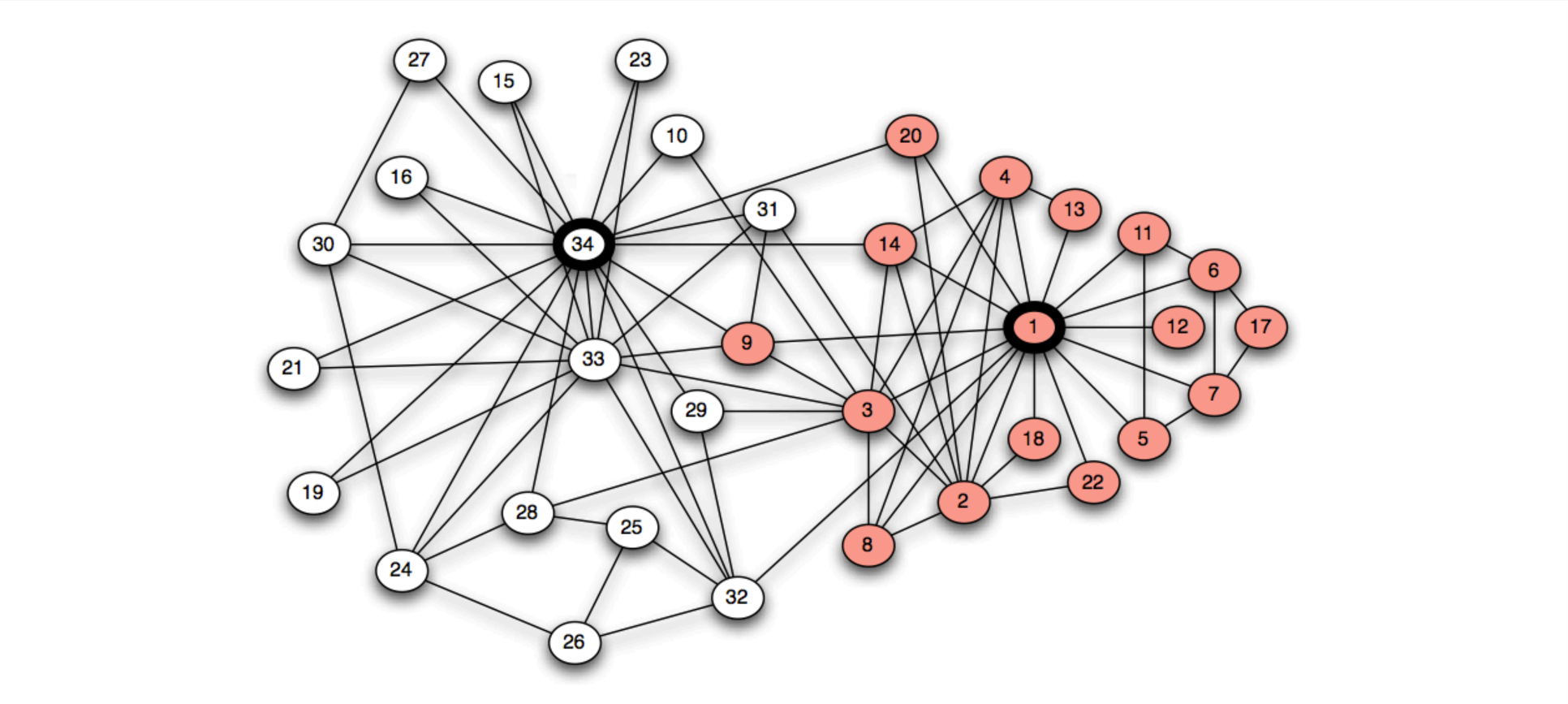
Ahmedabad University

Network Analysis in Telecom

Amit A. Nanavati
Ahmedabad University

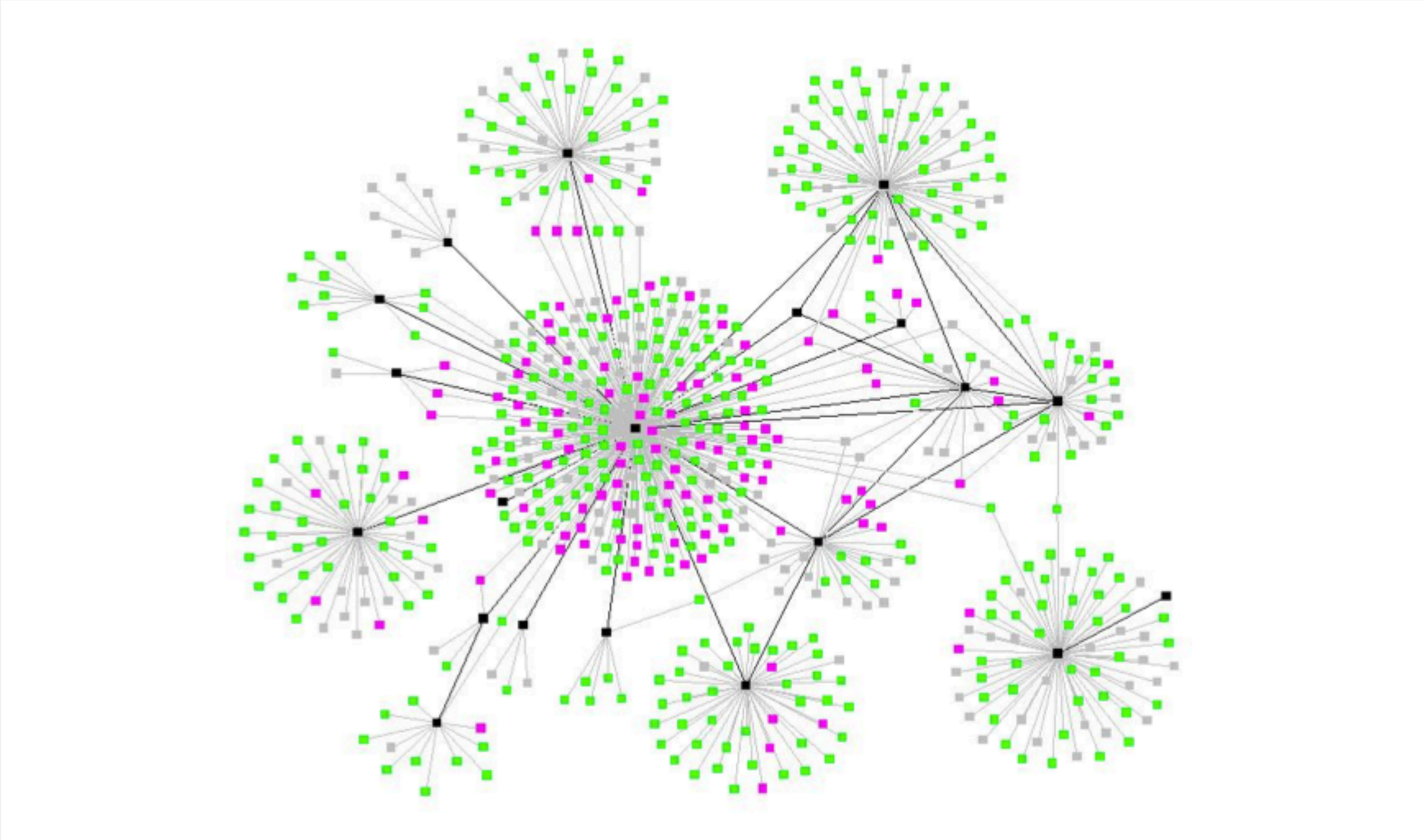
ACM Winter School on Network Science, 2023, Ahmedabad University



The social network of friendships within a 34-person karate club*

*Wayne Zachary. An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33(4):452–473, 1977.

Ahmedabad University

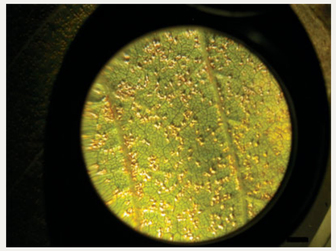


The spread of an epidemic disease (such as the tuberculosis outbreak shown here) is another form of cascading behaviour in a network*.

*McKenzie Andre, Kashef Ijaz, Jon D. Tillinghast, Valdis E. Krebs, Lois A. Diem, Bev- erly Metchock, Theresa Crisp, and Peter D. McElroy. Transmission network analysis to complement routine tuberculosis contact investigations. American Journal of Public Health, 97(3):470–477, 2007.

