On network analysis and user behavior

Ramayya Krishnan
iLab, The H. John Heinz III College
Carnegie Mellon University
Pittsburgh, PA
rk2x@cmu.edu

Outline

- **Two examples**
  - **Intra-organizational KM – the role of triadic closure or cliques in determining user behavior**
  - **Product adoption – the role of social influence vs. homophily**
- **Key points**
  - **Multi-disciplinary perspective that blends computational and social science is needed**
  - **New estimation methods to work with novel data sets**
  - **Need for new methods to design and conduct experiments in a networked world**

Example 1: Social Media and Knowledge Management in a Global Organization

Sample data posting of query and responses

threadid	associateid	postedtime	messagetype	subject	message
{20070110-	138242	2007-01-10 06:41:15	Query	Panel Creation in REXX	Hi,
{20070110-	122971	2007-01-10 07:42:54	Response	Re: Panel Creation in REXX	<p>For retaining the input panel
{20070110-	107246	2007-01-10 13:20:24	Response	Re: Panel Creation in REXX	<p> You are not creating the
{20070110-	128623	2007-01-17 07:19:18	Response	Re: Panel Creation in REXX	<p> No need to VPUT you can
{20070110-	129498	2007-03-01 12:31:42	Response	Re: Panel Creation in REXX	<p>it's simple .. if var1 var2 are the
{20070110-	107246	2007-03-01 13:49:16	Response	Re: Panel Creation in REXX	<p>TYPE(INPUT) is to define the
{20070110-	125034	2007-04-14 07:17:32	Response	Re: Panel Creation in REXX	<p>You can use the command
{20070110-	107246	2007-04-14 23:43:30	Response	Re: Panel Creation in REXX	<p> <em><strong>ADDRESS

Sample Query

- Query on: Singleton class and threads in Java
- Responses:

1. Singleton class means that any given time only one instance of the class is present, in one JVM. So, it is present at JVM level.
2. The thing is if two users(on two different machines which has separate JVMs) are requesting for singleton class then both can get one-one instance of that class in their JVM.

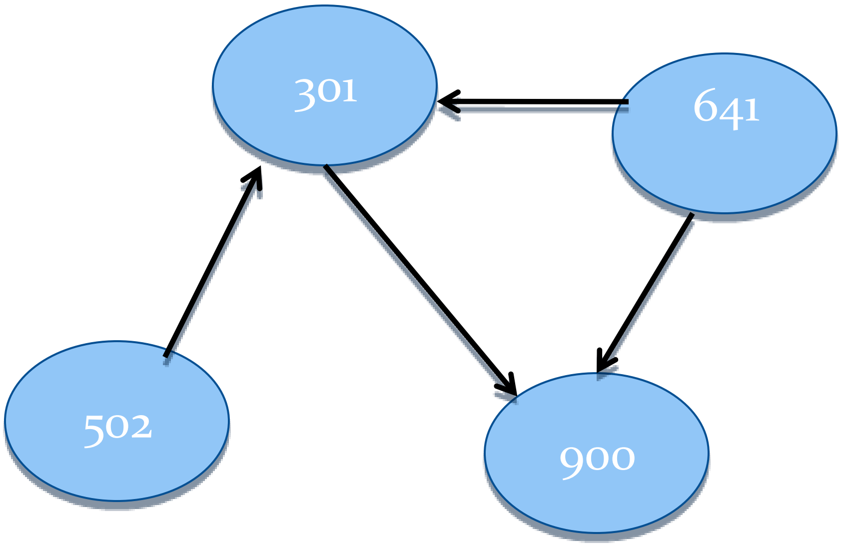
Data description

- Message level and thread-level data from forum
- Message characteristics
  - Posting time, EmployeeID, Thread, Type of message (query or response), content of message etc.
- User characteristics
  - EmployeeID, Tenure at firm, Age, Gender, Location, Division, Job Title

Network structure evolution

Sequence of Actions:

- *User 301 posts a query Q1000*
- *Users 502, 641 post responses*
- *User 900 posts a query Q1001*
- *Users 301, 641 post responses*



Directed Response Graph

Network structure

Asymmetric tie:

- A as responded to B’s query but B has not responded to A

Sole-symmetric tie:

- Users have responded to each other, but not as part of a clique

Simmelian Tie:

