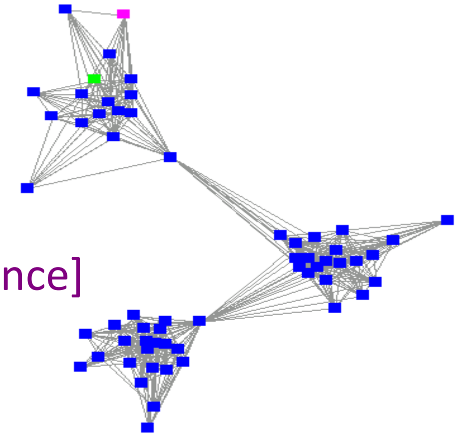
SNAzzy

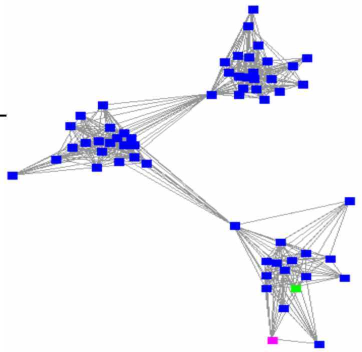


[A Social Network Analysis Suite for Business Intelligence]

Sougata Mukherjea & Amit A. Nanavati

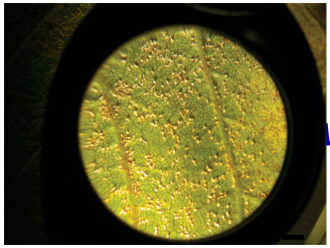
Social Network Analysis

Data Mining vs. SNA



- So far,
  - Individual calling pattern mining
    - when ? how often ? whom ? from where ?
    - Vertical ("tunnel") mining

Microscopic view

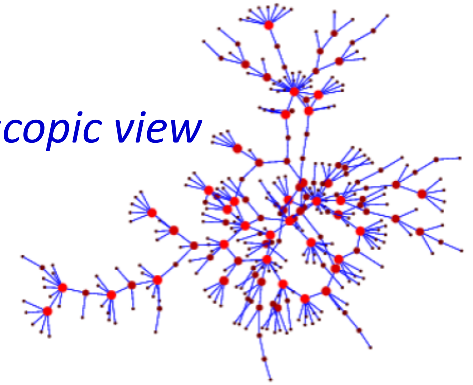


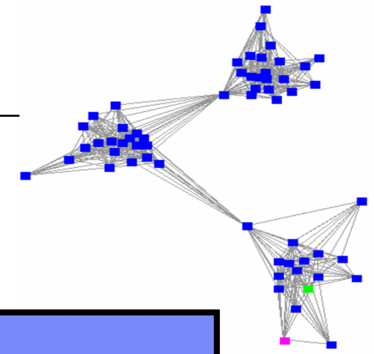
- But, we have a lot of behavioural data that is currently unused
  - who-calls-whom ?
  - how is everyone connected with each other ?
  - how is a new service usage spreading ?

- What is it that I do not know of my customers today ?

- Social Network Analysis can do this for you
  - The social network perspective:
    - *what also matters is the individual's location and interactions in the network*

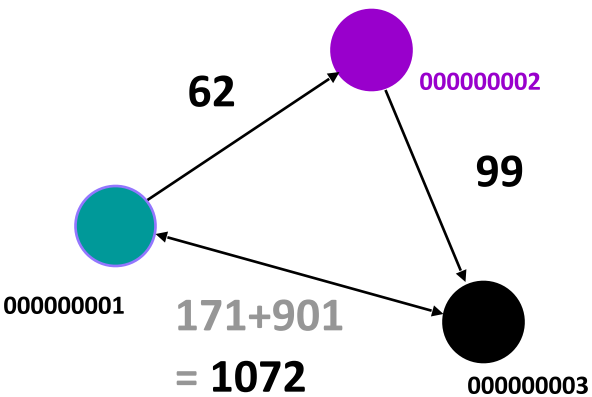
Telescopic view

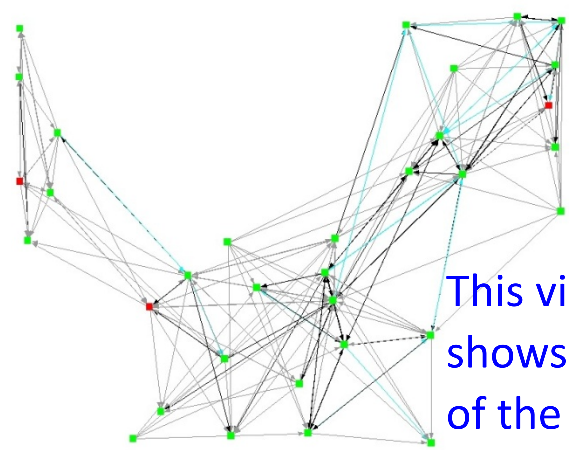




Construction of the Social Network from CDR data

Caller ID	Callee ID	Duration (seconds)
000000001	000000002	62
000000001	000000003	901
000000002	000000003	99
000000003	000000001	171

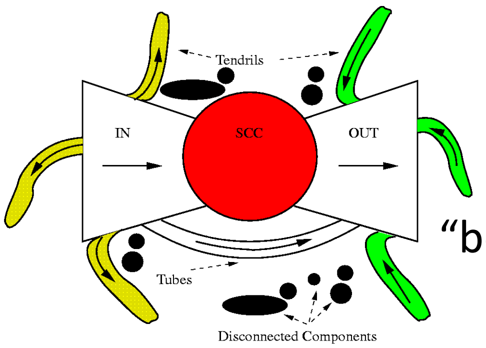




This visualization shows only part of the social network to be constructed.

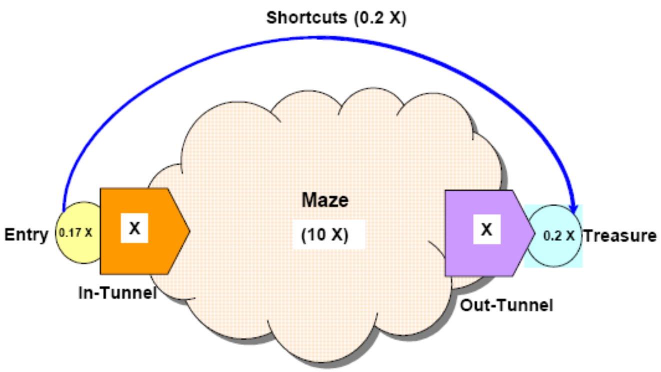
Related Work

- World-wide web graph model
  - Power laws in indegree / outdegree
  - Communities as *bipartite cores*
  - Hub, authority, pagerank
  - “Bow-tie” Model
    - IN, OUT, SCC (strongly connected component)



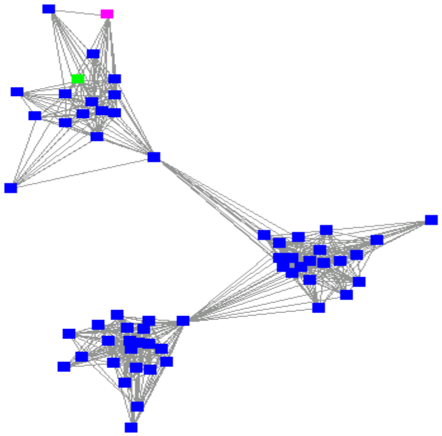
“bow-tie”

- Telecom network (AT&T Bell Labs)
  - Landline phones
  - Calls made in 1 day
  - 53 million vertices, 170 million edges
  - A giant component with 80% of the nodes found
  - 3.7 million separate components, most of them pairs

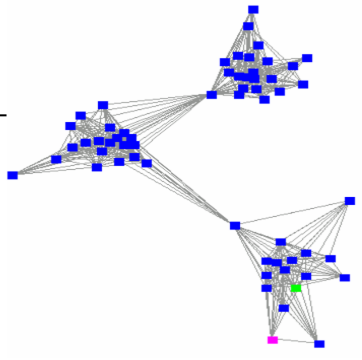


“treasure-hunt”

- Mobile call-graph (IBM Research)
  - Local mobile calls made in a month/week (4 regions)
  - 1.25 million vertices, 4.5 million edges
  - Edge-based model



Telecom Business Usecases



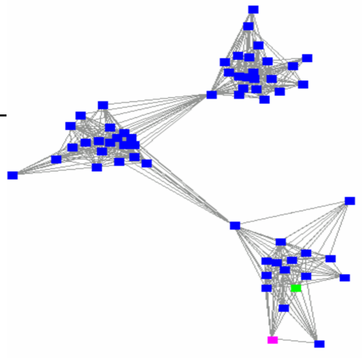
SNAzzy Capabilities

- Community Finding
  - Closed user groups, Stars, Dense groups
- Potential Acquisition Target Identification
  - Whom should we target from the competition?
- Rotational Churn Identification
  - How do we differentiate SIM changers from real churners?
- Customer Value & Influence Analysis
  - What is the value of this customer? What sort of social influence does he exert?
- VAS Usage Analysis
  - How is VAS adoption and usage? Is any of the adoption social? Who are the key customers?
- Churn Prediction and Analysis
  - Social churn: Does Raj churn because his friends do?

