the community of node x, and (iv) $\delta(a, b)$ takes value 1 if $a = b$ and 0 otherwise.

M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. Physical Review E 69, 026113, 2004.

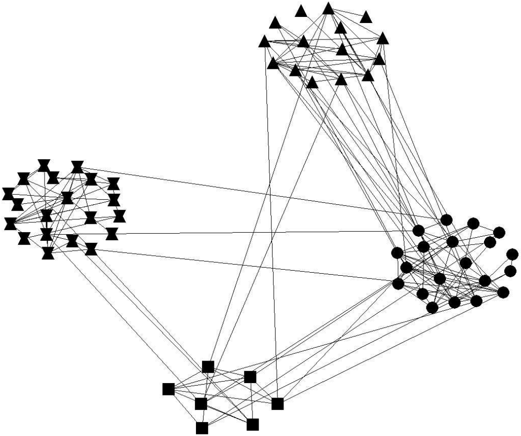
Experimental Results: Data Sets

Data Set	*Nodes*	*Edges*	*Triangles*
Karate	34	78	45
Dolphins	62	318	95
Les Miserables	77	508	467
Political Books	105	882	560
FootBall	115	1226	810
Jazz	198	2742	17899
Email	1133	5451	10687
Yeast	2361	6913	5999

Table: Description of various real world network data sets

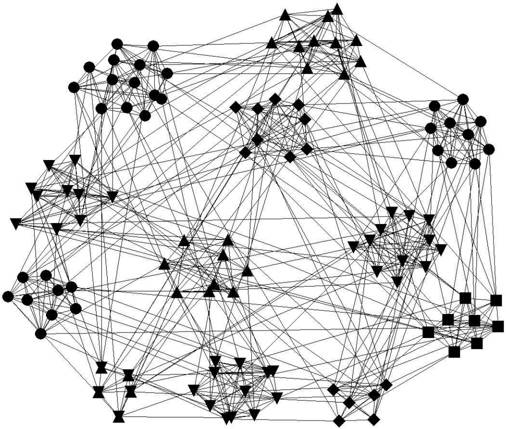
Comparison with GN Algorithm: Dolphins Data Set

- Modularity using GN Algorithm: 0.519.
- Modularity using NASHCoDe: 0.526.



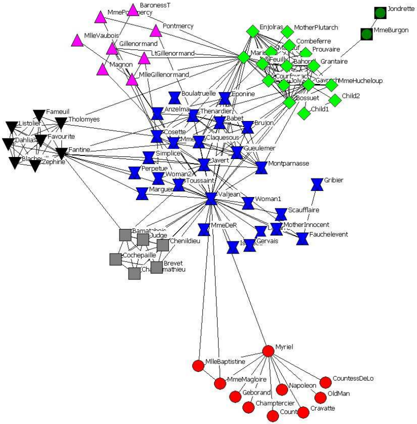
Comparison with GN Algorithm: Football Data Set

- Modularity using GN Algorithm: 0.598.
- Modularity using NASHCoDe: 0.6.



Comparison with GN Algorithm: Les Miserable Data Set

- Modularity using GN Algorithm: 0.538.
- Modularity using NASHCoDe: 0.538.



Experimental Results - Modularity

Table: Comparison of modularity obtained using NASHCoDe with that of four benchmark algorithms

Data Set	NASHCoDe ($\alpha = 1$)	NASHCoDe ($\alpha = 2$)	Greedy	Spectral	RGT
Karate	0.4	0.4	0.38	0.393	0.392
Dolphins	0.525	0.525	0.495	0.491	0.502
Les-Mis	0.545	0.545	0.5	0.532	0.54
Pol-Books	0.524	0.524	0.509	0.469	0.493
Football	0.60	0.60	0.566	0.539	0.581
Jazz	0.439	0.439	0.438	0.393	0.439
Email	0.479	0.51	0.494	0.498	0.509
Yeast	0.571	0.571	0.571	0.497	-

Experimental Results - Coverage

Table: Comparison of coverage obtained using NASHCoDe with that of four benchmark algorithms

Data Set	**NASHCoDe ($\alpha = 1$)**	**NASHCoDe ($\alpha = 2$)**	**Greedy**	**Spectral**	**RGT**
Karate	82.05	82.05	30.76	25.24	68.52
Dolphins	80.50	80.50	22.01	22.64	69.43
Les-Mis	75.19	75.19	35.03	25.19	72.83
Pol-Books	89.11	89.11	59.63	45.57	74.45
Football	69.00	69.00	16.15	12.39	67.92
Jazz	79.83	79.83	31.72	34.31	71.35
Email	72.46	73.18	21.61	18.33	58.12
Yeast	66.99	67.85	19.53	22.14	-

Next Part of the Presentation

Reference

Ramasuri Narayanam and Y. Narahari. *Topologies of strategically formed social networks based on a generic value function – Allocation rule model*. **Social Networks**, Volume 33, Number 1, 2011.