- Users are part of a ‘clique’, whose members have all responded to one another

Simmelian Ties

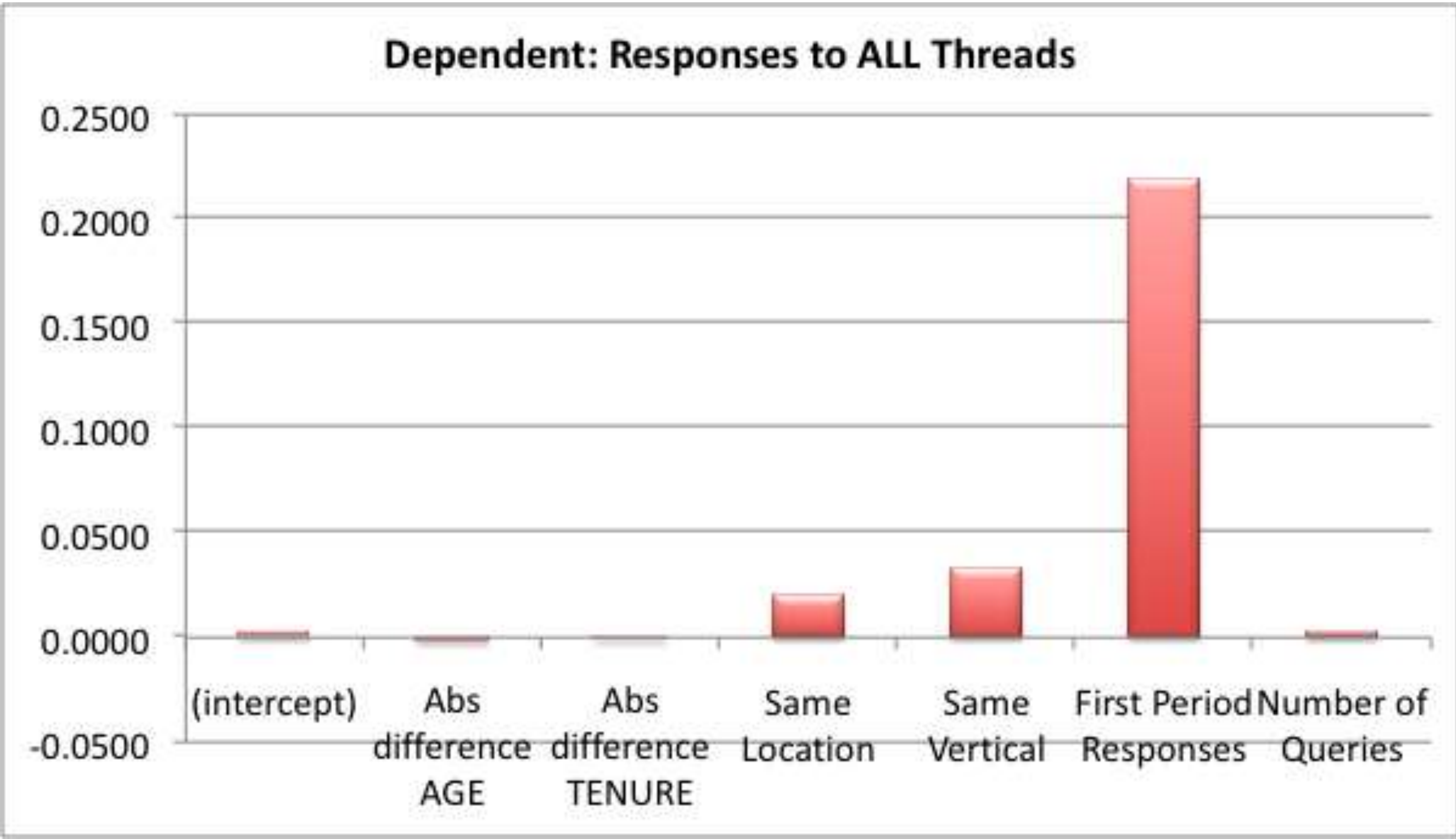
Research Questions

1. Can Simmelian ties be established in an electronic communications medium with repeated interactions? Will they matter?
2. Do these ties depend upon the context? Do more instrumental contexts result in weaker Simmelian ties or less effective Simmelian ties?
3. Do both current context (what type of query) or past context in which the tie was established matter?

Dyadic QAP Regression Results

Dependent variable:

Number of response by A to B in period two



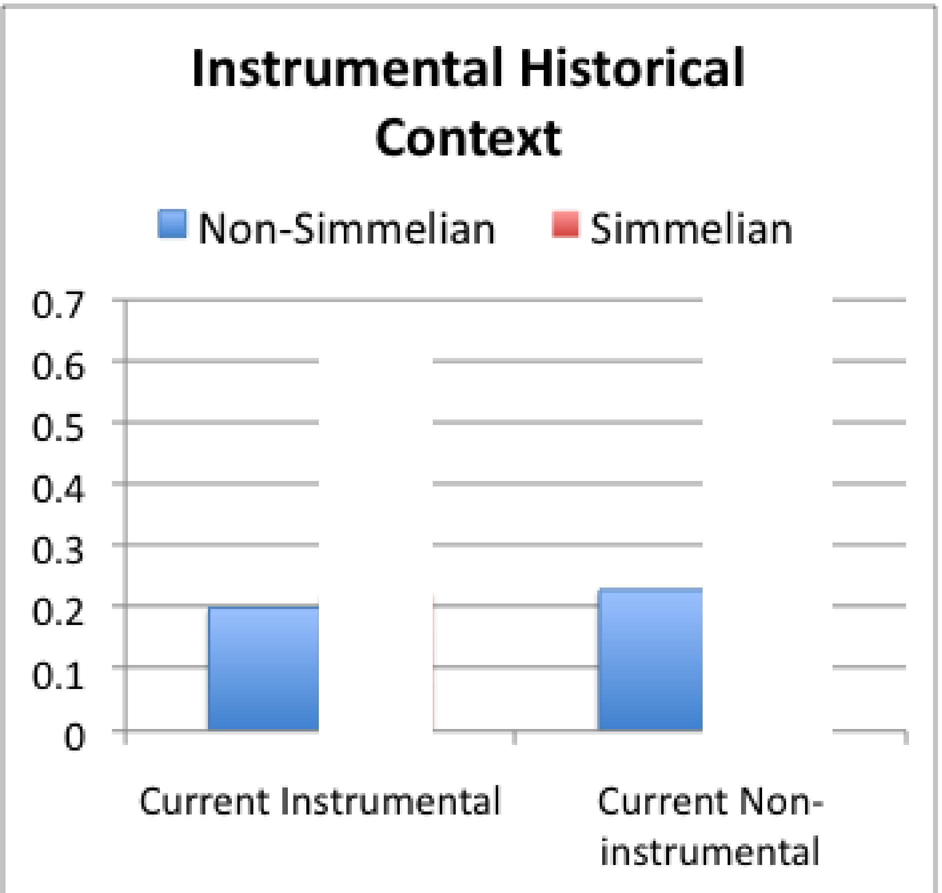
Dyadic QAP Regression Results

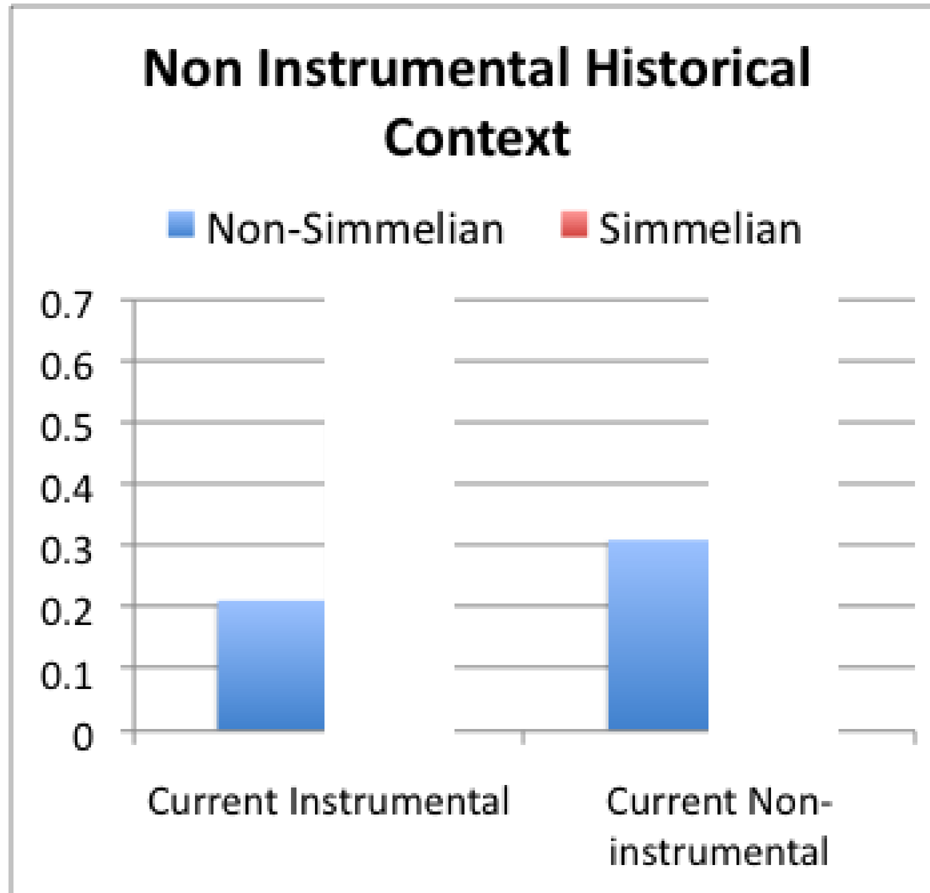
Dependent variable:

Number of response by A to B in period two

Explanatory Variables:

Dyadic Homophily Measures, **Structural Properties in period one**





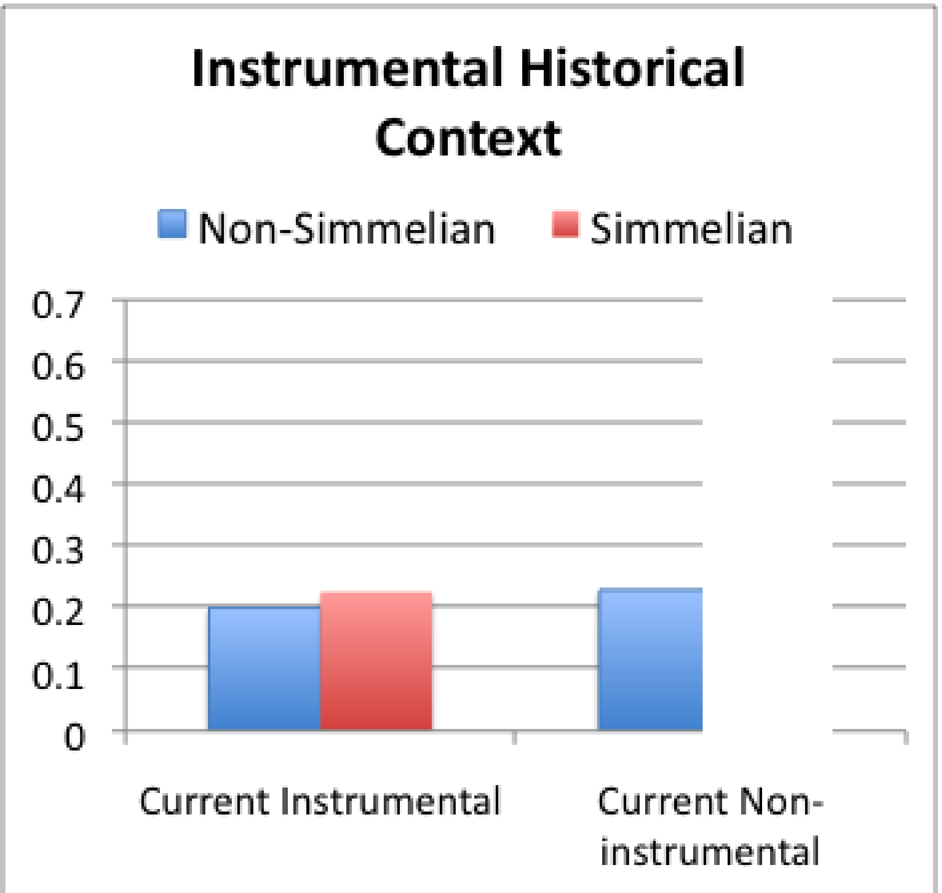
Dyadic QAP Regression Results

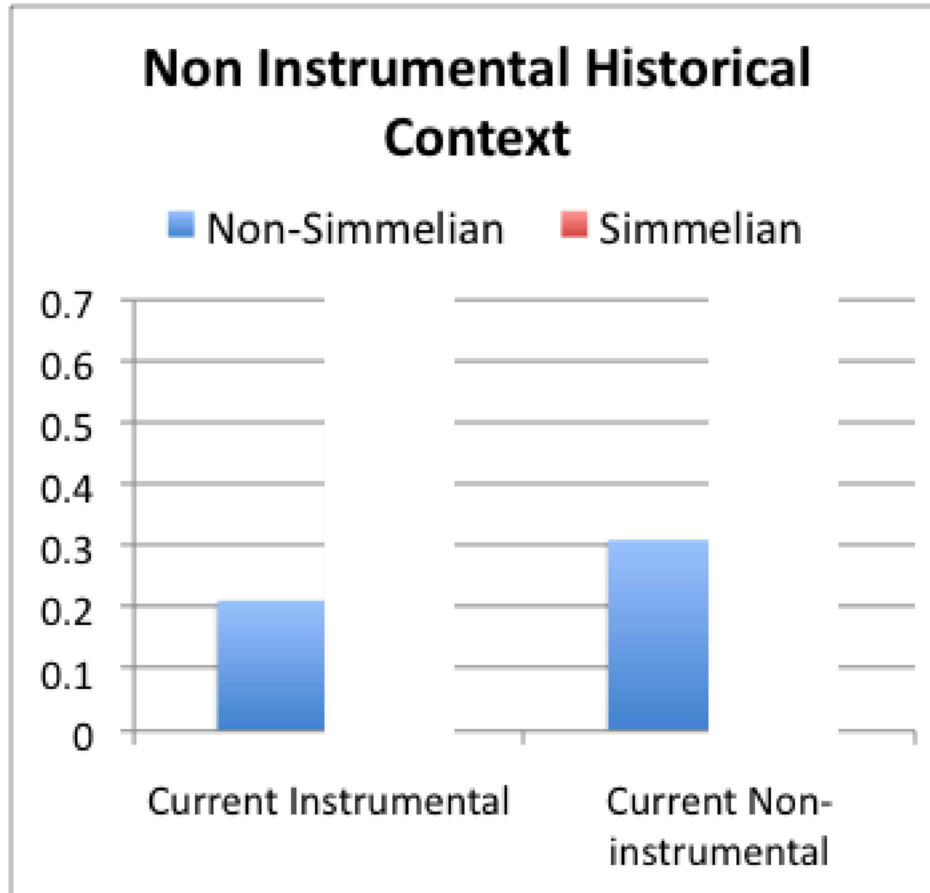
Dependent variable:

Number of response by A to B in period two

Explanatory Variables:

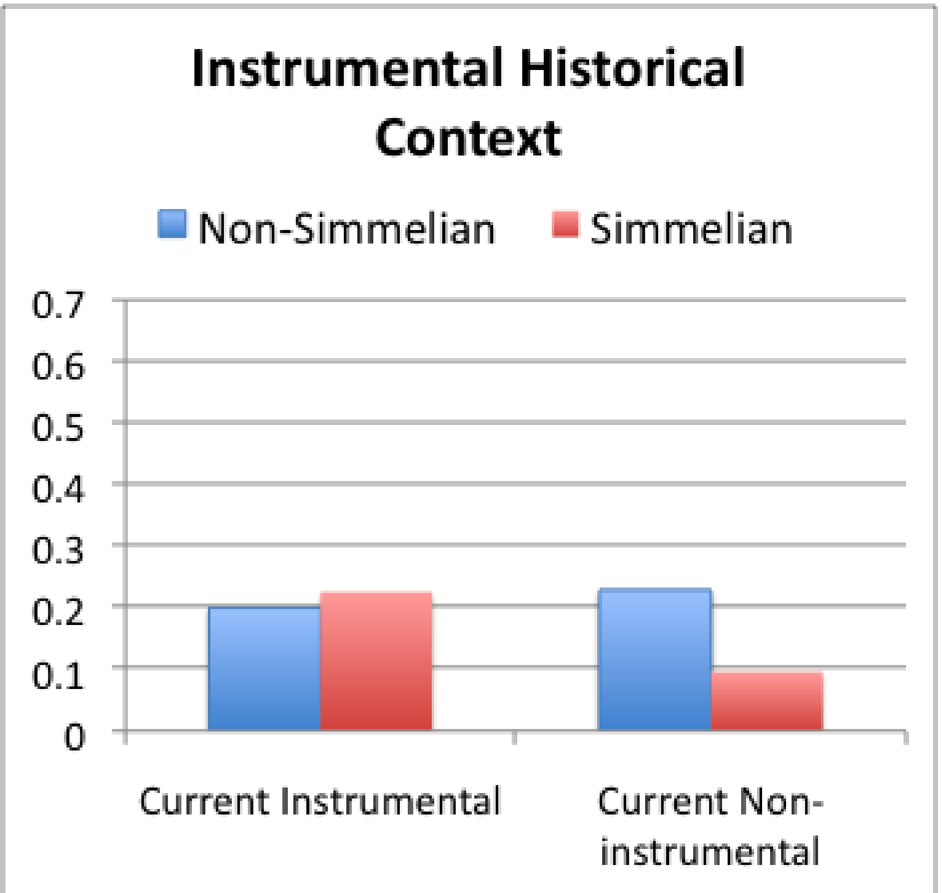
Dyadic Homophily Measures, **Structural Properties in period one**

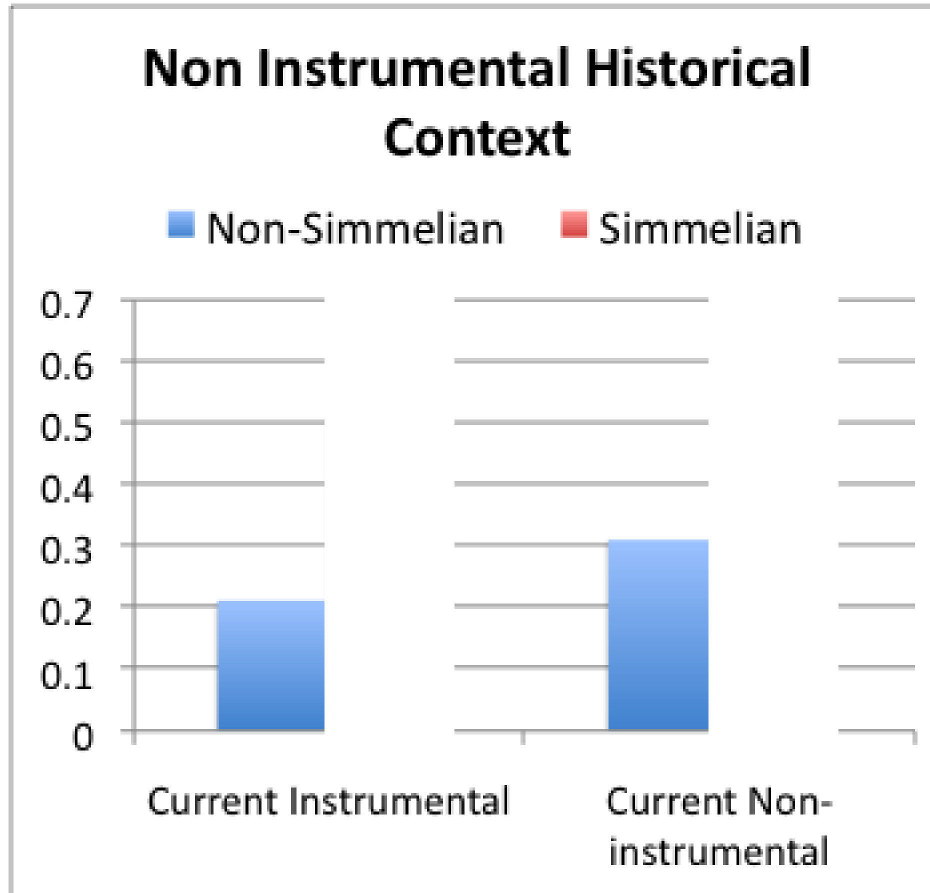




Dyadic QAP Regression Results

Dependent variable:
Number of response by A to B in period two
Explanatory Variables:
Dyadic Homophily Measures, **Structural Properties in period one**