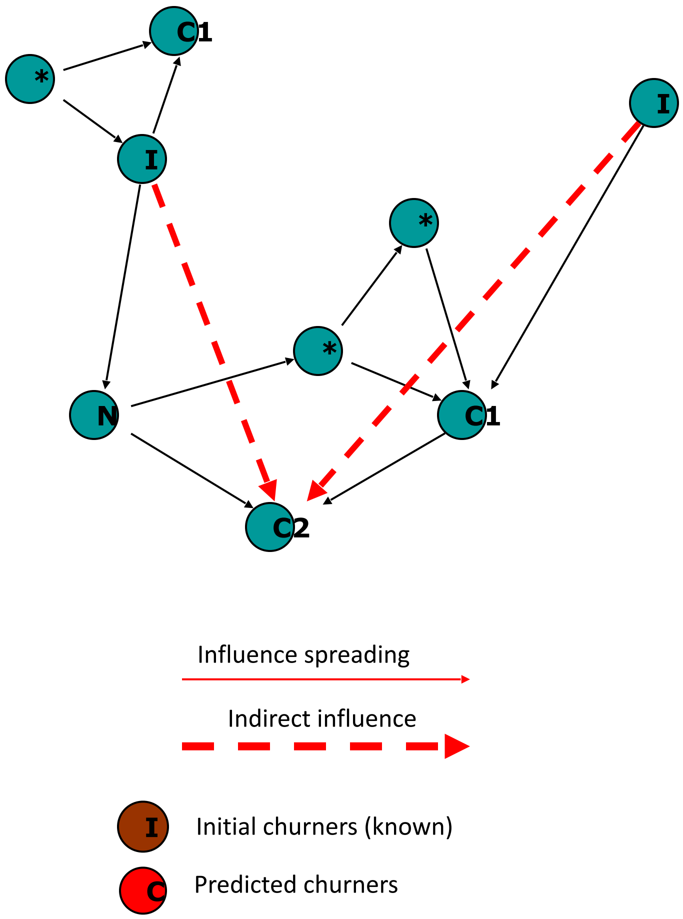
Telco Churn

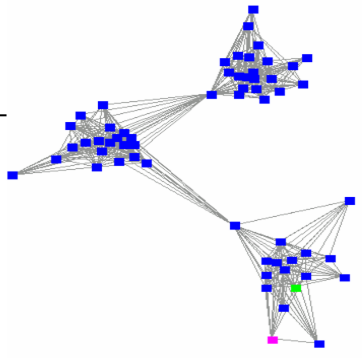
- In Mobile Telecom markets
  - Competition is ever-increasing
    - New players
    - Lower Average Revenue Per User (ARPU)
  - Customer "churn"
    - Low barriers to switching providers
      - Especially in Pre-paid segment
  - Customer acquisition → customer retention
- Telecoms must rely on business intelligence
  - Design the right incentives
  - Adopt right marketing strategies

Churn Prediction & Analysis



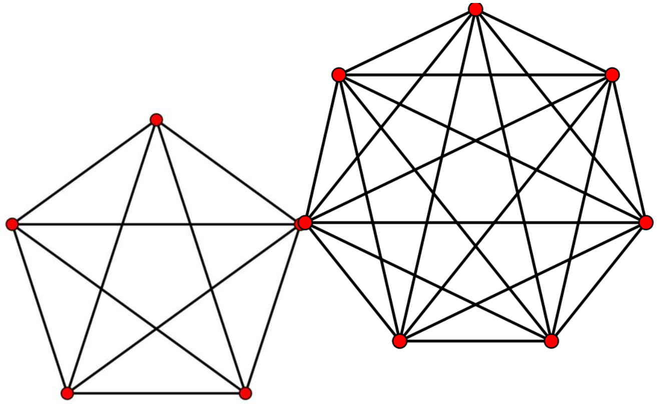
- SNAzzy constructs an influence propagation model:
  - Given persons who churned (initial churners)
  - Diffuse their influence into their social environment
  - Thus, their friends are at a larger churn risk..
    (The two C1s churn; N does not, since not enough influence)
  - And this propagates to some of their friends' friends as well.
    (C2 affected due to indirect, cumulative influence)
- Output
  - List of predicted churners
- Business Value
  - Unique model
    - Captures higher order social effects
    - Capture the effect of multiple churners on a subscriber
  - Does not require profile information.
  - Can be applied in post-paid and pre-paid markets as well.
  - Once the model is created, it can be run quickly and often.
  - Complements traditional churn models.



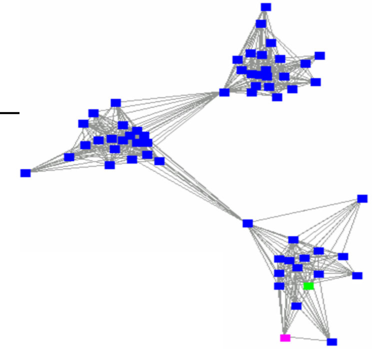


Community Identification | Cliques

- Closed user groups
  - Everyone knows everyone else
  - People could be members of multiple cliques
    - High school, college, company cliques

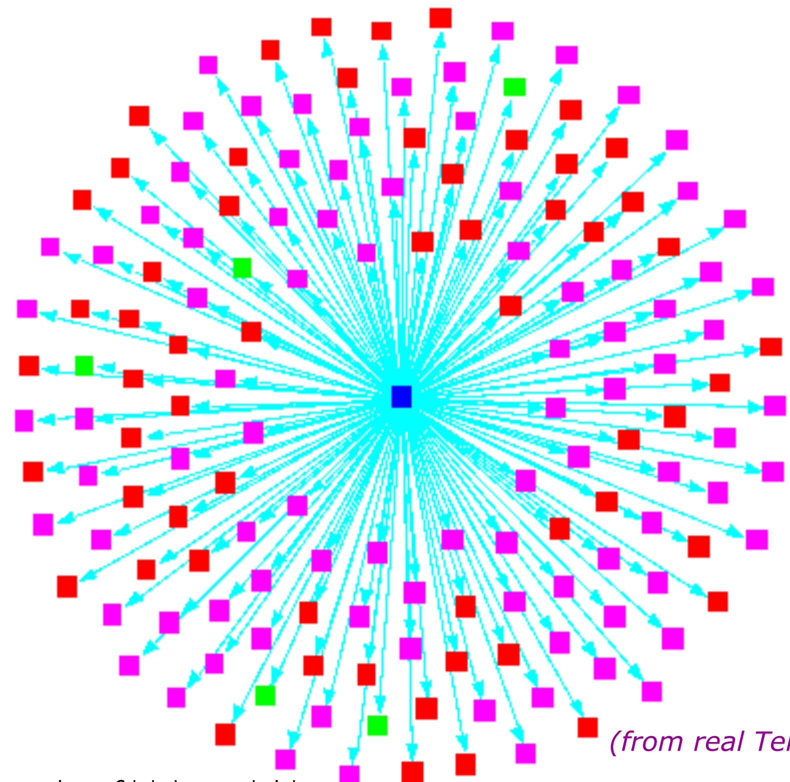


- Output
  - List of cliques with their members identified
  - Clique connectors identified
- Business Value
  - Telcos can offer automatic discounts for calls within a clique
  - To increase stickiness and loyalty
  - Can be leveraged by SNAzzy churn prediction