Networks in Telecom

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

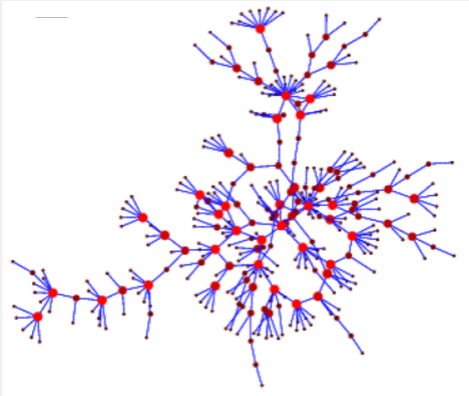


Microscopic view

- 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*

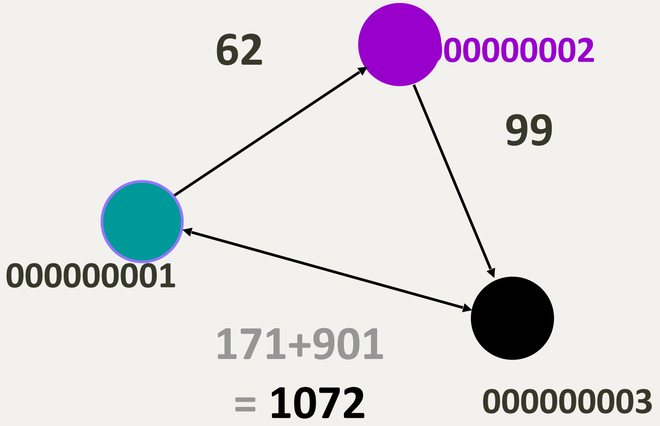


Telescopic view

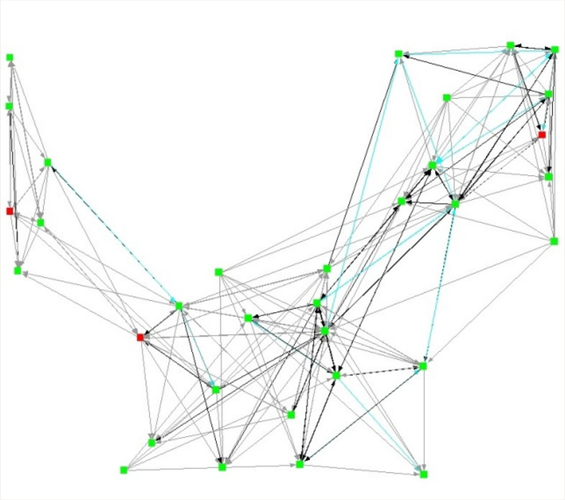
Ahmedabad University

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



Repeat for millions of rows of CDR data



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

* Call Detail Record

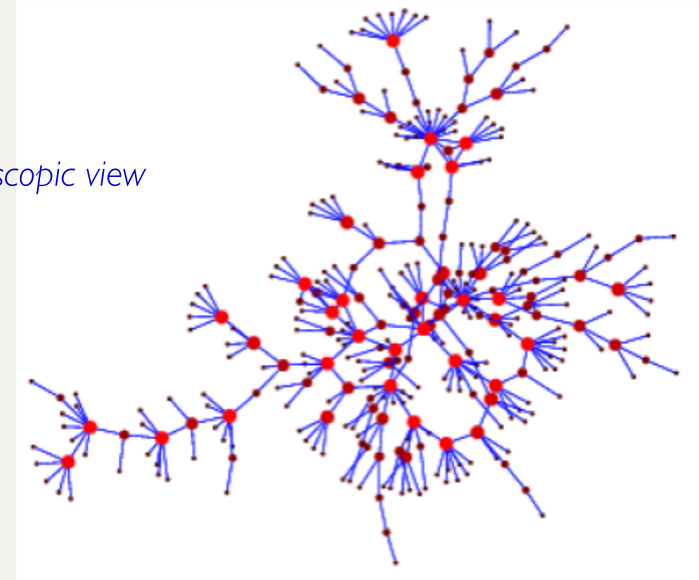
The Setting

- Telecom business analytics typically
  - Analyse only individual customer behaviour
  - "microscopic view"

- Analysing call-graphs
  - "who-calls-whom" graph
    - People are the vertices
    - If two people call each other, there is an edge between their corresponding vertices

  - How is everyone connected with each other ?
  - "telescopic view"

Telescopic view



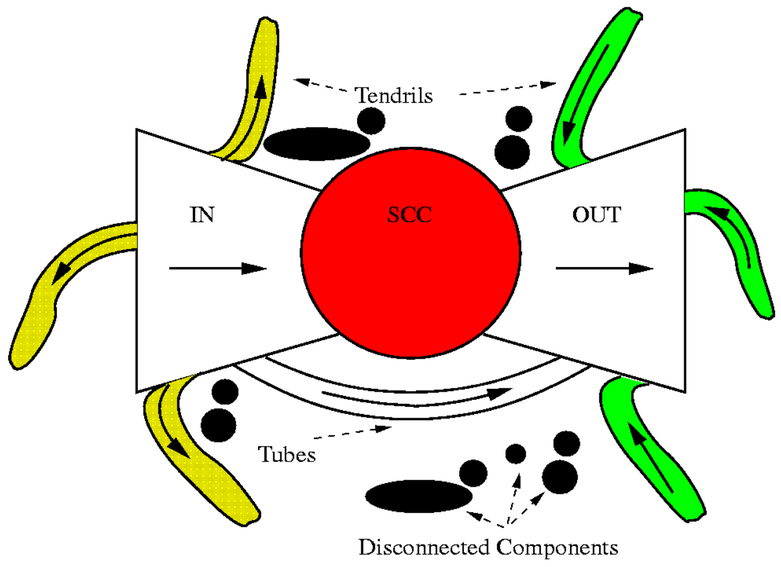
Goals and Contributions

- Studies a broad set of parameters that reveal various structural properties of mobile call graphs.
- Novel techniques to determine shape of large graphs
- The Treasure-Hunt Model, an edge-distribution based model, probably the first topological model for mobile call graphs.
- Attempt to emphasise the practical implications (business insights) of our findings

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)

- 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



"bow-tie"

Powers laws in Telephony and the WWW

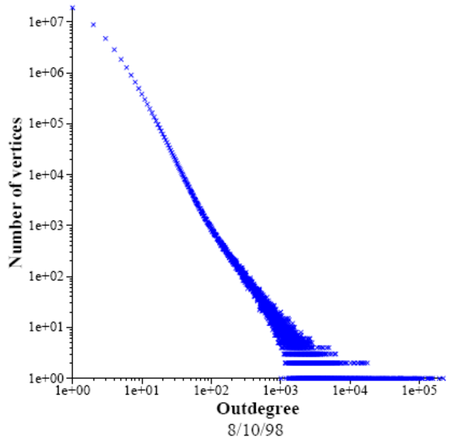


Figure 1: *The number of vertices for each possible outdegree for the call graph of a typical day.*

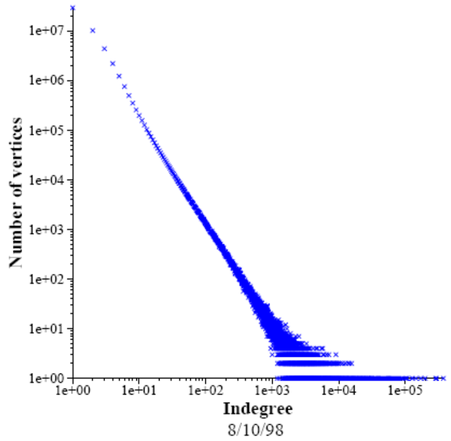


Figure 2: *The number of vertices for each possible indegree for the call graph of a typical day.*

Aiello et al, AT&T Labs (2000)

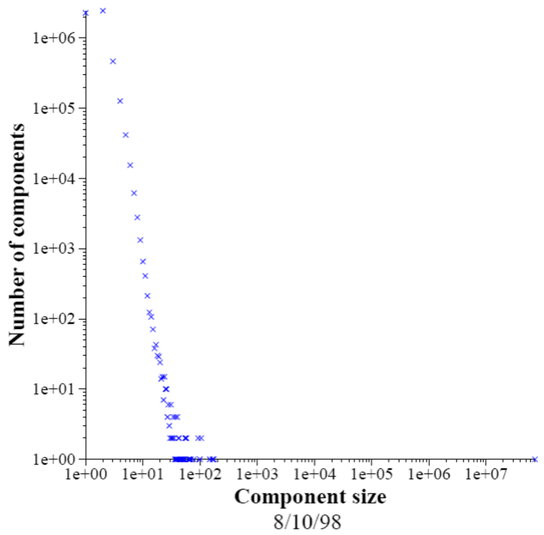


Figure 3: *The number of connected components for each possible component size for the call graph of a typical day.*

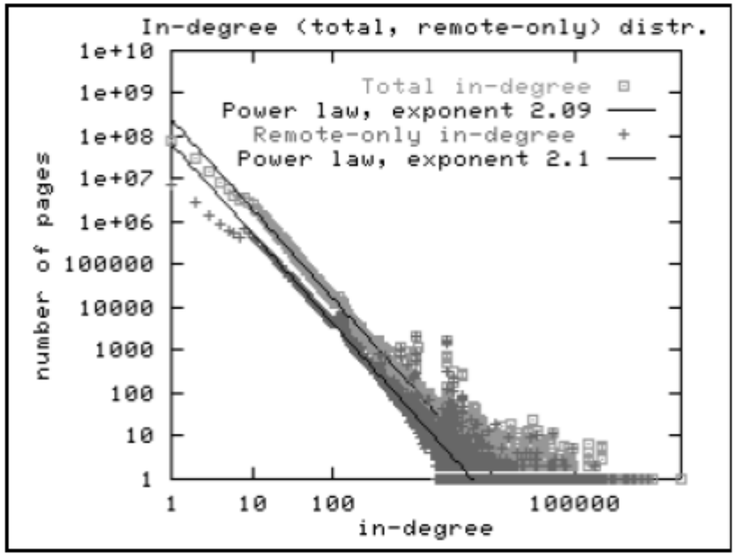


Figure 2: In-degree distribution.

Kumar et al, IBM Research (2000)

Data Sources / Preprocessing

- The study was done for intra-region calls, and does not include long distance or international calls.
- For two of the regions, we collected all the calls made in a week, and for the other two, we collected all the calls made in a month.
- Further, very short duration calls (less than 10 seconds) have been ignored as missed calls and wrong calls since they may yield incorrect results.
- Multiple calls between any two user or nodes is treated as a single edge.

Table 1: Details of data set used.

Region	Nodes	Edges	Period	Avg. Deg.	Type
$\mathcal{A}$-Region	224418	1816285	1 month	8.09	Dir.
$\mathcal{B}$-Region	1250656	4514528	1 week	3.6	Dir.
$\mathcal{C}$-Region	989573	4313797	1 week	4.35	Dir.
$\mathcal{D}$-Region	407332	1456645	1 month	3.57	Dir.

Structural Properties

- Degree distributions
  - All parameters quite close
  - Very few vertices with high in / out degree
  - Easier to target

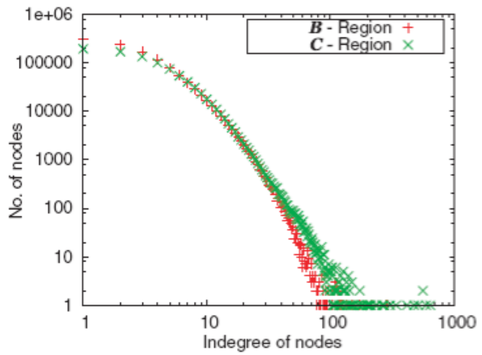


Figure 1: In-degree distribution (γ_{in} is 2.89961 and 2.89961)

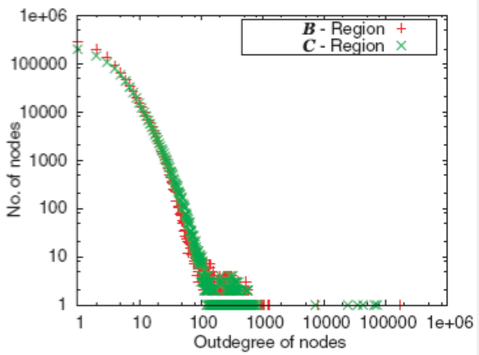


Figure 2: Out-degree distribution (γ_{out} is 1.70808 and 1.70808)

Results from our dataset (directed graphs)				
Region	$\mathcal{A}$-Reg.	$\mathcal{B}$-Reg.	$\mathcal{C}$-Reg.	$\mathcal{D}$-Reg.
In	2.76591	2.85924	2.89961	2.844
Out	1.50292	1.71374	1.70808	1.97345
Results from other datasets (directed graphs)				
Metric	Email	WWW	Software packages	
In	1.5	2.1	1.6	
Out	2.0	2.72	1.4	

Figure 3: Power law exponent values

Ahmedabad University

Structural Properties

- Degree correlations
  - Single vertex IN-OUT correlation
    - Indegree, outdegree correlated upto a point
    - Beyond that, high outdegree could be salesmen
    - High indegree micro businesses or highly influential people ?