Social Network Formation

In several social and economic situations, the specifics of network structure play crucial role in determining the outcome:

- Social networks play an important role in viral marketing.
- Personal contacts in friendship networks play critical role in job finding.
- Networks play important roles in the trade and exchange of goods in decentralized markets.
- Research and development alliances among corporations.

Social Network Formation (Cont.)

- In the process of information dissemination or in general value creation, individuals receive not only various kinds of benefits but also incur costs in terms of time, money, and effort.

Social Network Formation (Cont.)

- In the process of information dissemination or in general value creation, individuals receive not only various kinds of benefits but also incur costs in terms of time, money, and effort.
- Hence individual nodes do act strategically.

Social Network Formation (Cont.)

- In the process of information dissemination or in general value creation, individuals receive not only various kinds of benefits but also incur costs in terms of time, money, and effort.
- Hence individual nodes do act strategically.
- It is essential to study:
  - How to model the formation of social networks in the presence of strategic individuals that are interested in maximizing their payoffs from the social interactions?
  - What are the networks that will emerge due to the dynamics of network formation and what their characteristics are likely to be?

A General Approach

Game theoretic models are a natural choice over random graph models:

- Nodes often act strategically as link formation involves both benefits and costs
- More often links form by choice not by chance.
- Capture social and economic incentives while forming or severing links.

A Glimpse of State-of-the-Art

- M. O. Jackson. Social and Economic Networks. Princeton University Press, Princeton and Oxford, 2008.
- S. Goyal. Connections: An Introduction to the Economics of Networks. Princeton University Press, Princeton and Oxford, 2007.
- G. Demange and M. Wooders. Group Formation in Economics: Networks, Clubs, and Coalitions. Cambridge University Press, Cambridge and New York, 2005.
- M. Slikker and A. van den Nouweland. Social and Economic Networks in Cooperative Game Theory. Kluwer Academic Publishers, Massachusetts, USA and The Netherlands, 2001.

A Glimpse of State-of-the-Art (Cont.)

- M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. Journal of Economic Theory, 71(1):44-74, 1996.
- S. Goyal and F. Vega-Redondo. Structural holes in social networks. Journal of Economic Thoery, 137(1):460-492, 2007.
- V. Buskens and A. van de Rijt. Dynamics of networks if everyone strives for structural holes. American Journal of Sociology, 114(2):371-407, 2008.
- J. Kleinberg, S. Suri, E. Tardos, and T. Wexler. Strategic network formation with structural holes. In Proceedings of the 9th ACM Conference on Electronic Commerce (EC), pages 284-293, 2008.

Key Observations

- Several game theoretic models are proposed to capture the rational behavior of nodes.
- Some of these studies are able to yield sharp predictions on the network topologies that emerge, if stability and efficiency are to be satisfied.
- Various notions of stability and efficiency have been employed.
- Significant emphasis on the tradeoffs between stability and efficiency.

Research Gaps

- Existing models do not simultaneously capture all major determinants of network formation
- Significant gaps between analytical predictions and experimental findings

1. P. Doreian. Actor network utilities and network evolution. Social Networks, 28:137164, 2006.
2. N.P. Hummon. Utility and dynamic social networks. Social Networks, 22:221-249, 2000.

Stability

Stability: A network is stable if it is in a strategic equilibrium.
Examples: Nash equilibrium, Pairwise stability.

Nash Equilibrium: A network is said to be in Nash equilibrium if no node unilaterally forms or deletes a link to any other node.

Pairwise Stability: A network is said to be pairwise stable if

- if no node gains by deleting a link to any other node
- no pair of nodes wants to add a link between them

M.O. Jackson and A. Wolinsky. A strategic model of social and economic networks. In Journal of Economic Theory, 71:44–74, 1996.

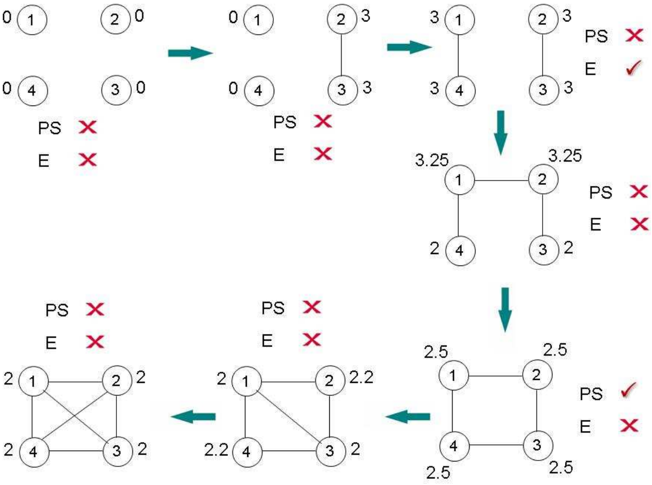
Efficiency

Efficiency: A network is efficient if a function of the utilities of the nodes is maximized.
Example: Pareto efficiency, Maximize sum of utilities.