Community Identification | Stars

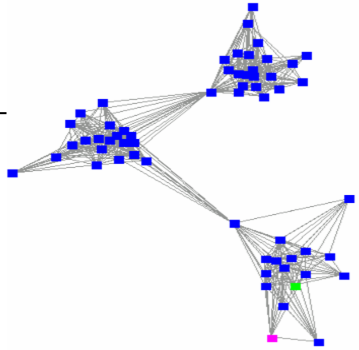
- "Hub" knows the "spokes"
  - Spokes may not know each other
  - Scenarios
    - Hub = buyer ; Spoke = seller (or vice-versa)
    - Hub is someone influential (a minister, celebrity) whom disconnected people are calling
    - Hub is a telemarketeer
    - Hub is a call-centre



(from real Telecom data)

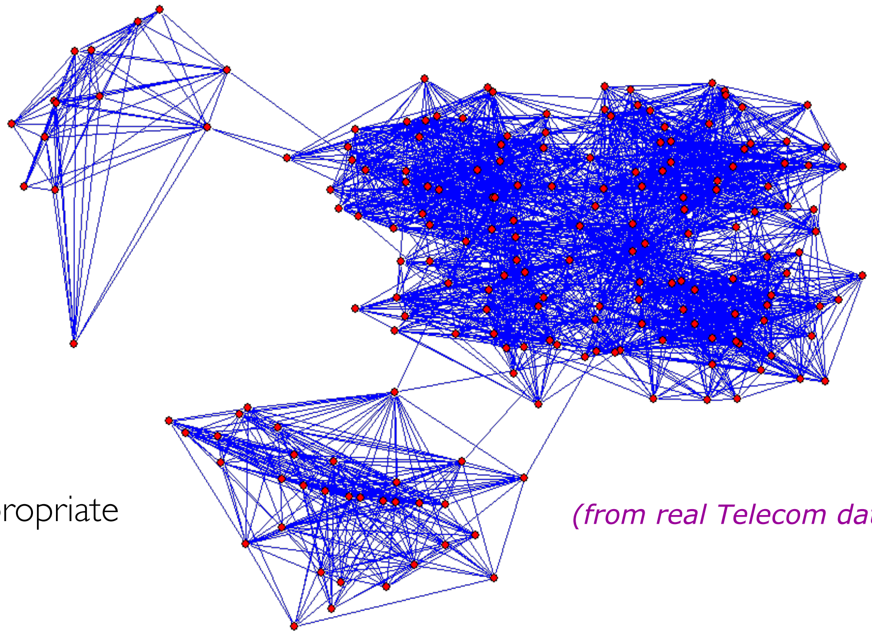
- Output
  - List of stars with the hubs and spokes identified

- Business Value
  - Telcos can identify call-centres, telemarketeers, and people of high social importance.
  - This is especially useful in prepaid where demographics/profile is not available.
  - Introduces social KPIs to measure the social value of a customer.
  - Can be leveraged by SNAzzy churn prediction.



Community Identification | Dense Communities

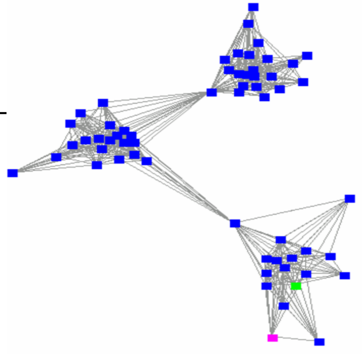
- Dense Communities are hotbeds of social activity
  - A lot of people are talking to a lot of people
  - Figure shows a group of 129 people where everyone knows and is calling at least 18 others!



- Output
  - Dense communities and members identified

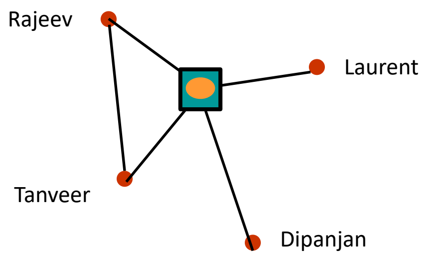
- Business Value
  - These communities can be targeted with appropriate calling plans to motivate more calling.

(from real Telecom data)



Detecting Internal/Rotational Churn

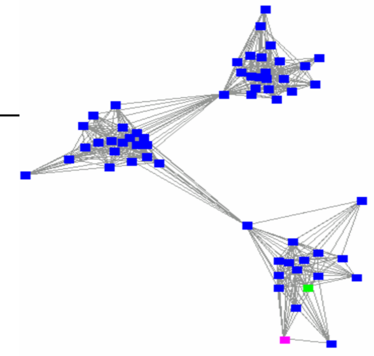
- Is my churn real?
  - Are people simply discarding old SIMS and replacing them with new ones?
  - Are any new plans cannibalising the existing ones?
  - If it is the same person, then it is likely that he is calling the same set of people with the same set of frequencies.



Same person ?

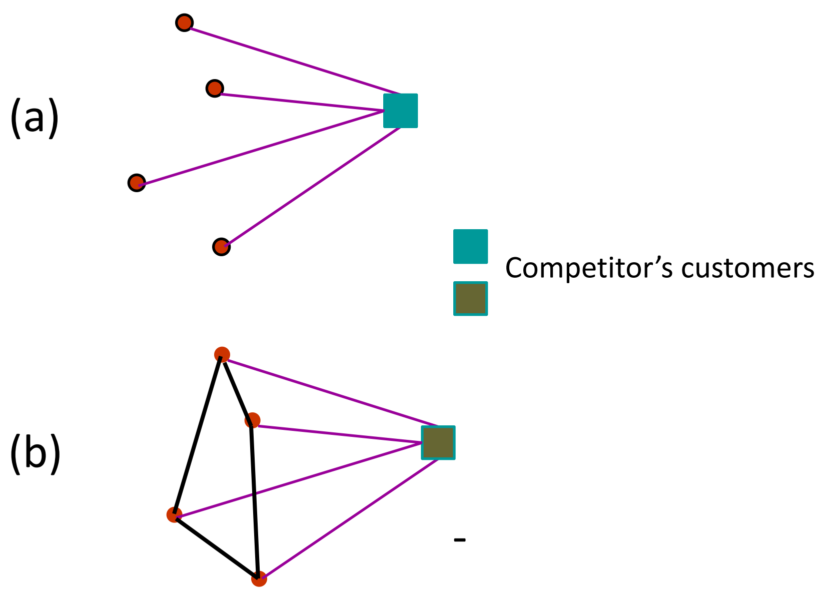
- Output
  - List of internal churners (old number, new number) identified.

- Business Value
  - Helps distinguish between real and internal churn

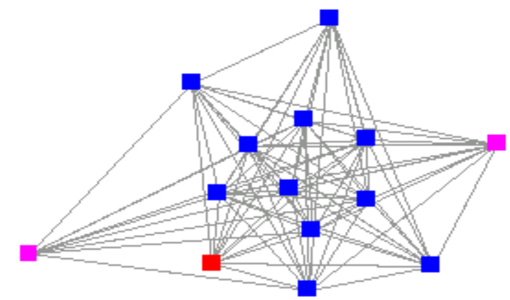


Potential Acquisition Target Identification

- Do I know high potential conversion targets from my competition?
  - (a) Competitor subscriber talking to unrelated people in my network.
  - (b) Competitor subscriber talking to friends in my network.
  - (b) is a better target than (a) due to its social connections.

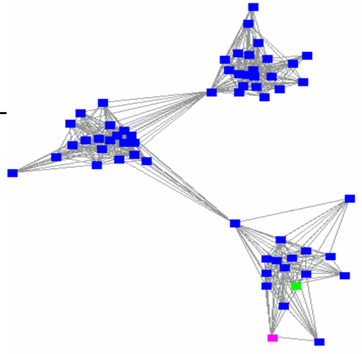


- Output
  - Ranked list of acquisition targets
- Business Value
  - Improves hit ratio for acquisition