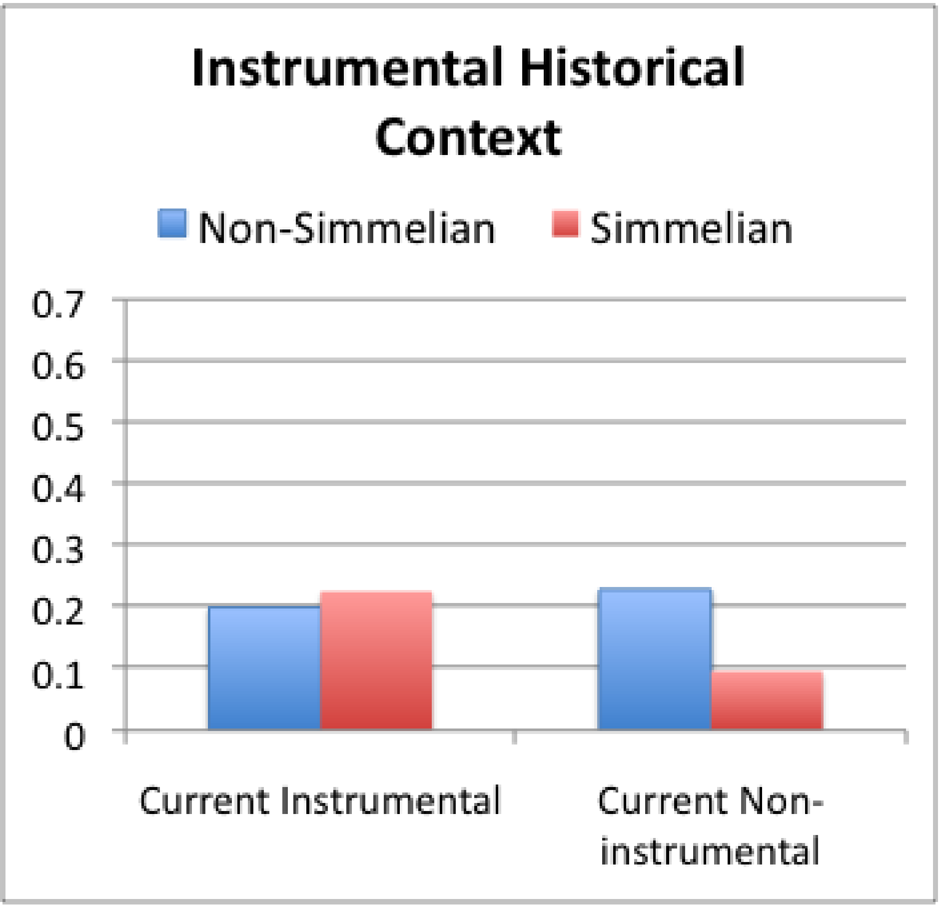
Dyadic QAP Regression Results

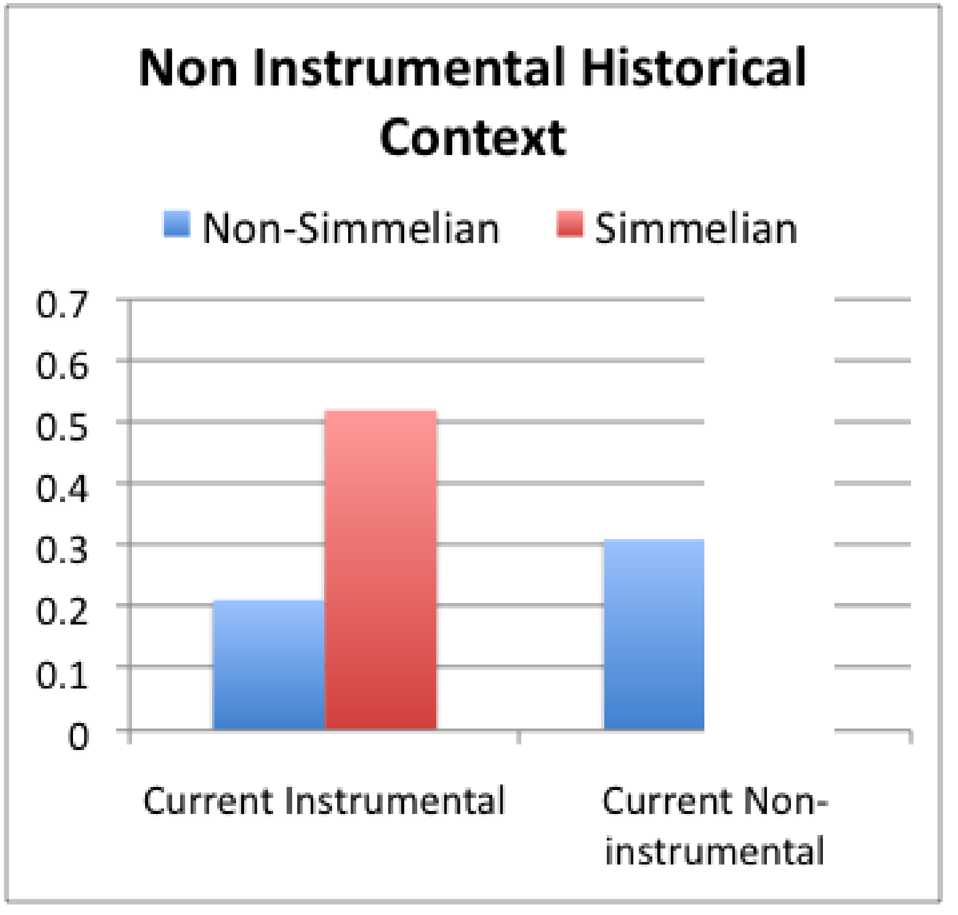
Dependent variable:

Number of response by A to B in period two

Explanatory Variables:

Dyadic Homophily Measures, **Structural Properties in period one**





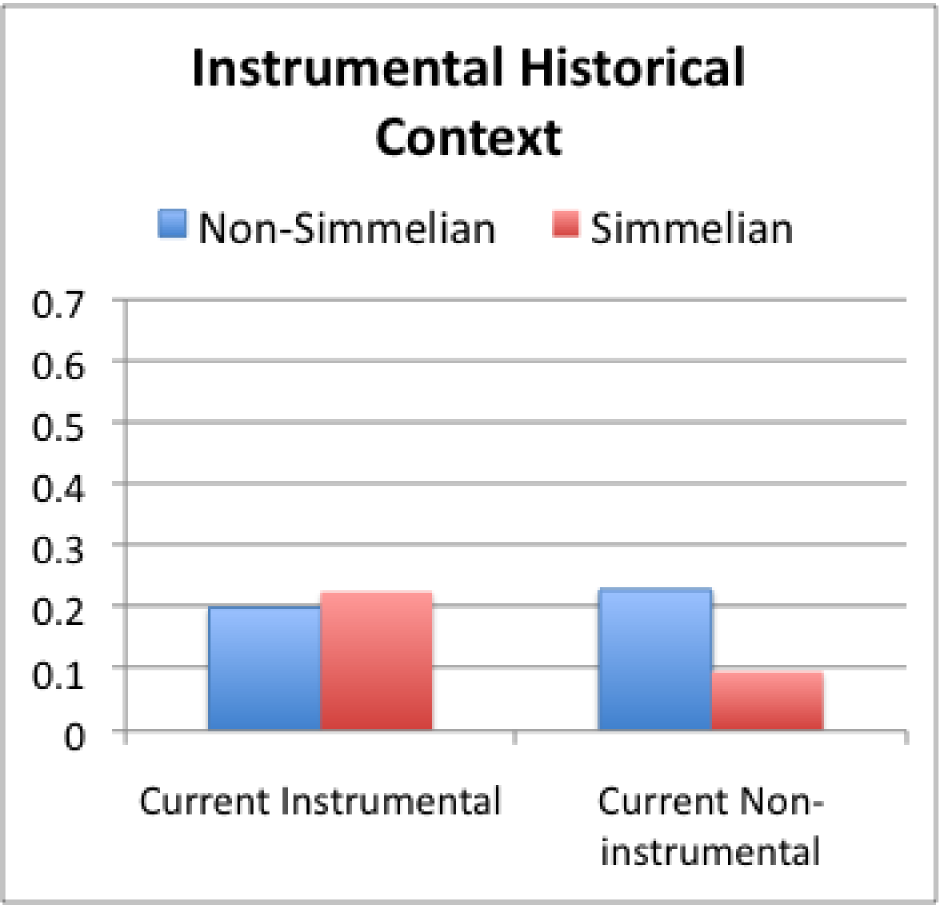
Dyadic QAP Regression Results

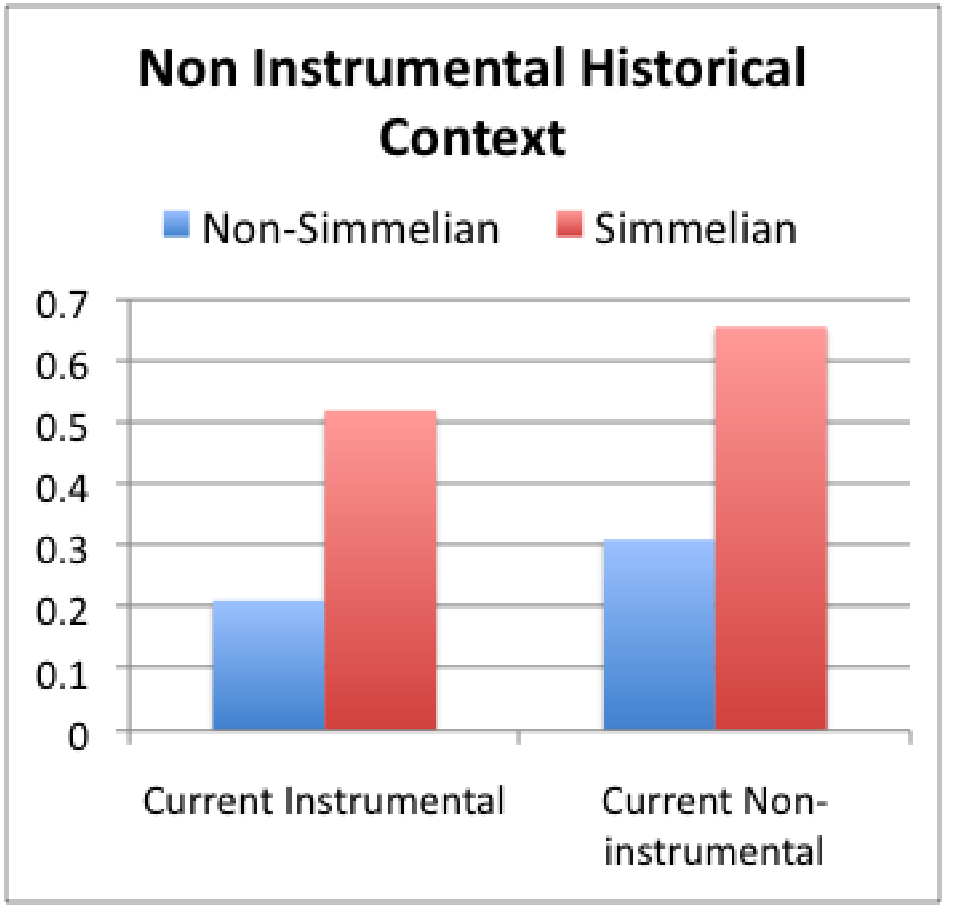
Dependent variable:

Number of response by A to B in period two

Explanatory Variables:

Dyadic Homophily Measures, **Structural Properties in period one**





Example 2: Social Influence vs. Homophily in product/service adoption

- Focus on identifying users that can help diffuse “information” over the network
- Learn about the power of “social influence” as trigger for the diffusion process
- Learn about how social influence is associated to “contagious churn”