Pareto Efficiency: A network g is said to be Pareto efficient if there is no network g' in which the utility of at least one node is strictly greater than that of in g and the utilities of the rest of the nodes are greater than or equal to that of in g.

Sum of Utilities: A network g is said to be efficient if the sum of utilities of the nodes in g is greater than or equal to that of any other network.

Stability and Efficiency Tradeoffs



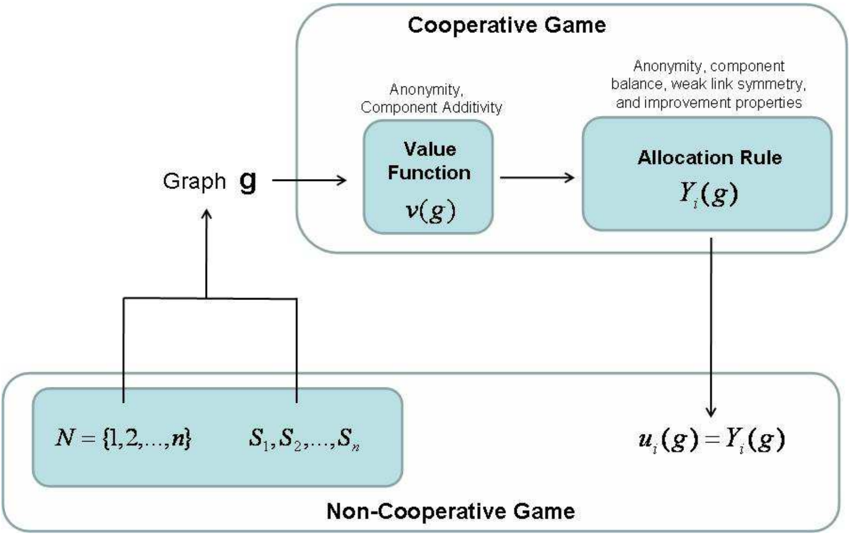
Our Contributions

1. Our study explores a generic *Value Function - Allocation Rule* model to capture the dynamics of network formation.
2. The value function considers 4 key determinants of network formation: (i) benefits, (ii) costs, (iii) decaying benefits from non-neighbors, and (iv) bridging benefits.
3. An axiomatic allocation rule is utilized. The Myerson value belongs to this class of allocation rules.
4. We characterize the structures of both socially efficient networks as well as pairwise stable networks.
5. Under the proposed model, we state and prove a necessary and sufficient condition for any efficient network to be pairwise stable.
6. We conduct experiments in some specific settings, leading to the unravelling of more specific topological possibilities.

Where Does Our Model Stand?

Model	Benefits	Costs	Decaying Benefits	Bridging Benefits	Analysis of Stability	Analysis of Efficiency
JW (1996)	✓	✓	✓	×	✓	✓
Jackson (2005)	✓	✓	✓	✓	×	×
GV (2007)	✓	✓	×	✓	✓	✓
BR (2008)	×	✓	✓ [Length 2]	×	✓	✓
KSTW (2008)	✓	✓	×	✓ [Length 2]	✓	×
Our Model	✓	✓	✓	✓	✓	✓

The Model


Cooperative Game
Anonymity, Component Additivity
Anonymity, component balance, weak link symmetry, and improvement properties
Graph g
Value Function
$v(g)$
Allocation Rule
$Y_i(g)$
$N = \{1,2,\ldots,n\}$
$S_1, S_2, \ldots, S_n$
$u_i(g) = Y_i(g)$
Non-Cooperative Game

The Model

- Let $N = \{1, 2, \ldots, n\}$ be the set of n (≥ 3) nodes.
- A strategy s_i of a node i is any subset of nodes with which it establishes links.
- Links are formed under mutual consent.
- S_i is the set of strategies of node i.
- Each $s = (s_1, s_2, \ldots, s_n)$ leads to an undirected graph and we represent it by $g(s)$.
- Let $\Psi(S)$ be the set of all such undirected graphs.
- When the context is clear, we use g and Ψ instead of $g(s)$ and $\Psi(S)$ respectively.

The Model (Cont.)