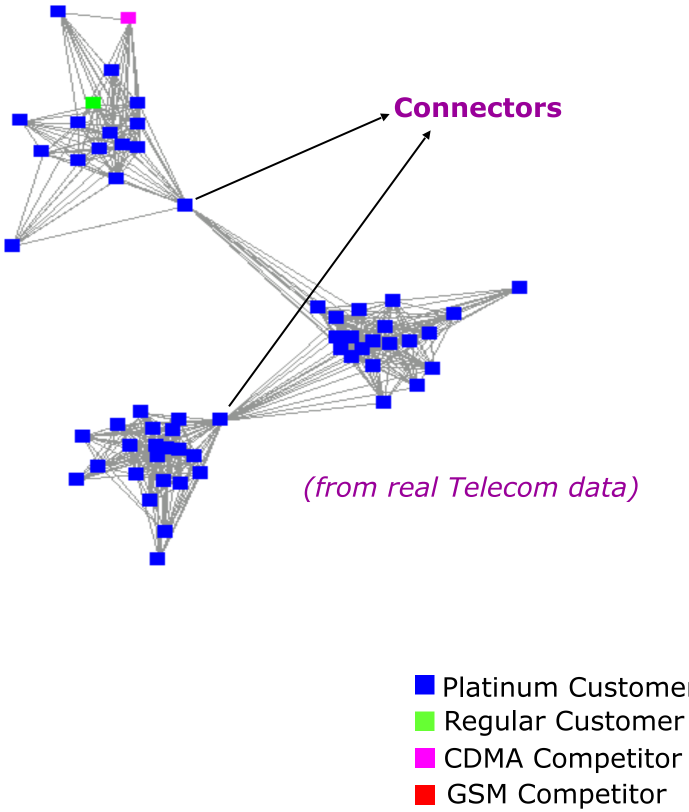
Potential acquisition targets from competitors
(from real Telecom data)

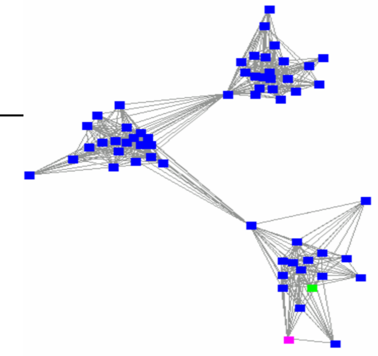
- Platinum Customer
- Regular Customer
- CDMA Competitor
- GSM Competitor



Customer Value & Influence Analysis

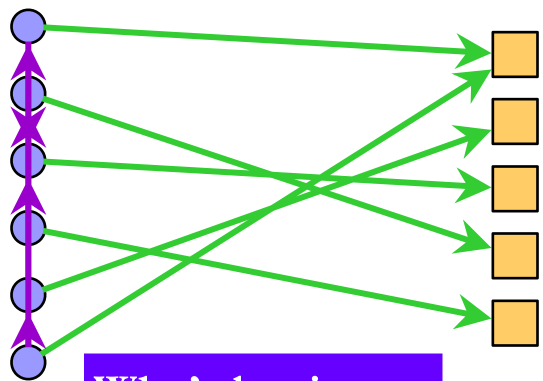
- Identify people of high social influence
  - Not just large bill payers
  - Includes people who are called by a lot of people
  - People who are hubs of 'stars'
  - Members of many communities
  - "Clique connectors"
- Output
  - A social importance value for every customer
  - List of viral marketing "seeds"
- Business Value
  - Potential targets for viral marketing
  - Provides social KPIs to calculate the value of a customer





VAS Usage Analysis

People
Social Network



Things (products, services,...)

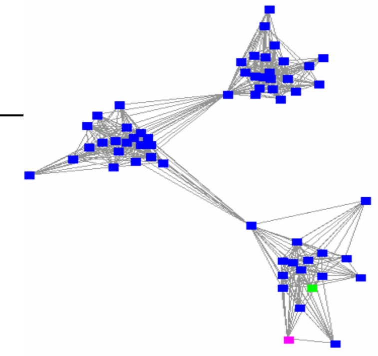
Who's buying what ?

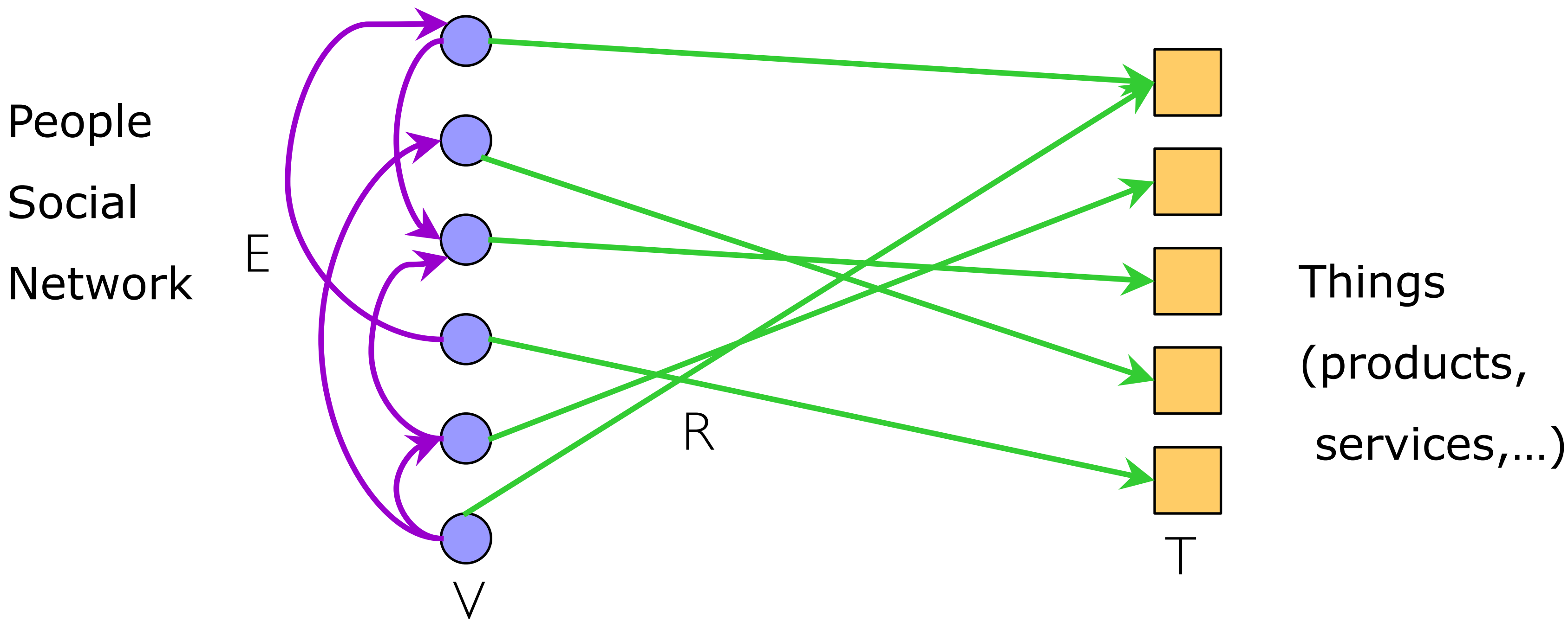
- What are the communities that are interested in a particular service?
- Which products are (not) being bought by communities?

Like-minded Communities

Bringing the Familiarity and Similarity Together

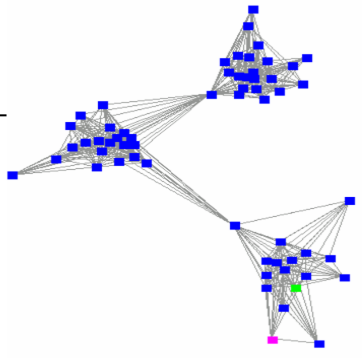
Setting





Who's buying (or interested in) what ?