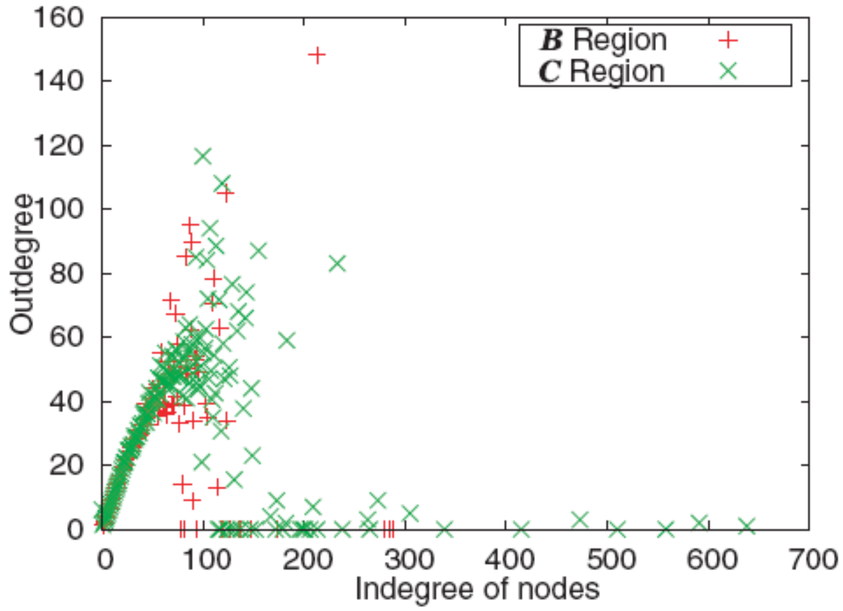
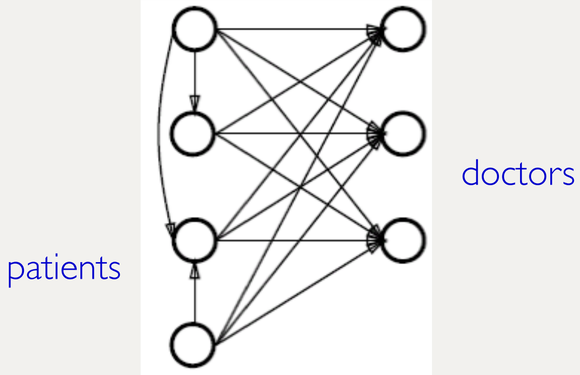


Figure 4: Correlation between in-degree and out-degree for regions $\mathcal{B}$ and $\mathcal{C}$

Structural Properties



- Bipartite cores
  - Community: one set of people calling another set of people, but may not necessarily know each other
  - Telecom service providers can target communities with appropriate incentives

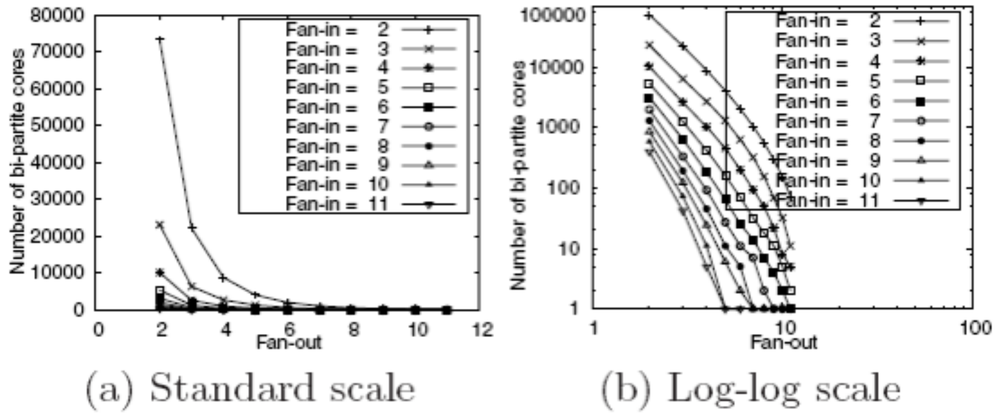


Figure 6: Distribution of bipartite cores for the $\mathcal{B}$-region

Structural Properties

- Cliques
  - Mobile Telecom operators can proactively target cliques with group incentives

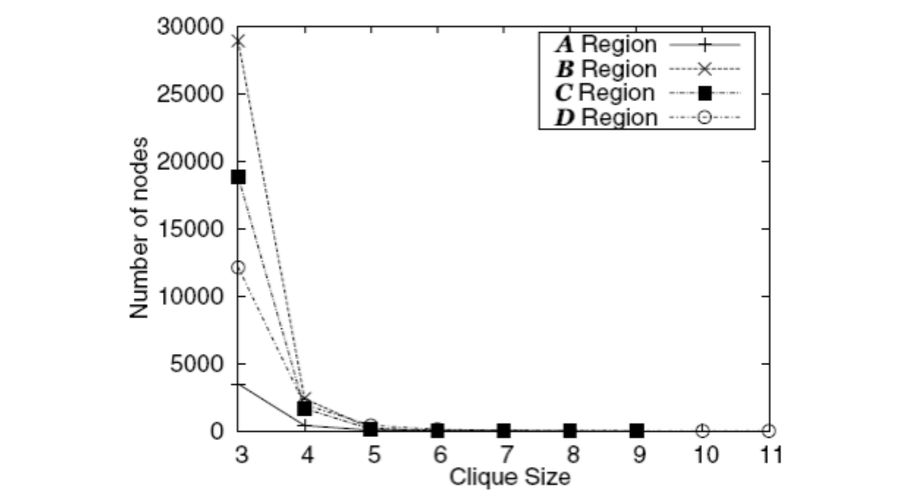


Figure 7: Clique size distribution for all the regions $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$

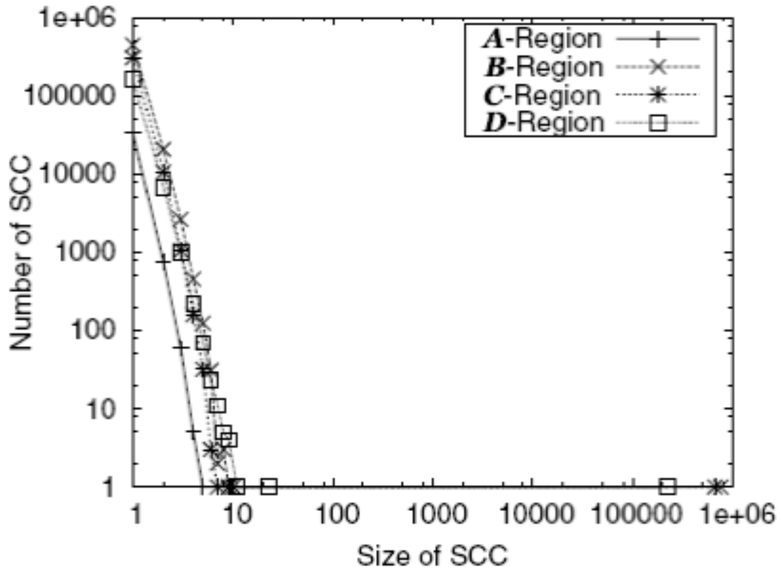
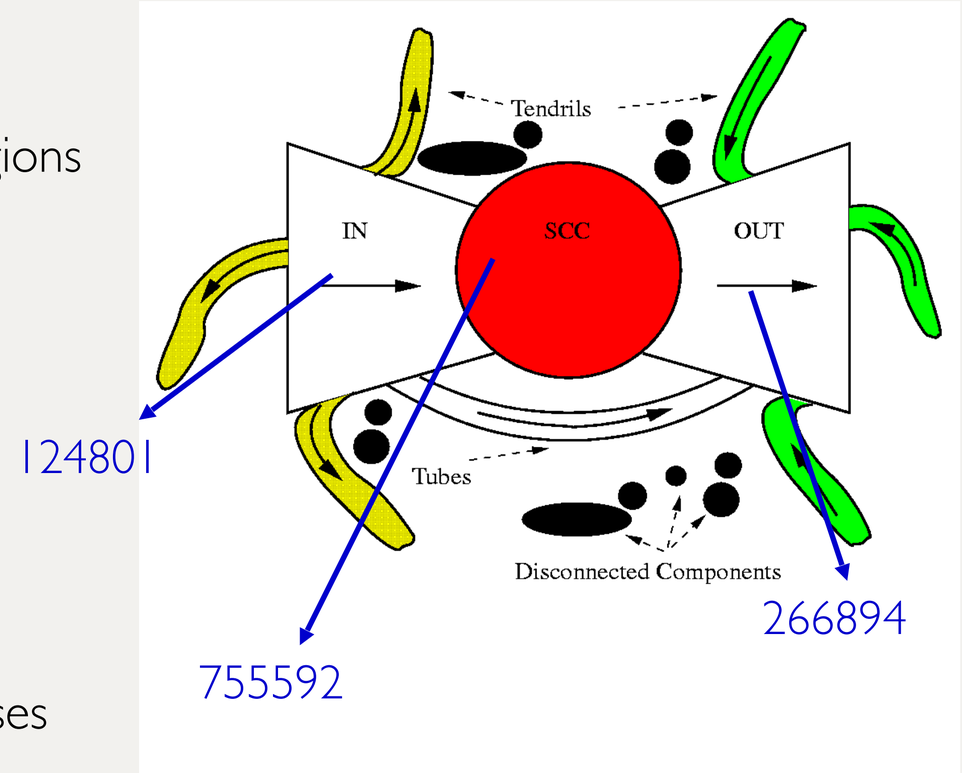


Figure 9: Distribution of SCCs in log-log scale

- Strongly connected components
  - A giant component exists in all graphs
  - Sizes of SCCs follow a power law distribution
  - The largest SCC is significantly larger than the next one
  - This is similar to the WWW graph