- $\forall i, j \in N$, $d_g(i,j)$ = length of shortest path between i and j.
- *Costs:* If nodes i and j are connected by a link in g, then we assume that the link incurs a cost $c > 0$.
- *Benefits:* The communication between i and j leads to a benefit of $b(d_g(i,j))$.
- We assume that $b(.)$ is a non-increasing function, implying that the benefit of communication decays as the length of shortest path increases.
- A value function $v : \Psi \to \mathbb{R}$ for a given graph $g \in \Psi$ is as follows:

$$v(g) = \sum_{\substack{x,y \in N, \\ (x,y) \in g}} [b(1) - c] + \sum_{i \in N} \sum_{\substack{j \in N, \\ j > i, \\ (i,j) \notin g}} b(d_g(i,j)) \tag{6}$$

The Model (Cont.)

- **Lemma:** The proposed value function $v(.)$ satisfies anonymity and component additivity.
- The network value $v(g)$ is divided among the nodes in g as utilities using an allocation rule. Allocation rule $Y : \Psi \to \mathbb{R}^n$ distributes the network value $v(g)$ among nodes as utilities such that

$$\sum_{i \in N} Y_i(g) = v(g), \ \forall g \in \Psi.$$

- We define $Y_i(g)$ to be the utility $(u_i(g))$ of node i.

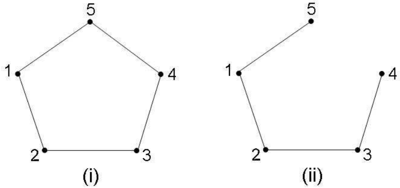
$$u_i(g) = Y_i(g), \ \forall i \in N.$$

- This framework clearly defines a strategic form game: $\Gamma = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$.

Axiomatic Allocation Rule

- **Anonymity:** Allocation rule Y is anonymous if for any v, $g \in \Psi$, and any permutation of the players π, $Y_{\pi(i)}(g^{\pi}) = Y_i(g)$.
- **Component Balance:** Allocation rule Y is *component balanced* if $\sum_{i \in C} Y_i(g) = v(C)$ for each component additive v, $g \in \Psi$, and for each component C in $\Omega(g)$ (the set of all components of the graph g).
- **Weak Link Symmetry:** Allocation rule Y satisfies *weak link symmetry* if for each link $e = (i, j) \notin g$, it holds that if $Y_i(g \cup \{e\}) > Y_i(g)$, then $Y_j(g \cup \{e\}) > Y_j(g)$.
- **Improvement Property:** Allocation rule Y satisfies *improvement property* if for each link $e = (i, j) \notin g$, whenever there exists a node $z \in N \setminus \{i, j\}$ such that $Y_z(g \cup \{e\}) > Y_z(g)$, then $Y_i(g \cup \{e\}) > Y_i(g)$ or $Y_j(g \cup \{e\}) > Y_j(g)$.

An Illustrative Example



$$v(g) = 5(b(1) - c) + 5b(2),$$
$$u_i(g) = (b(1) - c) + b(2) \quad \forall i \in \{1, 2, 3, 4, 5\}.$$

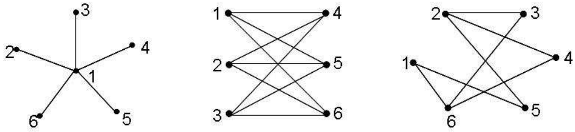
$$v(g') = 4(b(1) - c) + 3b(2) + 2b(3) + b(4),$$
$$u_1(g') = u_3(g') = (b(1) - c) + \tfrac{2}{3}b(2) + \tfrac{1}{2}b(3) + \tfrac{1}{5}b(4),$$
$$u_2(g') = (b(1) - c) + b(2) + \tfrac{1}{2}b(3) + \tfrac{1}{5}b(4),$$
$$u_4(g') = u_5(g') = \tfrac{1}{2}(b(1) - c) + \tfrac{1}{3}b(2) + \tfrac{1}{4}b(3) + \tfrac{1}{5}b(4).$$

Pairwise Stability and Efficiency

- **Pairwise Stability:** A network g is said to be *pairwise stable* with respect to the value function v and the allocation rule Y if (i) for each edge $e = (i, j) \in g$, $Y_i(g) \geq Y_i(g \setminus \{e\})$ and $Y_j(g) \geq Y_j(g \setminus \{e\})$, and (ii) for each edge $e' = (i, j) \notin g$, if $Y_i(g) < Y_i(g \cup \{e'\})$ then $Y_j(g) > Y_j(g \cup \{e'\})$.
- **Efficiency:** A network $g \in \Psi$ is said to be efficient if $v(g) \geq v(g') \;\; \forall g' \in \Psi$.

Minimal Edge Graphs with Diameter p $(1 < p < n)$

- **Definition:** The diameter of a graph is the length of a longest shortest path between any two vertices of the graph.
- **Definition:** A graph with diameter p is said to be a *minimal edge graph with diameter* p if the deletion of any edge in the graph results in a graph with diameter greater than p.
- Given a set of n nodes, there may be multiple minimal edge graphs with diameter p for $1 < p < n$. For example, the following figure shows three different minimal edge graphs with diameter 2.