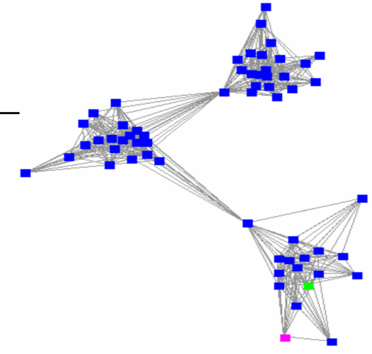
Like-minded Communities



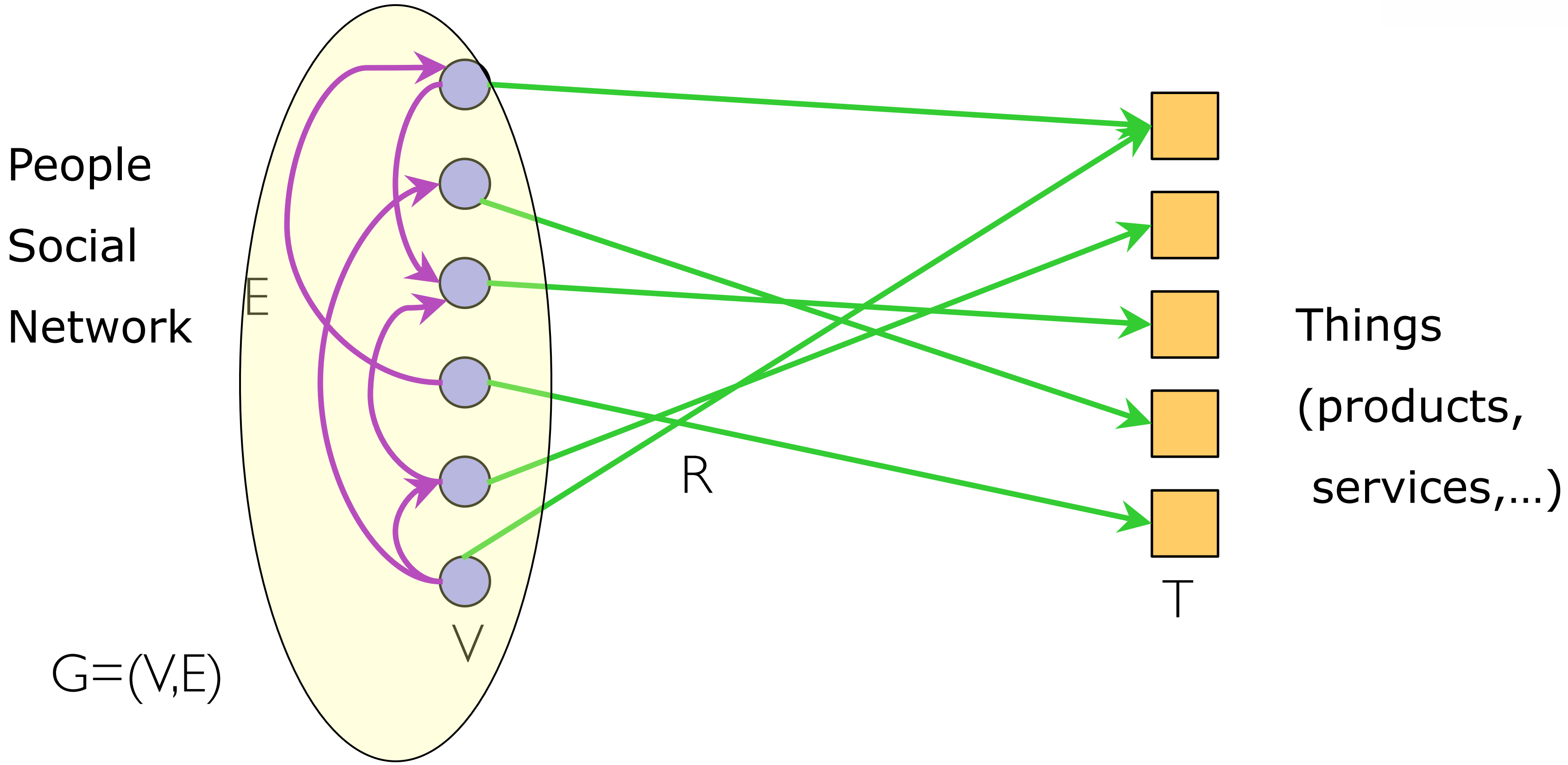
- Besides social interaction (familiarity), it is also desirable to factor in shared characteristics or interests (similarity) while finding communities
  - Why? Social (and viral) campaigns are more likely to succeed on communities with shared interests

- We provide a method to find communities where like-mindedness is an explicit objective
  - To formalize this notion, we define like-mindedness

- None of the approaches in the literature take like-mindedness (or any similar function) as an objective in the community finding process

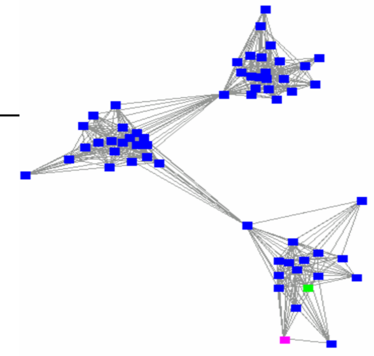


Setup and Notation



Who’s buying (or interested in) what ?

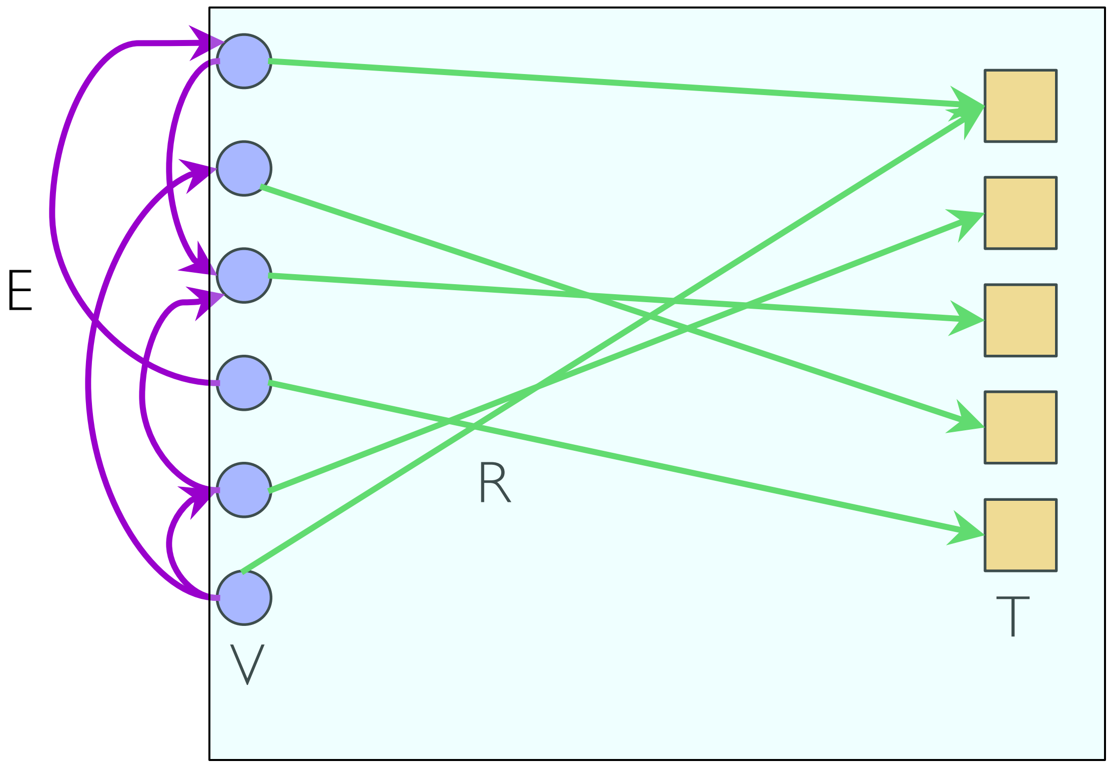
Setup and Notation



People

Social

Network



Things

(products,

services,…)

B=(V,T,R)

Who's buying (or interested in) what ?

A visit to the supermarket…

- Supermarket has a set of items
  - I = {**wheat, rice, salt, sugar,…, bread, butter**}

- You buy a subset of items
  - Your transaction **t** = {**milk, bread, cereal, jam**}
  - **t** $\subseteq$ **I**

- Instead of just looking at aggregate sales,
  - 1000 litres of milk, 523 breads…

- …let's analyse transactions.

A visit to the supermarket…

- Many people visited the market:
  - You t_1 = {milk, bread, cereal, jam}
  - Me t_2= {apple, cucumber, tomato, milk}
  - AB t_3= {oranges, cucumber, tomato, jam}
  - SRK t_4 = {cereal, bread, butter, jam}
  - …

- The total set of items is **I**
- The total set of transactions is D = $\{t_{1,\ldots,}t_m\}$
- We want to know which items are being bought together?