Shape of call-graphs

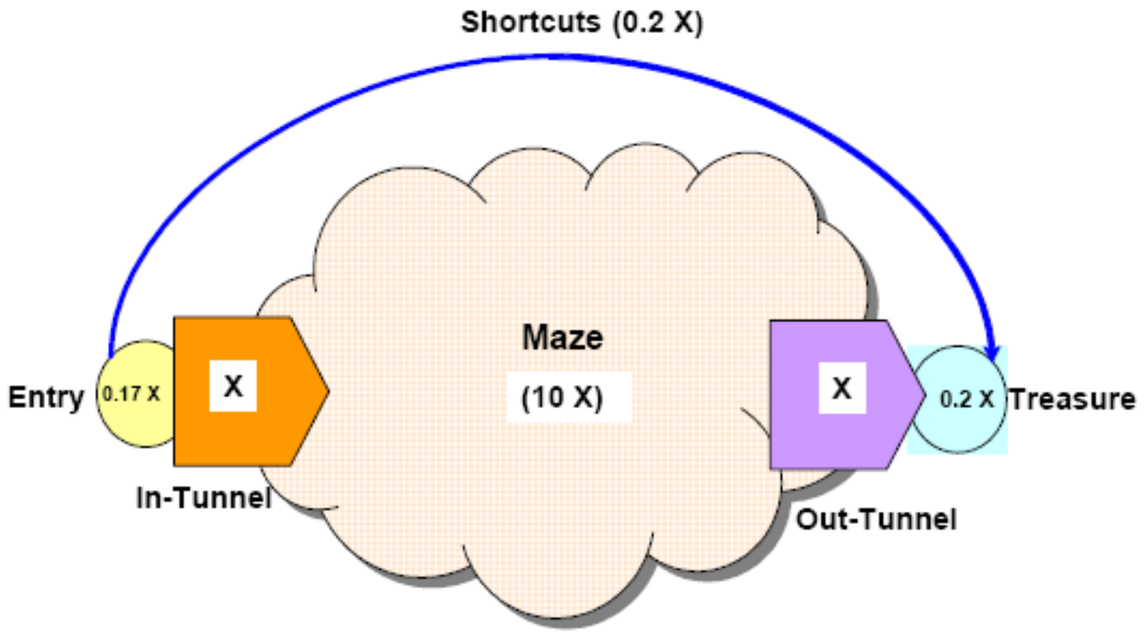
- "Bow-Tie" analysis
  - For all regions, the SCC was much bigger than the IN and OUT regions
    - For WWW, the sizes are nearly the same
      - 44 million, 56 million (SCC), 44 million
  - The "Bow-Tie" model does not fit our graphs
  - Perhaps, another vertex distribution based model
  - Instead, we found that an edge distribution based model characterises our graphs better



Shape of call-graphs

Table 7: Definition of subgraphs

Subgraph name	Definition
IN-IN	Subgraph containing the edges only between nodes of IN region
IN-SCC	Subgraph consisting edges one end from IN region and another from SCC
IN-OUT	Subgraph consisting edges one end from IN region and another from OUT
SCC-SCC	Subgraph containing the edges only between nodes of SCC region
SCC-OUT	Subgraph consisting edges one end from SCC region and another from OUT
OUT-OUT	Subgraph containing the edges only between nodes of OUT region



Treasure-Hunt Model

- "On the structural properties of massive telecom call graphs: findings and implications" ACM CIKM 2006. **(234 citations)**

Shape of call-graphs

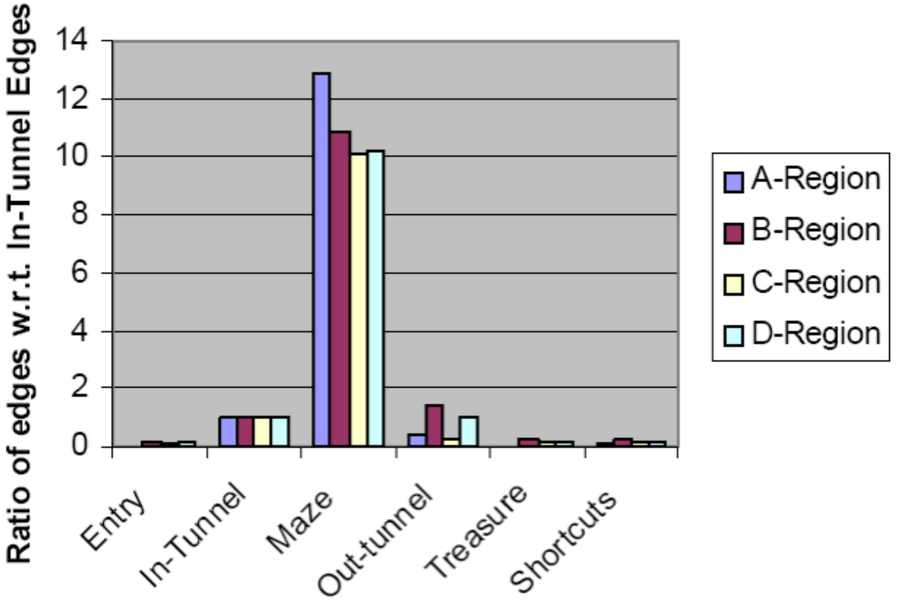
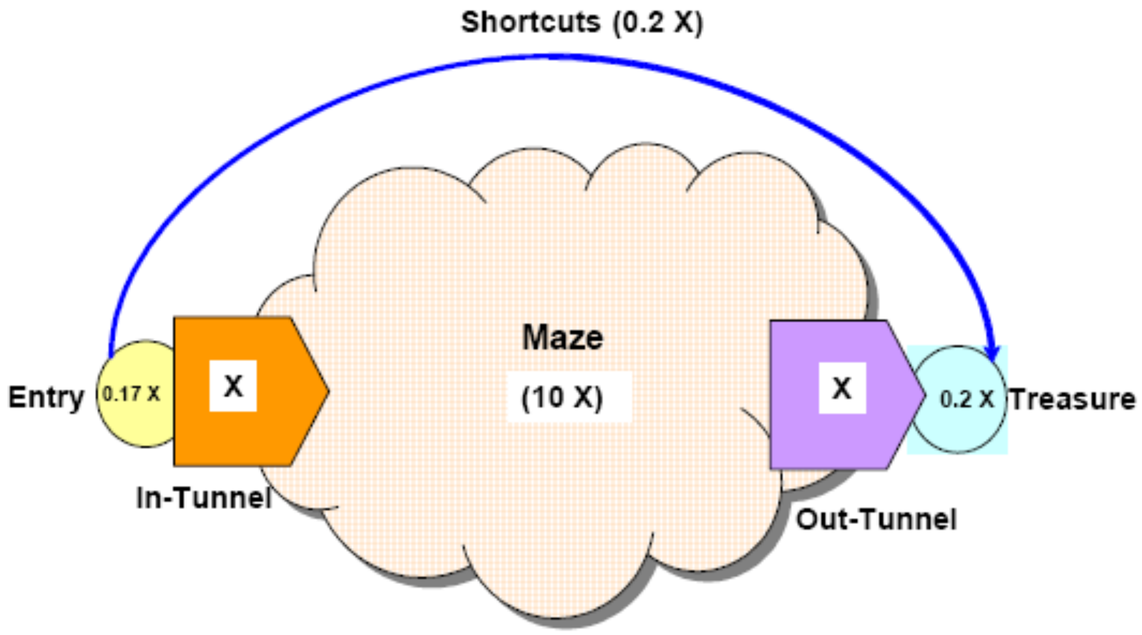


Figure 12: Distribution of edge fractions in different components



Treasure-Hunt Model

- Finding the shape can be helpful in many ways
  - Word-of-mouth advertising might work, since the graph is connected
  - For a new service rollout, the vertices in the Entry and In-Tunnel zones might qualify as better starting points
  - This provides another way to segment customers

Ahmedabad University

Telecom Business Usecases

- 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?

Ahmedabad University

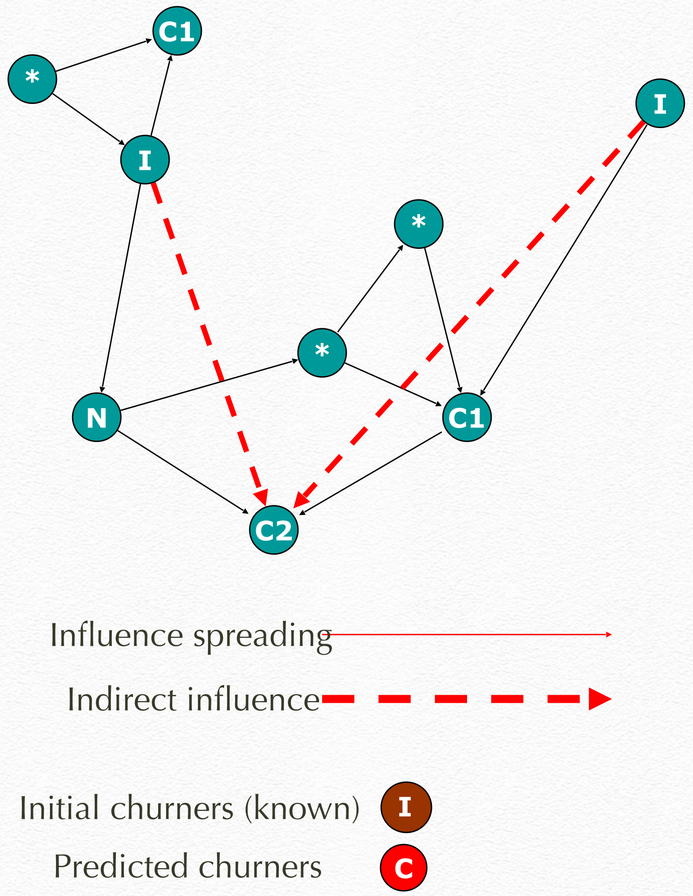
Churn in Telecom

- 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

Ahmedabad University

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.
  - Complements traditional churn models.