(i) (ii) (iii)

Analysis of Efficient Networks

The following two results useful in characterizing the topologies of efficient networks.

Lemma: Given a graph g, if $(b(1) - b(2)) < c < (b(1) - b(3))$ and there exists a pair of nodes x and y such that $d_g(x, y) > 2$, then forming a link between x and y strictly increases the value of g.

▸ Proof

Lemma: If $(b(1) - b(2)) < c < (b(1) - b(3))$, then every efficient network is a minimal edge graph with diameter 2.

▸ Proof

Analysis of Efficient Networks (Cont.)

Theorem: Following our proposed model,

(i) if $c < (b(1) - b(2))$, then the complete graph is the unique topology possible for an efficient network

(ii) if $(b(1) - b(2)) < c \leq b(1) + (\frac{n-2}{2})b(2)$, then the star network is the unique topology possible for an efficient network

(iii) if $c > b(1) + (\frac{n-2}{2})b(2)$, then the only efficient network is the empty graph.

Analysis of Pairwise Stable Networks

A few useful results are as follows.

- **Lemma:** For any graph g, if a pair of non-neighbor nodes i and j form a link (i, j) such that $v(g \cup \{(i, j)\}) > v(g)$, then it holds that both $Y_i(g \cup \{(i, j)\}) > Y_i(g)$ and $Y_j(g \cup \{(i, j)\}) > Y_j(g)$.
- **Lemma:** For any graph g, if a node i severs a link $e = (i, j) \in g$ with a node j such that $v(g \setminus \{(i, j)\}) \leq v(g)$, then it holds that $Y_i(g \setminus \{(i, j)\}) \leq Y_i(g)$.
- **Corollary:** For any graph g, under our model, if a node i severs a link $e = (i, j) \in g$ with a node j such that $v(g \setminus \{(i, j)\}) < v(g)$, then it must hold that $Y_i(g \setminus \{(i, j)\}) < Y_i(g)$.

Analysis of Pairwise Stable Networks (Cont.)

- **Lemma:** If $c < (b(1) - b(2))$, then the complete graph is the unique topology possible for a pairwise stable graph.

- **Regularity Condition (RC):** This involves a couple of conditions:
(a) If a pair of nodes i and j in a graph g are not neighbors and form a link (i,j) such that $v(g \cup \{(i,j)\}) \leq v(g)$, then it implies that either $Y_i(g \cup \{(i,j)\}) \leq Y_i(g)$ or $Y_j(g \cup \{(i,j)\}) \leq Y_j(g)$.
(b) $Y_i(g) \geq 0, \forall i \in N$.

- **Lemma:** If $c \in (b(1) - b(2), b(1)]$ and RC is satisfied, then any minimal edge graph with diameter 2 is pairwise stable.

Analysis of Pairwise Stable Networks (Cont.)

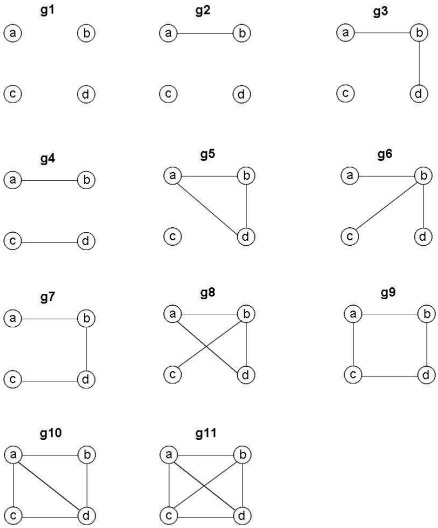
- **Corollary:** If $c \in (b(1) - b(2), b(1)]$ and RC is satisfied, then the star graph and the completely connected bi-partite graph are pairwise stable.
- **Lemma:** If $(b(1) - b(p)) < c < (b(1) - b(p+1))$ for any integer $p > 1$ and if g is a pairwise stable graph, then g is a graph with diameter p.
- **Lemma:** If $c > b(1) + b(2)$, then the empty graph is pairwise stable.

Efficiency versus Pairwise Stability

Theorem: Consider an anonymous and component additive value function v; and an anonymous, component balanced allocation rule $Y(.)$ satisfying weak link symmetry and improvement properties. Suppose g is an efficient graph relative to v. Then g is pairwise stable if and only if v, Y, and g satisfy the *regularity condition (RC)*.

Proof

Experiments: The Proposed Model



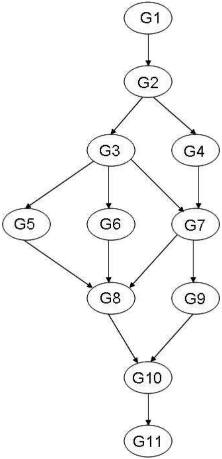
Experiments: The Proposed Model (Cont.)