A visit to the supermarket…

- Many people visited the market:
  - You t_1 = {milk, bread, cereal, jam}
  - Me t_2= {apple, cucumber, tomato, milk}
  - AB t_3= {oranges, cucumber, tomato, jam}
  - SRK t_4 = {cereal, bread, butter, jam}
- Does bread $\Longrightarrow$ butter?
  - "bread" occurs in 2 transactions, t_1 and t_4
  - "butter" occurs in 1 transaction, t_4
  - "bread" and "butter" occur in 1 transaction, t_4

Association Rules

- Many people visited the market:
  - You t_1= {milk, bread, cereal, jam}
  - Me t_2= {apple, cucumber, tomato, milk}
  - AB t_3= {oranges, cucumber, tomato, jam}
  - AK t_4 = {apple, bread, tomato}
  - SRK t_5 = {butter, $\underbrace{\text{jam, bread}}$}

frequent itemset

- bread $\Rightarrow$ jam
  - support = $\dfrac{\text{number of transactions in which both bread and jam occur}}{\text{total number of transactions}}$

    = 2/5 = 0.4 = 40%
  - confidence = $\dfrac{\text{number of transactions in which both bread and jam occur}}{\text{number of transactions in which bread occurs}}$

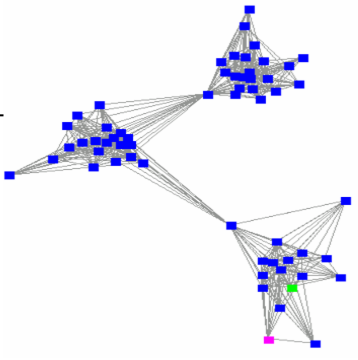
    = 2/3 = 0.6 = 60%

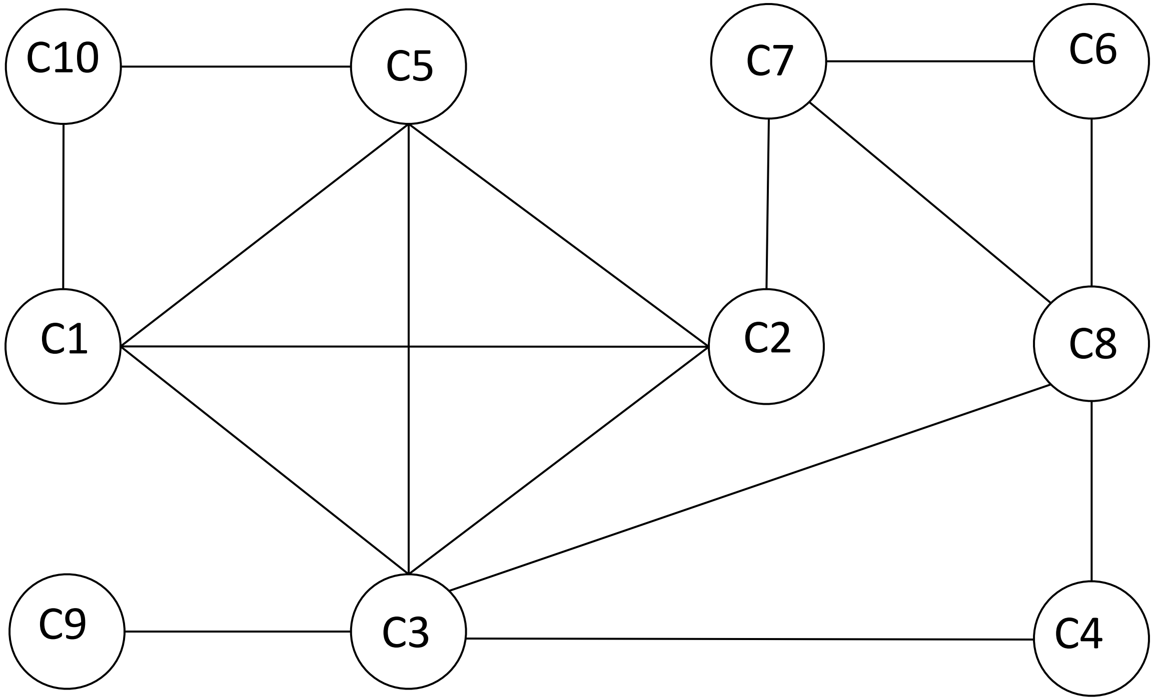
Association Rules

- A set of items $\mathbf{I} = \{ i_1, i_2, \ldots, i_N \}$
- A transaction t is a subset of $\mathbf{I}$
- The complete set of transactions D

- We say A $\Rightarrow$ B,
  - $$\frac{\text{Number of transactions including items in A and B}}{\text{Total number of transactions}} \geq \textit{minsup}$$
  - $$\frac{\text{Number of transactions including items in A and B}}{\text{Number of transactions including A}} \geq \textit{minconf}$$

An Example

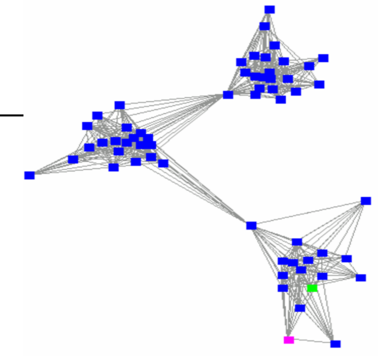




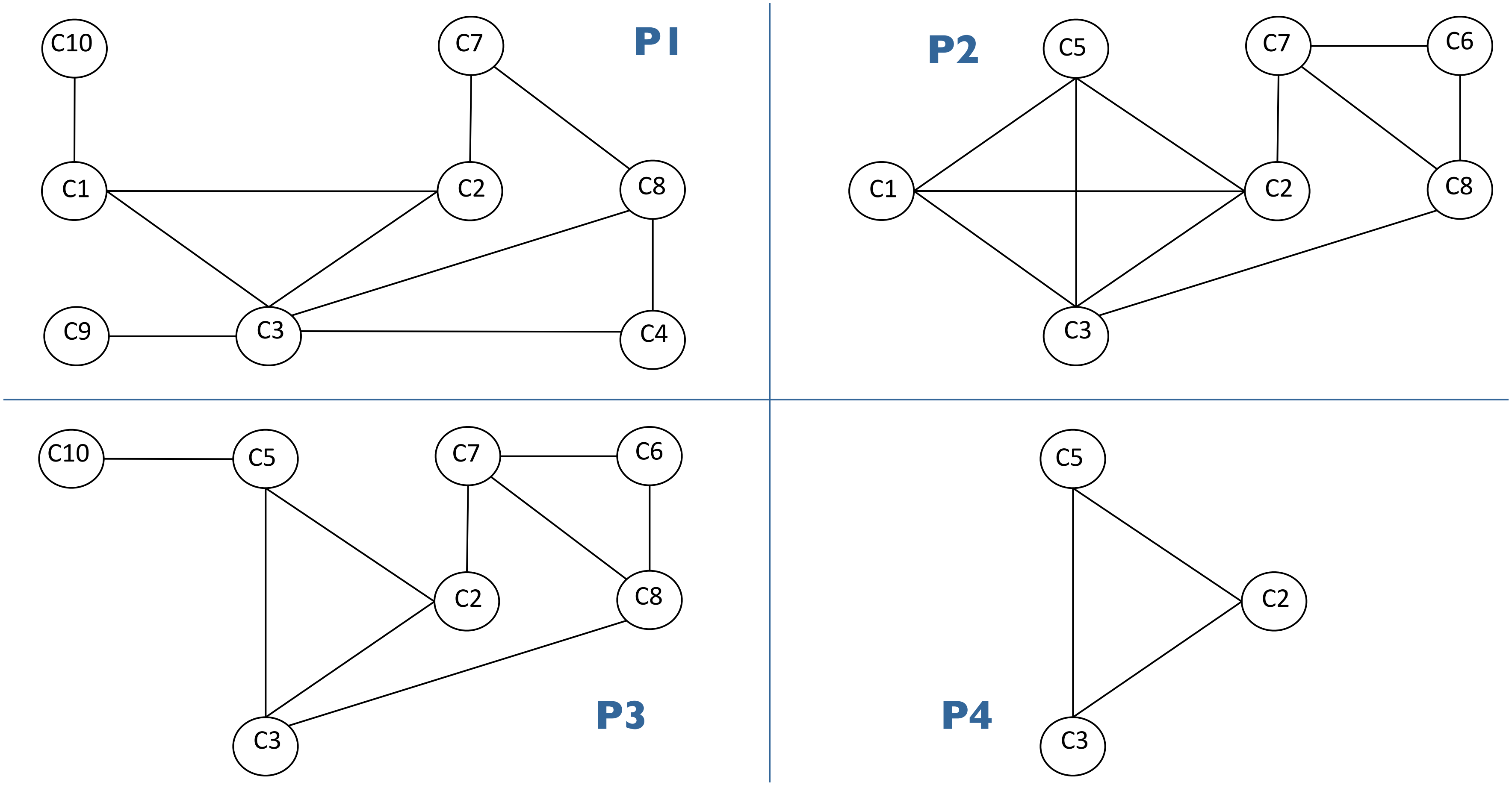
Example social network

	P1	P2	P3	P4
C1	Y	Y	N	N
C2	Y	Y	Y	Y
C3	Y	Y	Y	Y
C4	Y	N	N	N
C5	N	Y	Y	Y
C6	N	Y	Y	N
C7	Y	Y	Y	N
C8	Y	Y	Y	N
C9	Y	N	N	N
C10	Y	N	Y	N