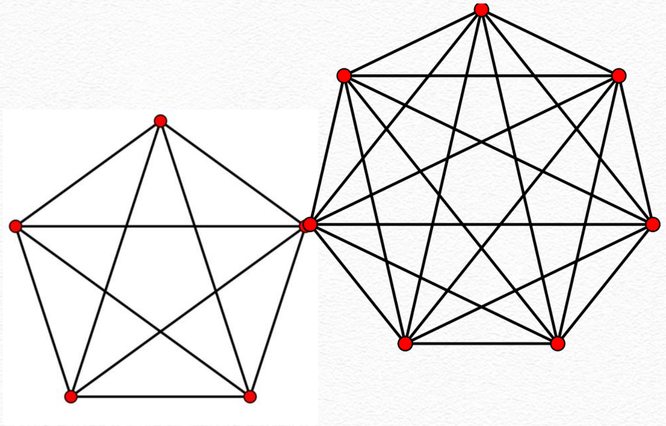


- "Social Ties and their Relevance to Churn in Mobile Telecom Networks", EDBT 2008. **(Test of Time Award, EDBT 2018)**

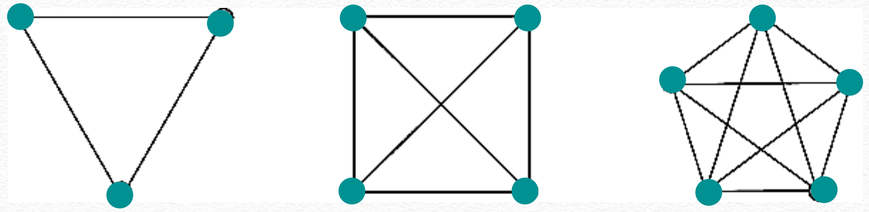
Ahmedabad University

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



K_3, K_4, K_5

- Business Value
  - Telcos can offer automatic discounts for calls within a clique
  - To increase stickiness and loyalty
  - Can be leveraged by SNAzzy churn prediction

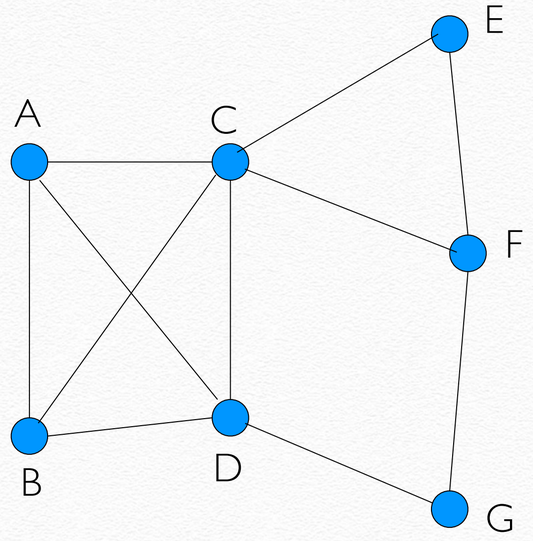
- "Large Maximal Cliques Enumeration in Large Sparse Graphs"COMAD, 2009 / CIKM 2008. **(68 citations)**

Community Identification | Cliques

- Identifying cliques is a computationally hard problem
- Suppose we are looking for cliques larger than a particular size (say, 4)
- Degree-based (DF) Filtering
  - *All nodes having less than 3 neighbours can be removed from consideration (why?)*
- Social Network Neighbour-based (SNN) filtering
  - *An edge between a pair of nodes (p, q) can be removed if they do not have 2 common neighbours (why?)*
- Repeat DF and SSNF till you can no longer do so.

- "Large Maximal Cliques Enumeration in Large Sparse Graphs"COMAD, 2009 / CIKM 2008. **(68 citations)**

Community Identification | Cliques

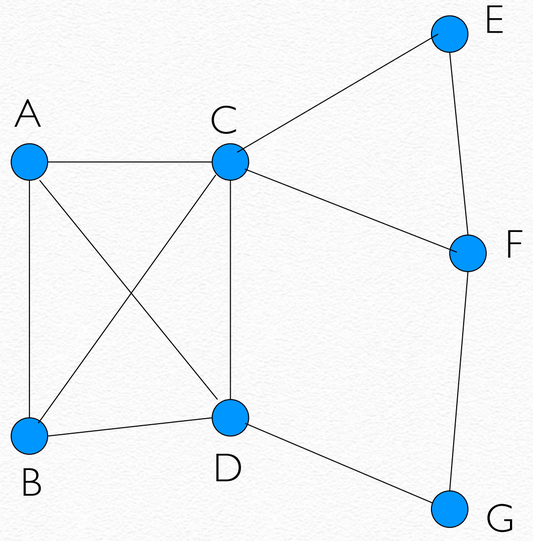


- Identifying cliques is a computationally hard problem
- Suppose we are looking for cliques larger than a particular size (say, 4)
- Degree-based (DF) Filtering
  - *All nodes having less than 3 neighbours can be removed from consideration (why?)*
- Social Network Neighbour-based (SNN) filtering
  - *An edge between a pair of nodes (p, q) can be removed if they do not have 2 common neighbours (why?)*
- Repeat DF and SSNF till you can no longer do so.

- "Large Maximal Cliques Enumeration in Large Sparse Graphs"COMAD, 2009 / CIKM 2008. **(68 citations)**

Ahmedabad University

Community Identification | Cliques



- Identifying cliques is a computationally hard problem
- Suppose we are looking for cliques larger than a particular size
  - *Let's take 4 (or more) as an example in the figure*
- Degree-based (DF) Filtering
  - *All nodes having less than 3 neighbours can be removed from consideration (why?)*
  - *E, G can be removed, along with the edges attached to them.*
- Social Network Neighbour-based (SNN) filtering
  - *An edge between a pair of nodes (p, q) can be removed if they do not have 2 common neighbours (why?)*
  - *The edge (F, C) can be removed.*
- Repeat DF and SSNF till you can no longer do so.

- "Large Maximal Cliques Enumeration in Large Sparse Graphs"COMAD, 2009 / CIKM 2008. **(68 citations)**

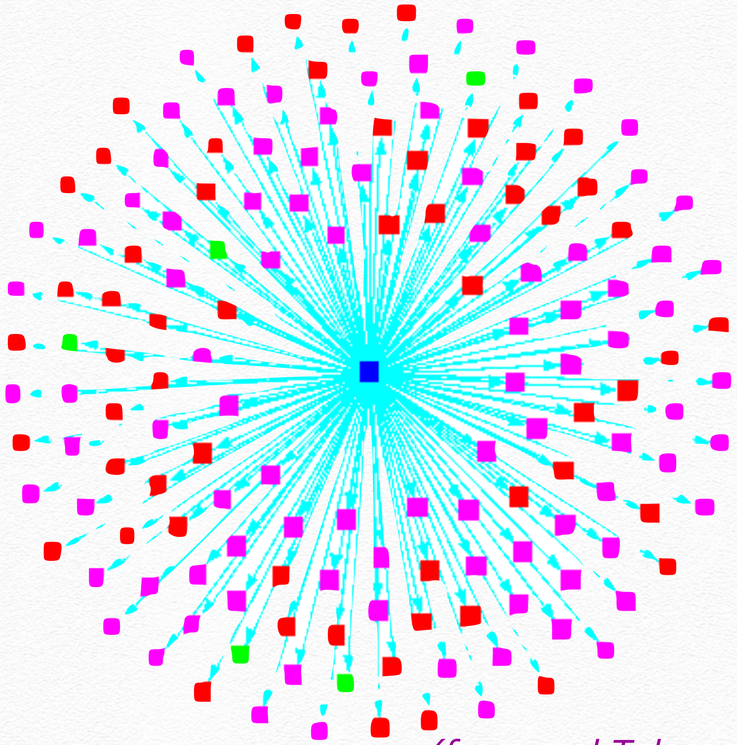
Ahmedabad University

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

- 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.



(from real Telecom data)

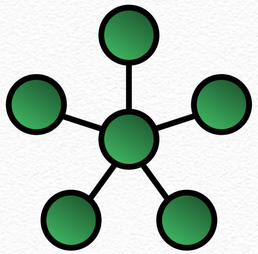
Platinum Customer
Regular Customer
CDMA Competitor
GSM Competitor

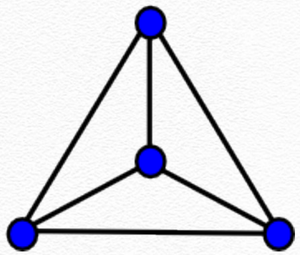
- "Systems and methods for finding star structures as communities in networks" US Patent 8,363,570. **(IBM high-value patent award)**

Star-like Communities

- Local, sparse sub-networks
  - The "hub" in a "star" network plays a central role in enabling shortest path communication between the "spokes"
  - "Hub" of a "clique" plays no role in enabling communication between members of the clique
  - Role of a vertex in enabling communication between its neighbours?

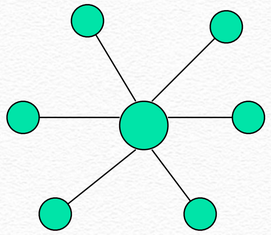
- Star value of a vertex measures how close the local adjacent sub-network of the vertex is to a star motif
  - For directed and undirected networks



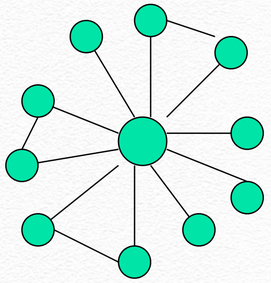


Star Value: *communication workload on the hub*

undirected



"perfect" star



"imperfect" star

Communication Workload $L(v_k) = \frac{\| C_k \| (\| C_k \| - 1)}{2} - m_k.$

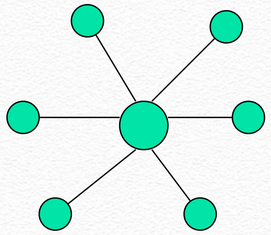
- C_k is the number of spokes
- C_k = 6 for the perfect star
- The communication workload of the hub (centre):
  - $\frac{6 \times 5}{2}$ - 0 = 15

- C_k is the number of spokes
- C_k = 10 for the perfect star
- The communication workload of the hub (centre):
  - $\frac{10 \times 9}{2}$ - 3 = 42

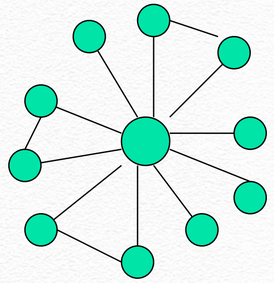
Ahmedabad University

Star Value: *communication workload on the hub*

undirected



"perfect" star



"imperfect" star

Communication Workload

$$L(v_k) = \frac{\| C_k \| (\| C_k \| - 1)}{2} - m_k.$$

Star Value

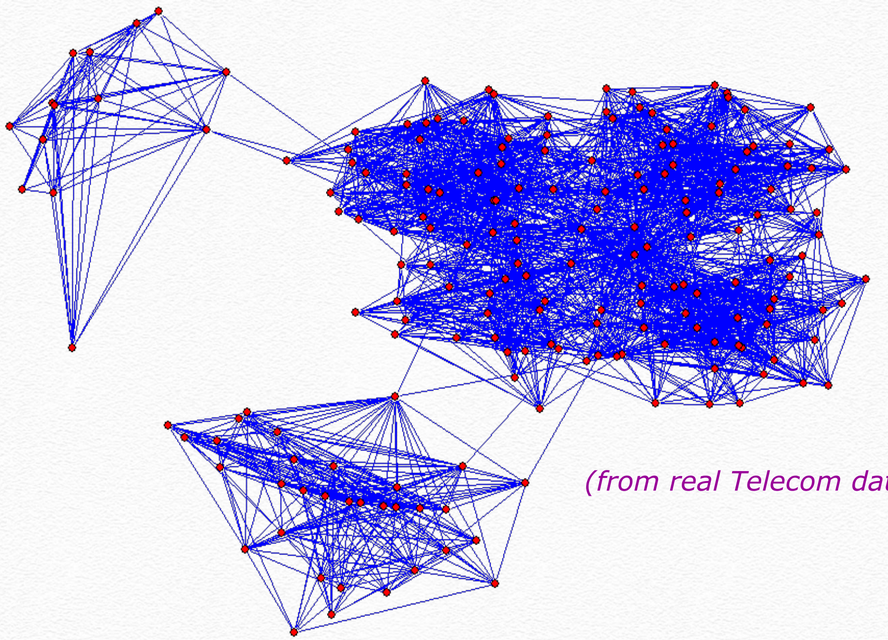
$$SV(v_k) = \frac{2 * L(v_k)}{||C_k||^2} = 1 - \frac{2 * m_k + \| C_k \|}{\| C_k \|^2}.$$

The star value $SV(v_k)$ is the normalized workload such that for a clique it is zero and for a star it tends to one as $||C_k||$ grows

Ahmedabad University

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!