Graph	Value of Graph
$g1$	0
$g2$	$b(1) - c$
$g3$	$2(b(1) - c) + b(2)$
$g4$	$2(b(1) - c)$
$g5$	$3(b(1) - c)$
$g6$	$3(b(1) - c) + 3b(2)$
$g7$	$3(b(1) - c) + 2b(2) + b(3)$
$g8$	$4(b(1) - c) + 2b(2)$
$g9$	$4(b(1) - c) + 2b(2)$
$g10$	$5(b(1) - c) + b(2)$
$g11$	$6(b(1) - c)$

Table: Value of each graph in the set of all possible graphs with 4 nodes

Experiments: The Proposed Model (Cont.)

Lattice of graphs with 4 nodes.



Experiments: The Proposed Model (Cont.)

Graph	Node a	Node b	Node c	Node d
$g1$	0	0	0	0
$g2$	$\frac{b(1)-c}{2}$	$\frac{b(1)-c}{2}$	0	0
$g3$	$\frac{b(1)-c}{2}+\frac{b(2)}{3}$	$(b(1)-c)+\frac{b(2)}{3}$	0	$\frac{b(1)-c}{2}+\frac{b(2)}{3}$
$g4$	$\frac{b(1)-c}{2}$	$\frac{b(1)-c}{2}$	$\frac{b(1)-c}{2}$	$\frac{b(1)-c}{2}$
$g5$	$b(1)-c$	$b(1)-c$	0	$b(1)-c$
$g6$	$\frac{b(1)-c}{2}+\frac{2b(2)}{3}$	$\frac{3(b(1)-c)}{2}+b(2)$	$\frac{b(1)-c}{2}+\frac{2b(2)}{3}$	$\frac{b(1)-c}{2}+\frac{2b(2)}{3}$
$g7$	$\frac{b(1)-c}{2}+\frac{b(2)}{3}+\frac{b(3)}{4}$	$(b(1)-c)+\frac{2b(2)}{3}+\frac{b(3)}{4}$	$\frac{b(1)-c}{2}+\frac{b(2)}{3}+\frac{b(3)}{4}$	$(b(1)-c)+\frac{2b(2)}{3}+\frac{b(3)}{4}$
$g8$	$(b(1)-c)+\frac{b(2)}{3}$	$\frac{3(b(1)-c)}{2}+\frac{2b(2)}{3}$	$\frac{b(1)-c}{2}+\frac{2b(2)}{3}$	$(b(1)-c)+\frac{b(2)}{3}$
$g9$	$(b(1)-c)+\frac{b(2)}{2}$	$(b(1)-c)+\frac{b(2)}{2}$	$(b(1)-c)+\frac{b(2)}{2}$	$(b(1)-c)+\frac{b(2)}{2}$
$g10$	$\frac{3(b(1)-c)}{2}+\frac{b(2)}{12}$	$(b(1)-c)+\frac{5b(2)}{12}$	$(b(1)-c)+\frac{5b(2)}{12}$	$\frac{3(b(1)-c)}{2}+\frac{b(2)}{12}$
$g11$	$\frac{3(b(1)-c)}{2}$	$\frac{3(b(1)-c)}{2}$	$\frac{3(b(1)-c)}{2}$	$\frac{3(b(1)-c)}{2}$

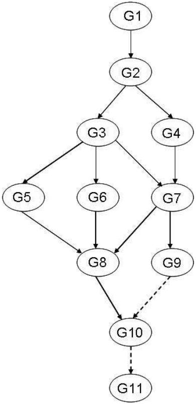
Table: Utilities of nodes for all graphs with 4 nodes, computed using the Myerson value

Experiments: The Proposed Model (Cont.)