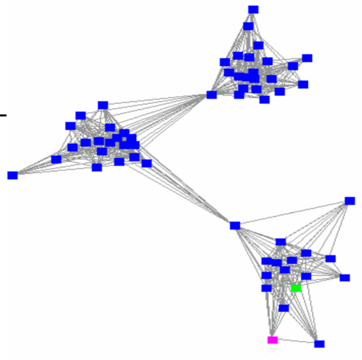
Example purchase history



Example continued
Induced subgraphs for products

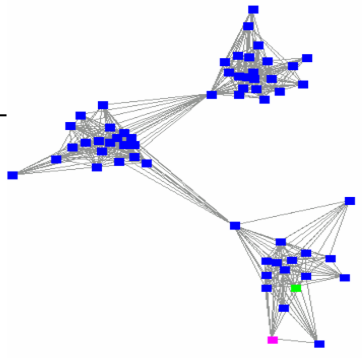


Example continued



- Cliques for product (minimum size 3)
  - P1: {c1,c2,c3}, {c3,c4,c8}
  - P2: {c1,c2,c3,c5}, {c6,c7,c8}
  - P3: {c2,c3,c5}, {c6,c7,c8}
  - P4: {c2,c3,c5}
- Treating each clique as a transaction and perform frequent itemset mining with minimum support as 2
- The frequent itemsets found are {c1,c2,c3} (support 2), {c2,c3,c5} (support 3) and {c6,c7,c8} (support 2)
  - c4, c9 and c10 are not part of any frequent itemsets

Example continued



Like-minded

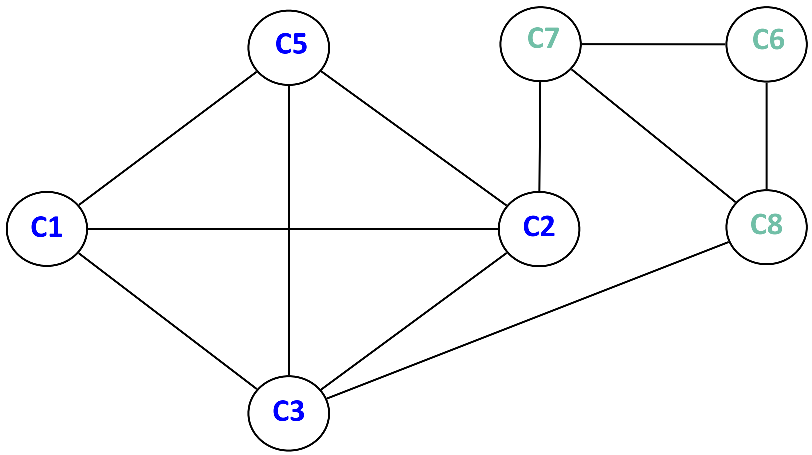
Communities found

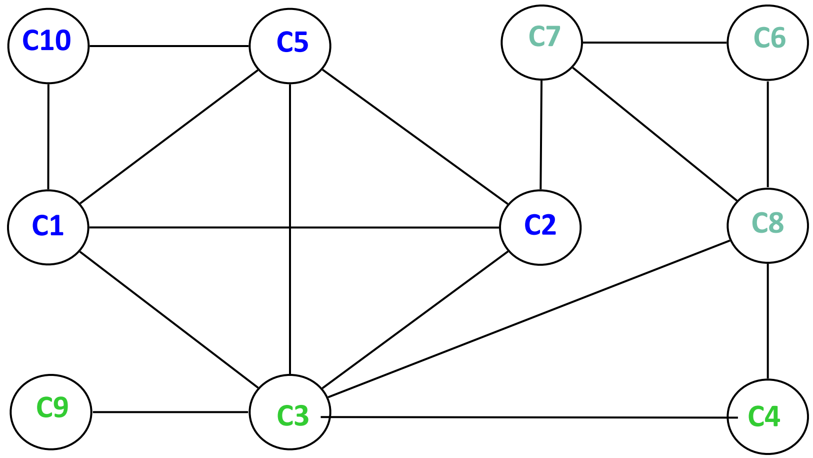
{c1,c2,c3,c5}

{c6,c7,c8}

Like-mindedness: 0.77157

Modularity: 0.28099





Communities found

Using CNM method

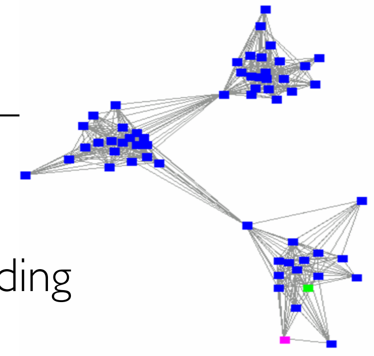
{c1,c2,c5,c10}

{c3,c4,c9}

{c6,c7,c8}

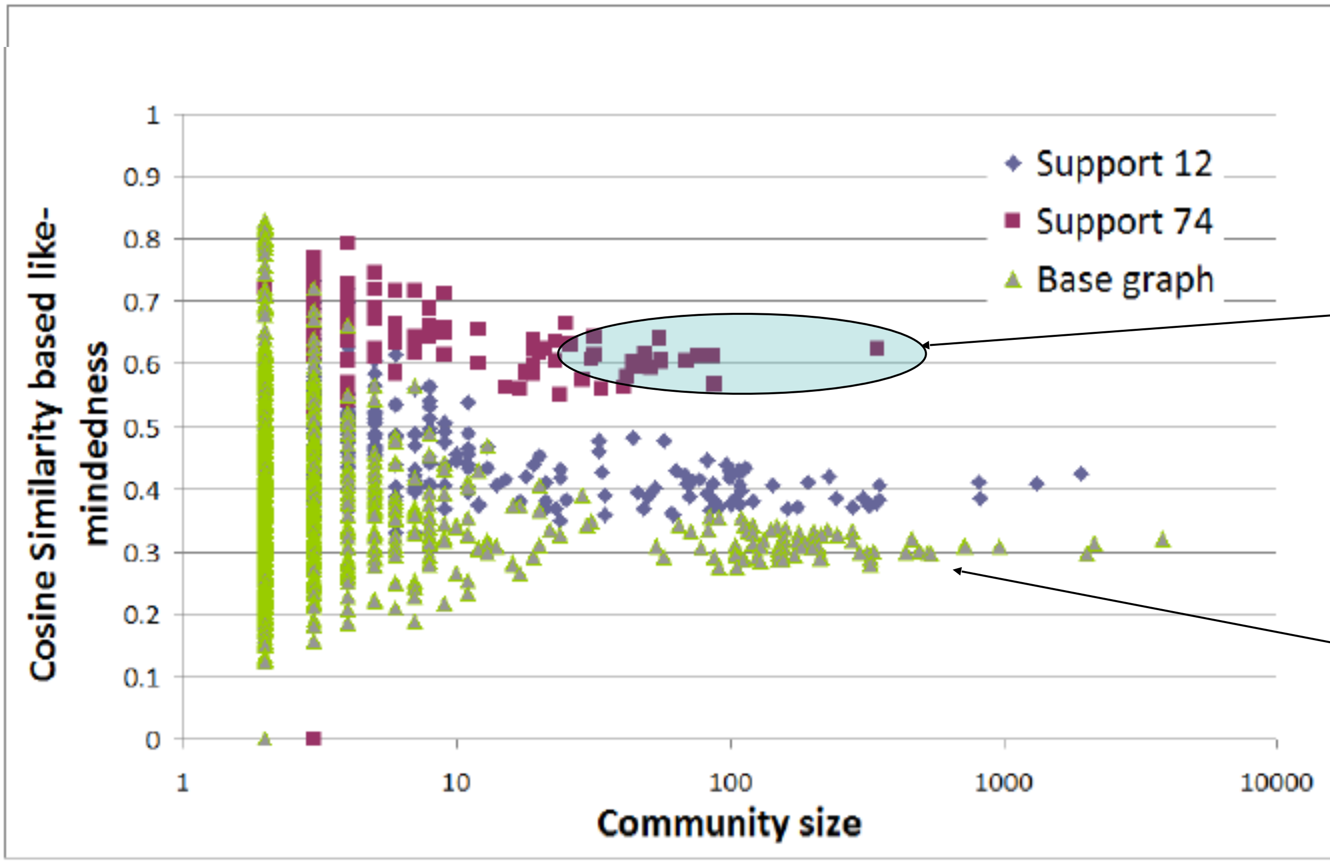
Like-mindedness: 0.66553

Modularity: 0.27539



The Like-minded Communities: Benefit

We now compare the performance of the usual SNA community finding algorithm with our algorithm at the individual community level.