(from real Telecom data)

- Output
  - Dense communities and members identified

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

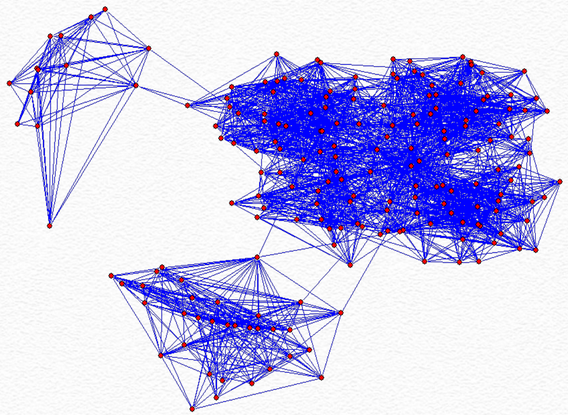
- "Extracting dense communities from telecom call graphs." COMSWARE 2008. **(18 citations)**

Ahmedabad University

Community Identification | Dense Communities

- Given an undirected graph $G = (V, E)$, a subset of vertices $S \subseteq V$, $E(S)$ is defined to the set of edges induced by S,

$$E(S) = \{(i, j) \mid i, j \in S\}$$



- The density of a given subset of vertices $S \subseteq V$ is given by

$$f(S) = \frac{|E(S)|}{|S|}$$

- The density of the densest subgraph is defined to be the density of the graph G, denoted by $f(G)$

$$f(G) = \max_{S \subseteq V} \{f(S)\}$$

- "Extracting dense communities from telecom call graphs." COMSWARE 2008. **(18 citations)**

Community Identification | Dense Communities

- Charikar* showed that the densest subgraph can be computed in polynomial time.
- He also gave a greedy heuristic which is guaranteed to compute a 2-approximation for $f(G)$
  - Initially, $S_0 = V$. Compute $f(S_0)$.
  - Delete the vertex with minimum degree to obtain S_1. Compute $f(S_1)$.
  - Runs n iterations where $n = |V|$.
  - Now, pick the maximum among $f(S_0), \ldots, f(S_{n-1})$
- Another related notion of dense community is a subgraph in which the minimum degree of a node is maximized, call it as max-min degree problem.

*M. Charikar, "Greedy approximation algorithms for finding dense components in a graph," in (APPROX), 2000.

Detecting Internal/Rotational Churn

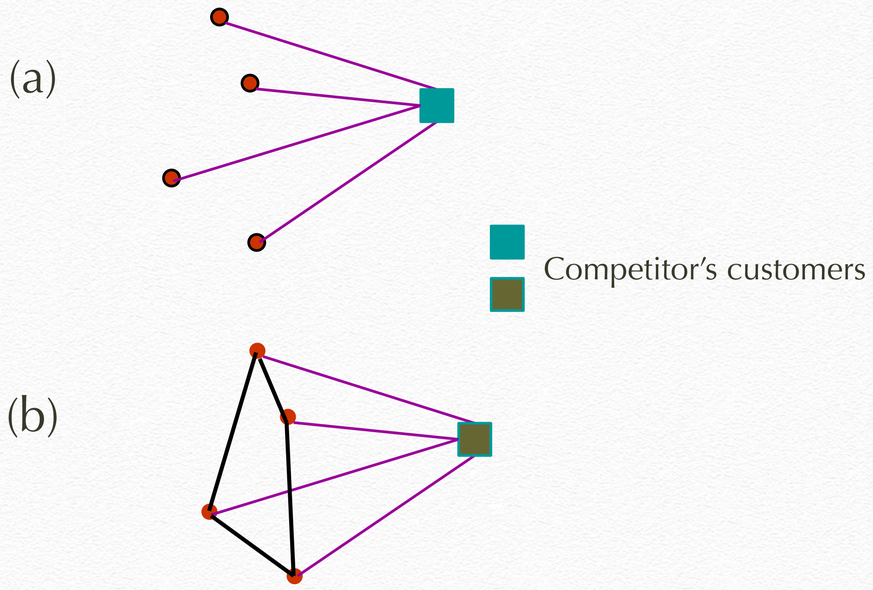
- Is my churn real?
  - Are people simply discarding old SIMS and replacing them with new ones?
  - Are any new plans cannibalising the existing ones?
  - If it is the same person, then it is likely that he is calling the same set of people with the same set of frequencies.

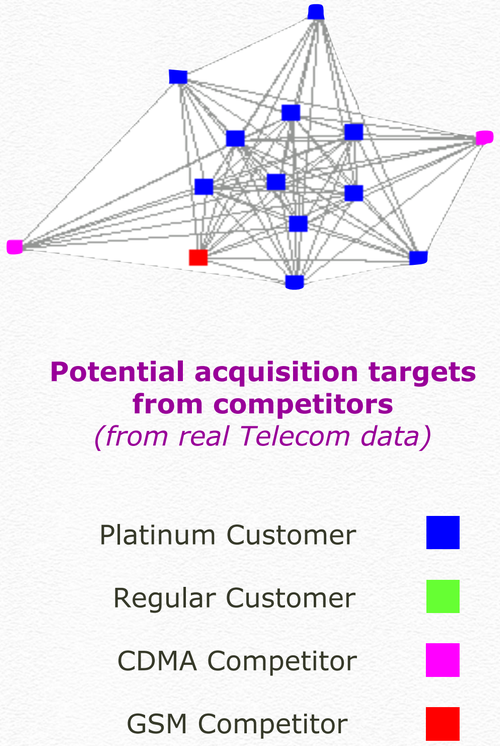
- Output
  - List of internal churners (old number, new number) identified.

- 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.



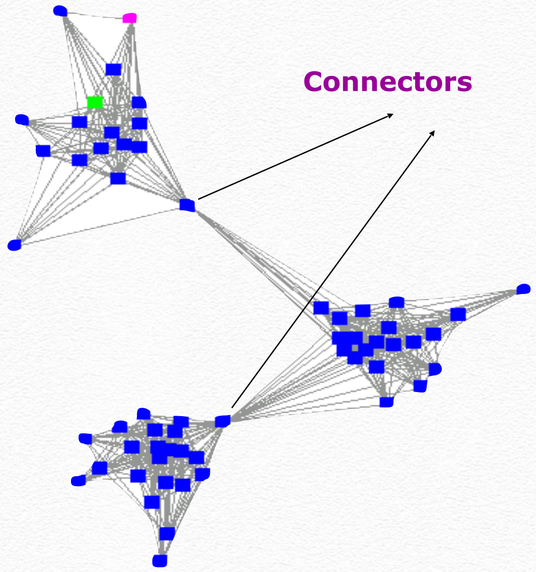


- Output
  - Ranked list of acquisition targets
- Business Value
  - Improves hit ratio for acquisition

Ahmedabad University

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



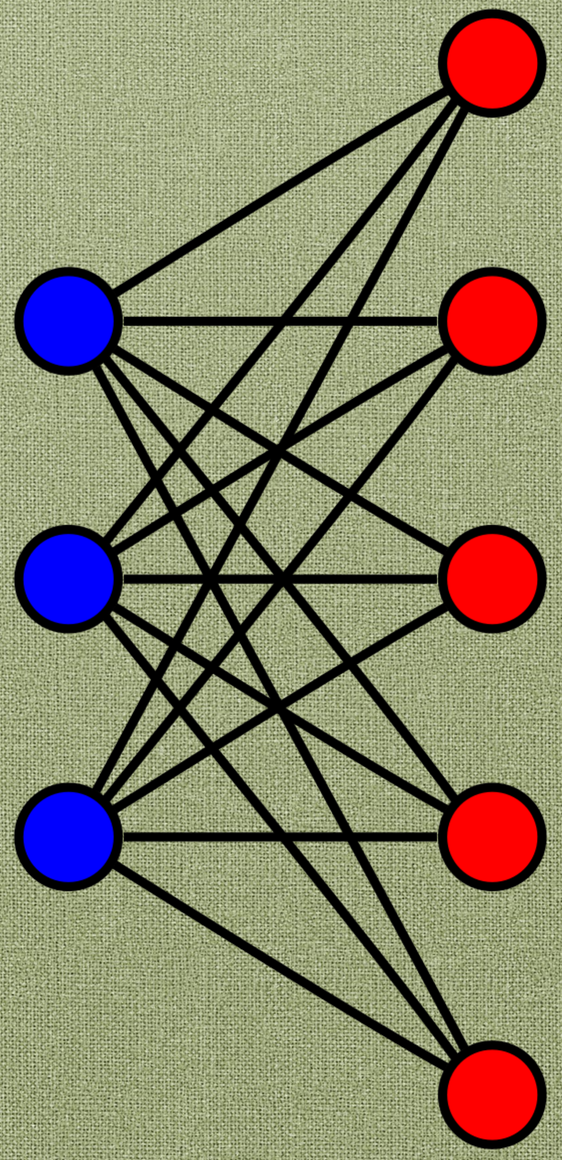
(from real Telecom data)