From Graph	To Graph	Link added	Condition for First Node	Condition for Second Node
$g1$	$g2$	(a, b)	$c < b(1)$	$c < b(1)$
$g2$	$g3$	(b, d)	$c < b(1) + \frac{2b(2)}{3}$	$c < b(1) + \frac{2b(2)}{3}$
$g2$	$g4$	(c, d)	$c < b(1)$	$c < b(1)$
$g3$	$g5$	(a, d)	$c < b(1) - \frac{2b(2)}{3}$	$c < b(1) - \frac{2b(2)}{3}$
$g3$	$g6$	(b, c)	$c < b(1) + \frac{4b(2)}{3}$	$c < b(1) + \frac{4b(2)}{3}$
$g3$	$g7$	(c, d)	$c < b(1) + \frac{2b(2)}{3} + \frac{b(3)}{2}$	$c < b(1) + \frac{2b(2)}{3} + \frac{b(3)}{2}$
$g4$	$g7$	(b, d)	$c < b(1) + \frac{4b(2)}{3} + \frac{b(3)}{2}$	$c < b(1) + \frac{4b(2)}{3} + \frac{b(3)}{2}$
$g5$	$g8$	(b, c)	$c < b(1) + \frac{4b(2)}{3}$	$c < b(1) + \frac{4b(2)}{3}$
$g6$	$g8$	(a, d)	$c < b(1) - \frac{2b(2)}{3}$	$c < b(1) - \frac{2b(2)}{3}$
$g7$	$g8$	(b, d)	$c < b(1) - \frac{b(3)}{2}$	$c < b(1) - \frac{b(3)}{2}$
$g7$	$g9$	(a, c)	$c < b(1) + \frac{b(2)}{3} - \frac{b(3)}{2}$	$c < b(1) + \frac{b(2)}{3} - \frac{b(3)}{2}$
$g8$	$g10$	(c, d)	$c < b(1) - \frac{b(2)}{2}$	$c < b(1) - \frac{b(2)}{2}$
$g9$	$g10$	(a, d)	$c < b(1) - \frac{5b(2)}{6}$	$c < b(1) - \frac{5b(2)}{6}$
$g10$	$g11$	(b, c)	$c < b(1) - \frac{5b(2)}{6}$	$c < b(1) - \frac{5b(2)}{6}$

Experiments: The Proposed Model (Cont.)

Transitions for $(b(1) - \frac{5b(2)}{6}) < c < (b(1) - \frac{2b(2)}{3})$ in lattice of graphs with 4 vertices. Solid lines indicate possible transitions and Dotted lines indicate transitions that are not possible between graphs.



Experiments: The Proposed Model (Cont.)

Comparison of experimental results with that of formal analysis of the model.

Condition	Topology of Graph	Prediction from Analysis
$c < b(1) - b(2)$	$g11$	$g11$
$(b(1) - b(2)) < c < (b(1) - \frac{5b(2)}{6})$	$g11$	–
$(b(1) - \frac{5b(2)}{6}) < c < (b(1) - \frac{2b(2)}{3})$	$g9,\ g10$	–
$(b(1) - \frac{2b(2)}{3}) < c < (b(1) - \frac{b(2)}{2})$	$g6,\ g10$	$g6$
$(b(1) - \frac{b(2)}{2}) < c < (b(1) - b(2))$	$g6,\ g8,\ g9$	$g6,\ g9$
$(b(1) - b(2)) < c < (b(1) - \frac{b(3)}{2})$	$g6,\ g8,\ g9$	$g6,\ g9$
$(b(1) - \frac{b(3)}{2}) < c < (b(1) - b(3))$	$g6,\ g9$	–
$(b(1) - b(3)) < c < b(1),\ b(1) < \frac{2}{3}$	$g6,\ g9$	–
$(b(1) - b(3)) < c < b(1),\ b(1) > \frac{2}{3}$	$g6,\ g7$	–
$c > b(1)$	$g1$	$g1$

Directions to Future Work

- More often, in real life social networks, links get form or severe based on local information. Thus network formation games with local information are much interesting to explore
- The ideal goal for game theoretic models for network formation is to come up stable and efficient network topologies that exhibit heavy tail like distribution for degree distributions and small-world phenomenon

Next Part of the Presentation

Important Research Directions

- **Altruistic Game Theoretic Models:**
  - Often individuals in social networks are not only rational and intelligent but also altruistic
  - Social network formation and Bargaining on social networks
- **Probabilistic Graphs:**
  - Complex networks often entail uncertainty and thus can be modeled as probabilistic graphs
  - M. Potamias, F. Bonchi, A. Gionis, and G. Kollios. *k*-nearest neighbors in uncertain graphs In VLDB Endowment, Vol. 3, No. 1, 2010

Important Research Directions (Cont.)

- Exploit games with special structure such as convex games, potential games, matrix games, etc. to problems in SNA
- Designing scalable approximation algorithms with worst case guarantees
- Problems such as incentive compatible learning and social network monetization are at the cutting edge
- Explore numerous solution concepts available in the ocean of game theory literature

Important Text Books

- D. Easley and J. Kleinberg. Networks, Crowds, and Markets. Cambridge University Press, 2010.
- 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.