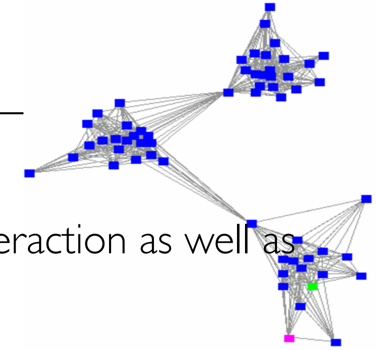


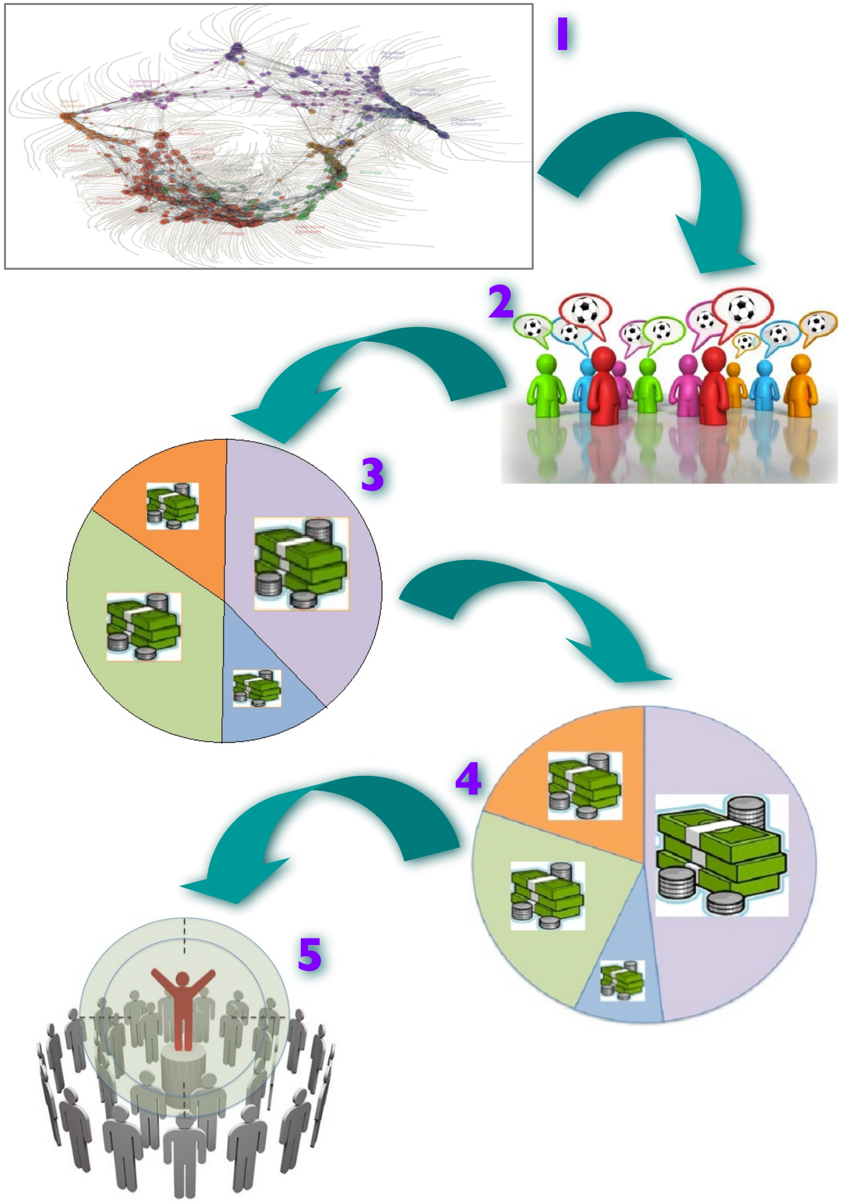
Goldmine communiti
Found using Vibes

High like-mindedness
And (relatively) large size

No such large and high
Like-mindedness
Communities using usual
SNA algorithms

Viral Marketing Campaign Target Selection





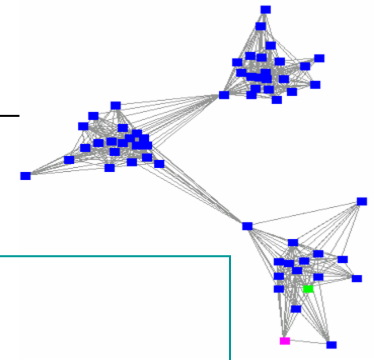
1. Given: A population, and its social interaction as well as purchase/interest information

2. To identify the campaign targets, first we divide the population into 'Like-Minded Communities'
   - Such communities are more likely to adopt products virally

3. Given the overall budget, we decide how many people to target from each like-minded community
   - Based on the size, like-mindedness and other parameters of the communities

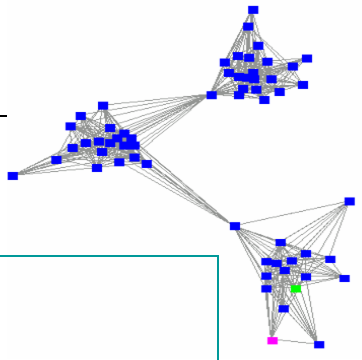
4. The Marketer can choose to increase or decrease the system suggested budget allocation

5. The final step is to determines the specific targets within the groups with the budget constraint given at the group level
   - Based on an optimization formulation
   - Considering the role and social reach of the individuals



SNAzzy Papers

- ☆ "On the Structural Properties of Massive Telecom Call Graphs: Findings and Implications", ACM CIKM 2006.
- ☆ "Analyzing the Structure and Evolution of Massive Telecom Graphs", IEEE TKDE, 2008.
- ☆ "Social Ties and their Relevance to Churn in Mobile Telecom Networks", EDBT 2008.
- ☆ "Large Maximal Cliques Enumeration in Large Sparse Graphs" COMAD, 2009 / CIKM 2008.
- ☆ "Extracting dense communities from telecom call graphs." COMSWARE 2008.
- ☆ "Leveraging Social Networks for Corporate Staffing and Expert Recommendation", IBM Systems Journal, 2010.
- ☆ "Discovery and Analysis of Tight-Knit Communities in Telecom Social Networks", IBM Journal of R&D, 2010.
- ☆ "A shapley value-based approach to discover influential nodes in social networks", IEEE Transactions on Automation Science and Engineering 99 (2010): 1-18.
- ☆ "Topologies of strategically formed social networks based on a generic value function—Allocation rule model." Social Networks 33.1 (2011): 56-69.
- ☆ "Like-minded Communities: Bringing the Familiarity and Similarity Together", WISE 2012.



SNAzzy Papers

- ☆ "Game theoretic models for social network analysis."
  20th international conference companion on World wide web. ACM, 2011.
- ☆ "A game theory inspired, decentralized, local information based algorithm for community detection in social graphs." IEEE Pattern Recognition (ICPR), 2012.
- ☆ "Density Functions subject to a Co-Matroid Constraint",
  FSTTCS 2012.
- ☆ "Viral Marketing for Product Cross-Sell through Social Networks",
  ECML/PKDD 2012.
- ☆ "Design of Viral Marketing Strategies for Product Cross-sell through Social Networks",
  KAIS Journal, 2013.
- ☆ "Bug Resolution Catalysts: Identifying Essential Non-committers from Bug Repositories",
  Mining Software Repositories, 2013.
- ☆ "Link Label Prediction in Signed Social Networks",
  IJCAI 2013.
- ☆ "Computational Analysis of Connectivity Games with Applications to the Investigation of Terrorist Networks", IJCAI 2013.
- ☆ "A Novel and Model Independent Approach for Efficient Influence Maximization in Social Networks."
  WISE 2013.
- ☆ "Strategic Network Formation with Localized Payoffs."
  Studies in Microeconomics 2.1 (2014): 63-119.