Platinum Customer
Regular Customer
CDMA Competitor
GSM Competitor

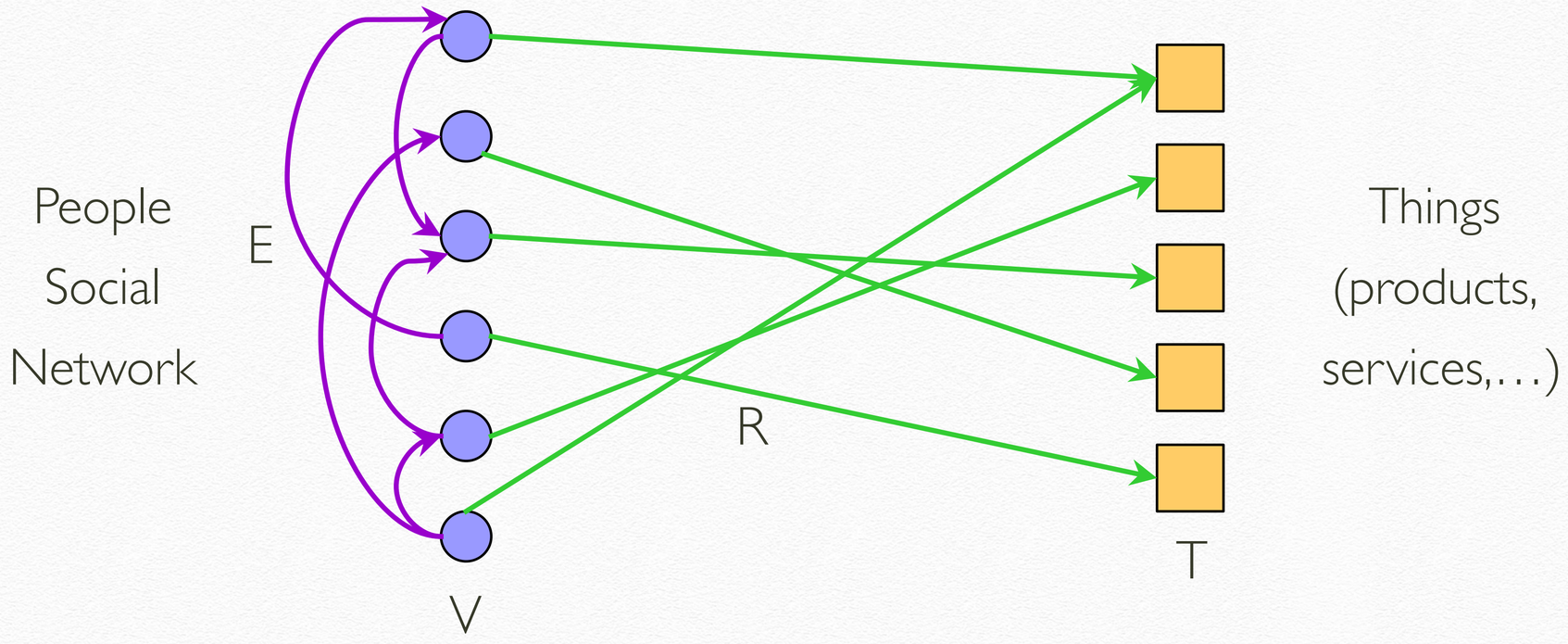
Vaule-Added-Service Usage Analysis

What are the communities that are interested in a particular service?

Which products are (not) being bought by communities?



Setting



Who's buying (interested in) what ?

Ahmedabad University

Like-minded Communities

Bringing the Familiarity and Similarity Together

- 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

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 $\Rightarrow$ 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, jam, bread}

- bread ⇒ jam

frequent itemset

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

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

    = 2/3 = 0.6 = 60%

Association Rules

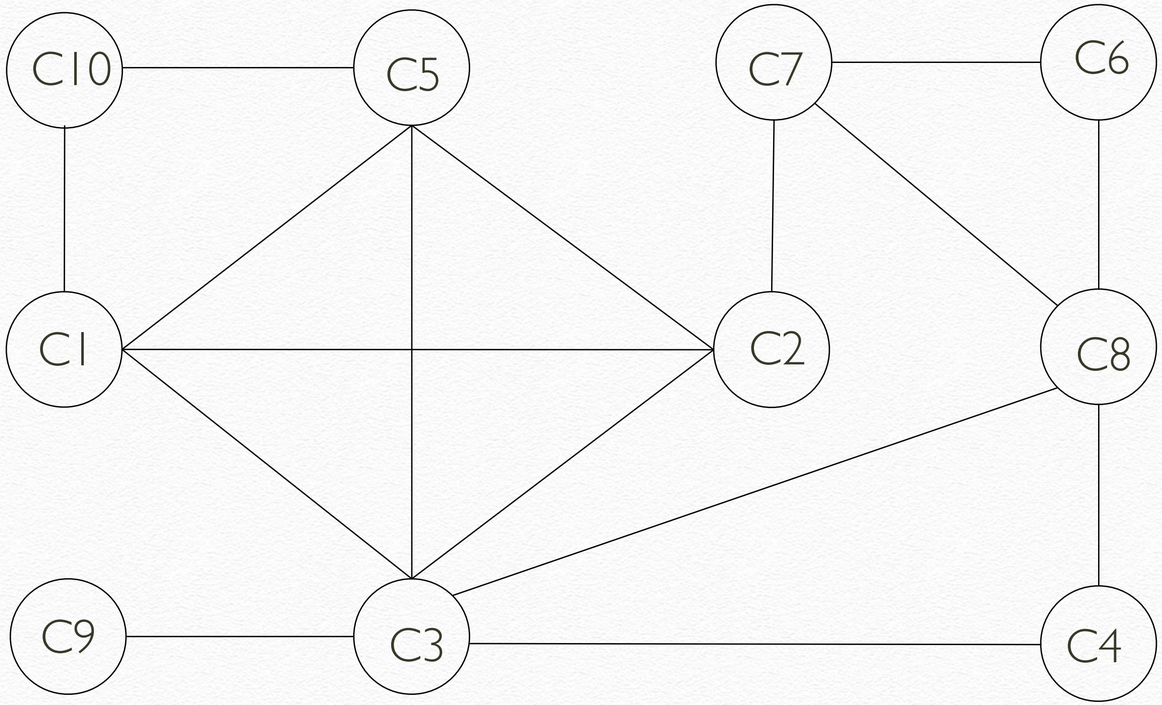
- A set of items $I = \{ i_1, i_2, \ldots, i_N \}$
- A transaction t is a subset of I
- The complete set of transactions D

- We say $A \Rightarrow B$,

  - $\dfrac{\text{Number of transactions including items in A and B}}{\text{Total number of transactions}} \geq$ *minsup*

  - $\dfrac{\text{Number of transactions including items in A and B}}{\text{Number of transactions including A}} \geq$ *minconf*