Important References

- Ramasuri Narayanam and Y. Narahari. A Shapley Value based Approach to Discover Influential Nodes in Social Networks. In IEEE Transactions on Automation Science and Engineering (IEEE TASE), 2011.
- 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.
- Ramasuri Narayanam and Y. Narahari. Determining Top-k Nodes in Social Networks using the Shapley Value. In AAMAS, pages 1509-1512, Portugal, 2008.
- Ramasuri Narayanam and Y. Narahari. Nash Stable Partitioning of Graphs with Application to Community Detection in Social Networks. Under Review, 2010.
- D. Dikshit and Y. Narahari. Truthful and Quality Conscious Query Incentive Networks. In Workshop on Internet and Network Economics (WINE), 2009.
- Mayur Mohite and Y. Narahari. Incentive Compatible Influence Maximization in Social Networks with Application to Viral Marketing. AAMAS 2011.

Leading Researchers

- Jon M. Kleinberg
- Christos Faloutsos
- Matthew O. Jackson
- Sanjeev Goyal
- Eva Tardos
- Jure Leskovec
- Nicole Immorlica
- David Kempe
- Krishna P. Gummadi
- Tanya Berger-Wolf
- ...

Network Data Sets

- 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/
- ...

Softwares

- 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)

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
- ...

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
- ...

Some More Useful Resources

- Y. Narahari, Dinesh Garg, Ramasuri Narayanam, Hastagiri Prakash. Game Theoretic Problems in Network Economics and Mechanism Design Solutions. In Series: Advance Information & Knowledge Processing (AIKP), Springer Verlag, London, 2009.
- Blog on Social Networks: http://cs2socialnetworks.wordpress.com/

Thank You

Proof:

- Given that $(b(1) - b(2)) < c < (b(1) - b(3))$.
- Consider any network g and assume that there exists a pair of nodes x and y such that $d_g(x, y) > 2$.
- Recall that the communication between nodes x and y in g leads to a benefit of $b(d_g(x, y))$.
- Assume that x and y form a link and call the link $e = (x, y)$ and also call the new graph $g' = g \cup \{e\}$.

Go Back

- Link (x, y) leads to a net benefit of $(b(1) - c)$.
- Note that the length of a shortest path between any pair of nodes in g' either remains same or decreases when compared to that in g.
- From the above observations, we get that $v(g') - v(g) \geq [(b(1) - c) - b(d_g(x, y))] > 0$, since $d_g(x, y) > 2$ and $(b(1) - c) > b(3) \geq b(d_g(x, y))$.
- That is, $v(g') > v(g)$.

Go Back

Proof:

- Consider that g is an efficient graph.
- Due to previous lemma, the shortest distance between any pair of nodes is at most 2 in g.
- That is, g is a graph with diameter 2.
- Suppose that g is not a minimal edge graph with diameter 2.

Go Back

- Then g contains a link (x, y) such that severing the link (x, y) does not lead the diameter to exceed 2.
- Thus, if we remove the link (x, y), only the shortest distance between the nodes x and y increases to 2.
- Since $(b(1) - b(2)) < c$, the value of g strictly increases if the link (x, y) is severed.
- Contradiction to the fact that g is efficient.

Go Back

Proof Part 1:

- Given that g is efficient. Assume that g is pairwise stable.
- *Claim:* Regularity condition (RC) is holds.
- Let i and j be a pair of non-neighbor nodes in g and form a link, (i, j). Call the new graph $g' = g \cup \{(i, j)\}$.
- Since g is efficient and v is component additive, we get that $v(g') \leq v(g)$.
- If $Y_i(g') > Y_i(g)$ (or $Y_j(g') > Y_j(g)$), then due to weak link symmetry, we get that $Y_j(g') > Y_j(g)$ (or $Y_i(g') > Y_i(g)$).
- Contradicts the fact that g is pairwise stable.
- Also, since g is pairwise stable, we have that $Y_i(g) \geq 0$, $\forall i \in N$.

Go Back

Proof Part 2:

- Given that g is efficient. Assume that the regularity condition is satisfied.
- *Claim:* g is pairwise stable for Y relative to v.
- *Severing a Link:* Suppose a node x_1 severs a link (x_1, y_1) with node y_1 in g. Call the new graph $g_1 = g \setminus \{(x_1, y_1)\}$.
- Since g is efficient and v is component additive, we get that $v(g_1) \leq v(g)$.
- From previous results, it is clear that node x_1 is not strictly better off by severing the link.

Go Back

- *Adding a Link:* Suppose two non-neighbor nodes i and j form a link (i,j) in g and call the new graph $g' = g \cup \{(i,j)\}$.
- Since g is efficient and v is component additive, it holds that $v(g') \leq v(g)$.
- If $Y_i(g') > Y_i(g)$ (or $Y_j(g') > Y_j(g)$), then due to weak link symmetry, we get that $Y_j(g') > Y_j(g)$ (or $Y_i(g') > Y_i(g)$). Contradiction to *RC*!
- Implies that neither i nor j is strictly better off. Again since *RC* is satisfied, we get that $Y_i(g) \geq 0$, $\forall i \in N$.
- Hence, g is pairwise stable.

Go Back