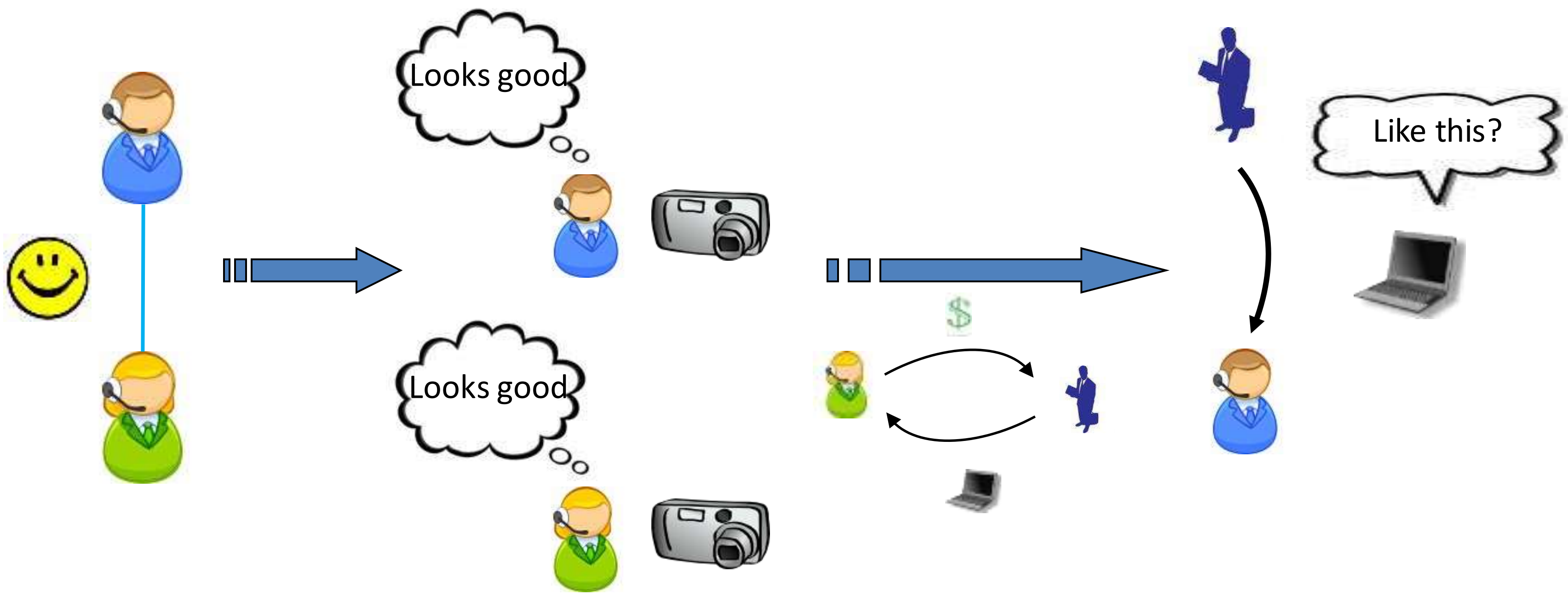
Research Question

- Can we predict consumers' product purchase decisions…

- Using social network information?

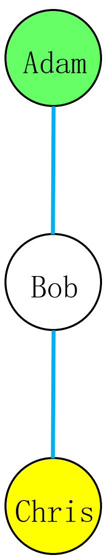
Theoretical Foundation

- ➢ Homophily (Mcpherson et al. 2001)
  - ➢ "Birds of a feather flock together"



The Challenge

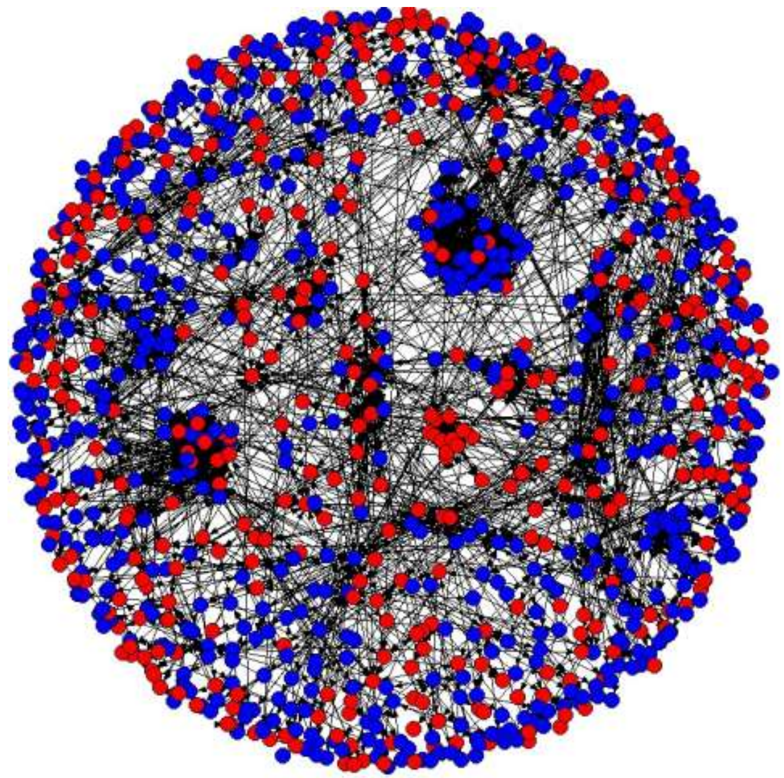
- Large-scale network



I like it

?

No, I don't



Literature

- A rich literature on networks from various fields (e.g. Kleinberg 1999, Brin and Page 1998)
- Network-based marketing
  - Network Neighbors: Hill, Provost, Volinsky (2006)
  - Viral Marketing: Richardson and Domingos (2002)
  - Classification: Macskassy and Provost (2003, 2007)
- What about *unobserved product taste*?
  - For small, tightly connected groups: Hartmann (2010)
  - But what about large-scale networks of arbitrary connection structure?

This Study

- Model correlated purchase behaviors of consumers in a large social network…
- Using Gaussian Markov Random Field (GMRF) to characterize latent product taste
  - Handle networks of arbitrary topology
  - Encapsulate conditional independence
- Estimation result confirms the positive taste correlation among connected people
- Predictive performance better than existing LR based models, and better than SVM based models, too.