An Example



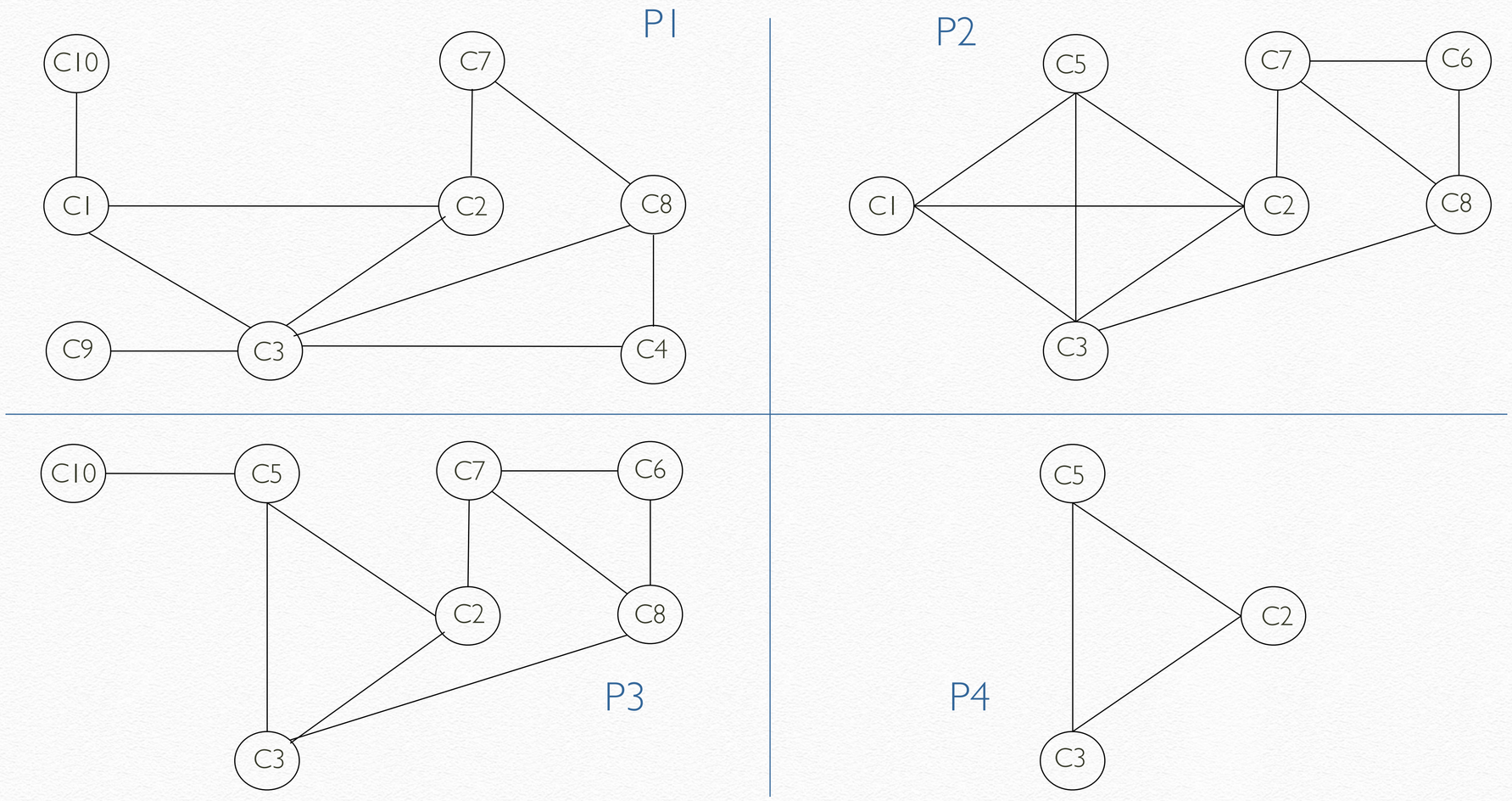
Sample social network

	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

Sample purchase history

Ahmedabad University

Induced subgraphs for products

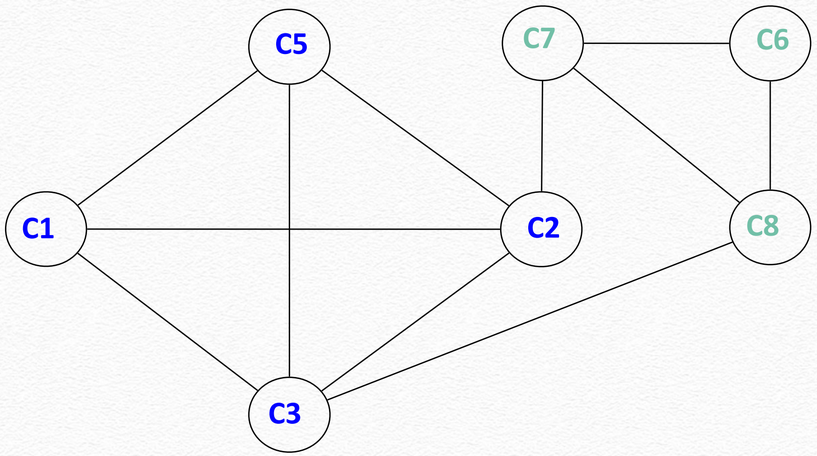


Ahmedabad University

Frequent Itemset Analysis

- 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

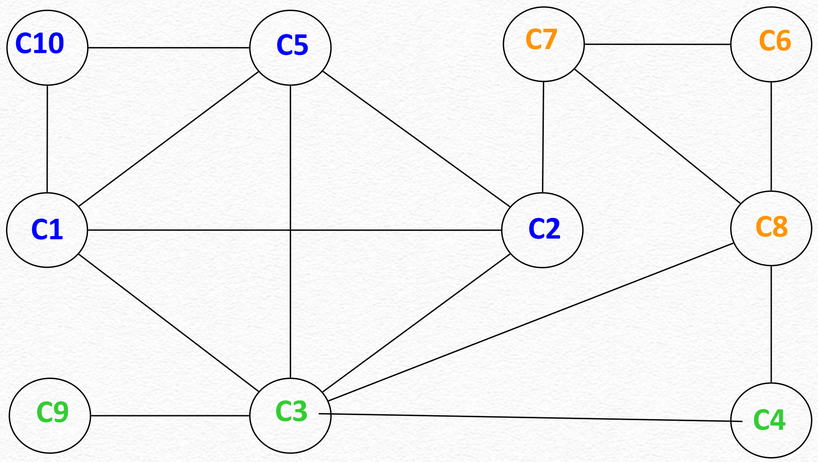
Example continued



Like-minded Communities:
{C1, C2, C3, C5}
{C6, C7, C8}

Like-mindedness: 0.77157

Modularity: 0.28099



Communities:
{C1, C2, C5, C10}
{C3, C4, C9}
{C6, C7, C8}

Like-mindedness: 0.66553

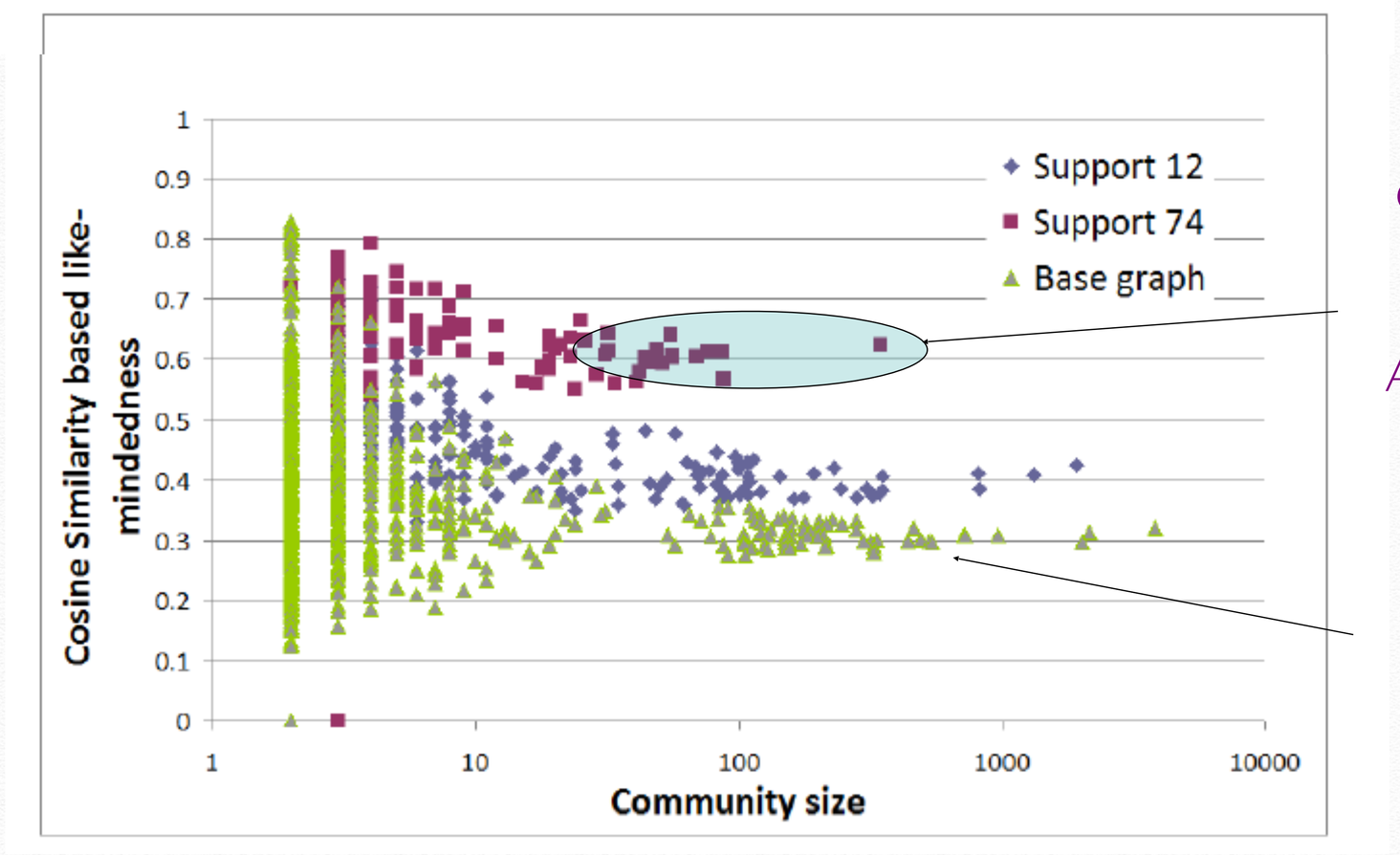
Modularity: 0.27539

- "Like-minded Communities: Bringing the Familiarity and Similarity Together", WISE 2012. **("Runner-up" Award)**

Ahmedabad University

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.



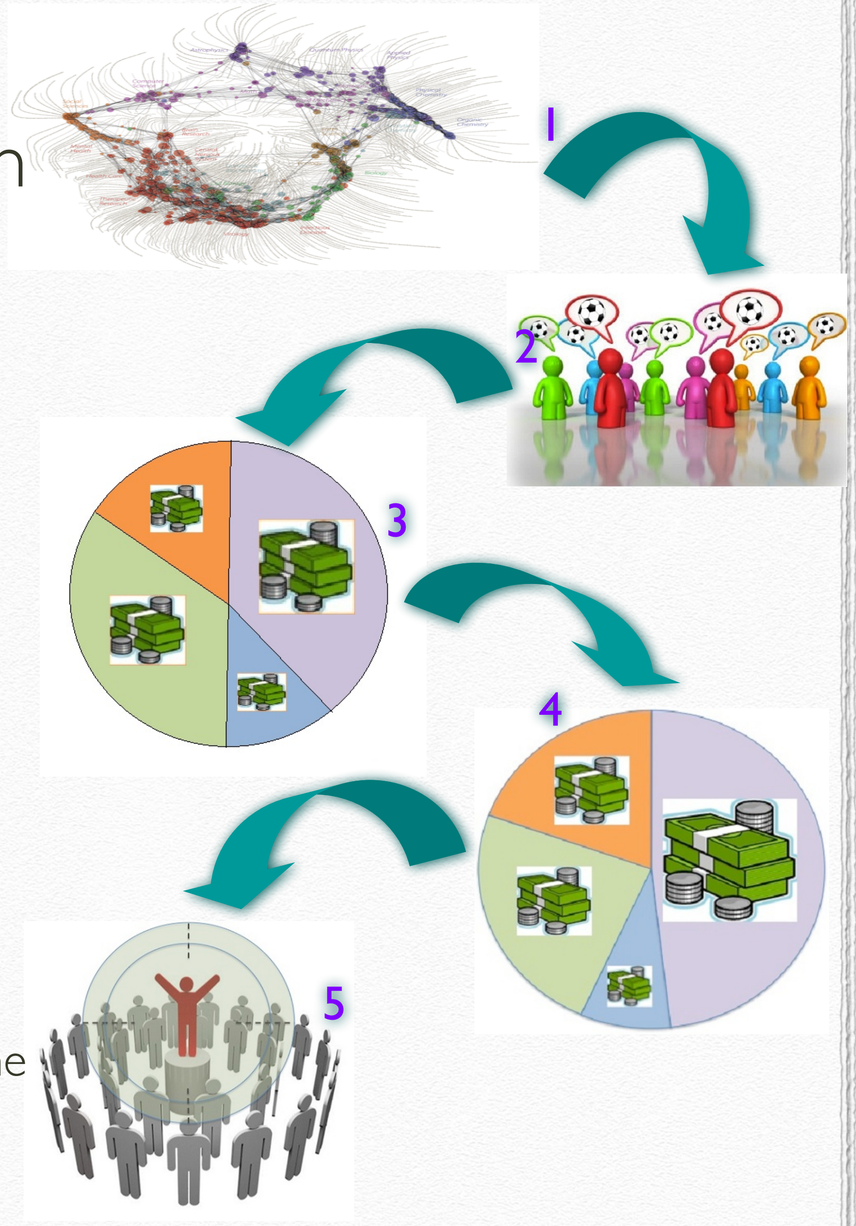
Goldmine communities Found using Vibes

High like-mindedness
And (relatively) large size

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

Ahmedabad University

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

Ahmedabad University

Thank you!

Amit A. Nanavati
Ahmedabad University