Data

- Obtained from a large Asian telecom company
  - 231,416 customers
  - 6 month period
  - Detailed phone call data
    - Who called whom, when
  - Demographics information: gender, age
  - Purchase records of caller ringback tone (CRBT)
    - Who purchased what, when

- *Can we predict CRBT adoption decisions?*

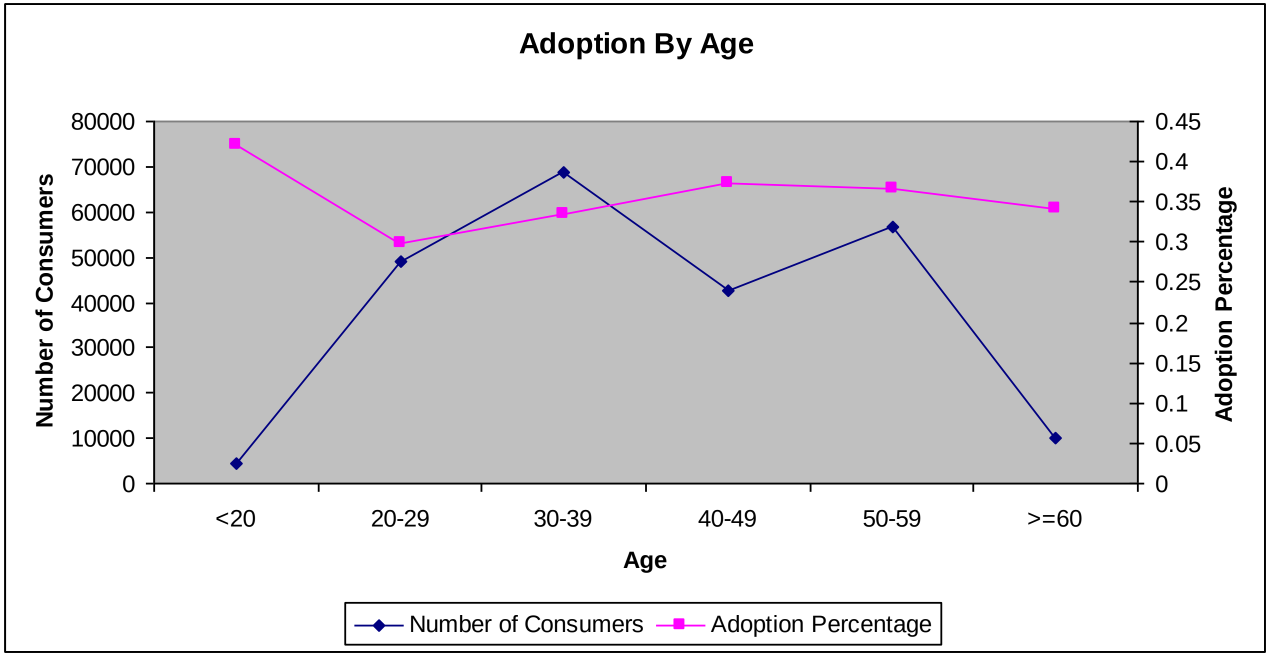
Descriptive Statistics

	Mean	SD	Min	Max
Gender	Male	218017	Female	13399
Age	40.56	13.67		
Number of Consumers Called by Each Consumer	13.73	22.9	1	2858
Number of Phone Calls Per Consumer	410.4	942.7	1	59016
	Number	Adoption Percentage		
Number of Consumers	231416			
Number of Consumers Who Adopted CRBT	79505	34.36%		
Adoption Percentage by Gender	Male	34.50%	Female	31.89%

Preliminary analysis: gender doesn't help much in prediction…

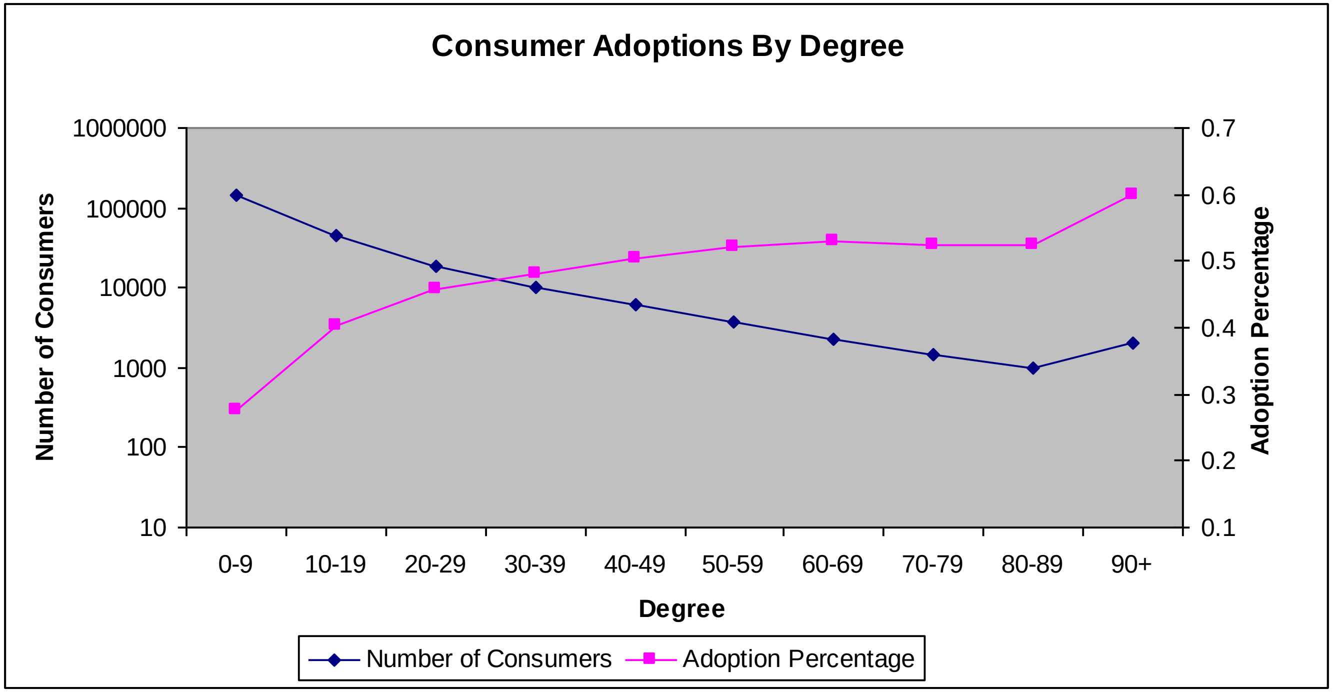
Data – Preliminary Analysis

Age doesn't help much, either…



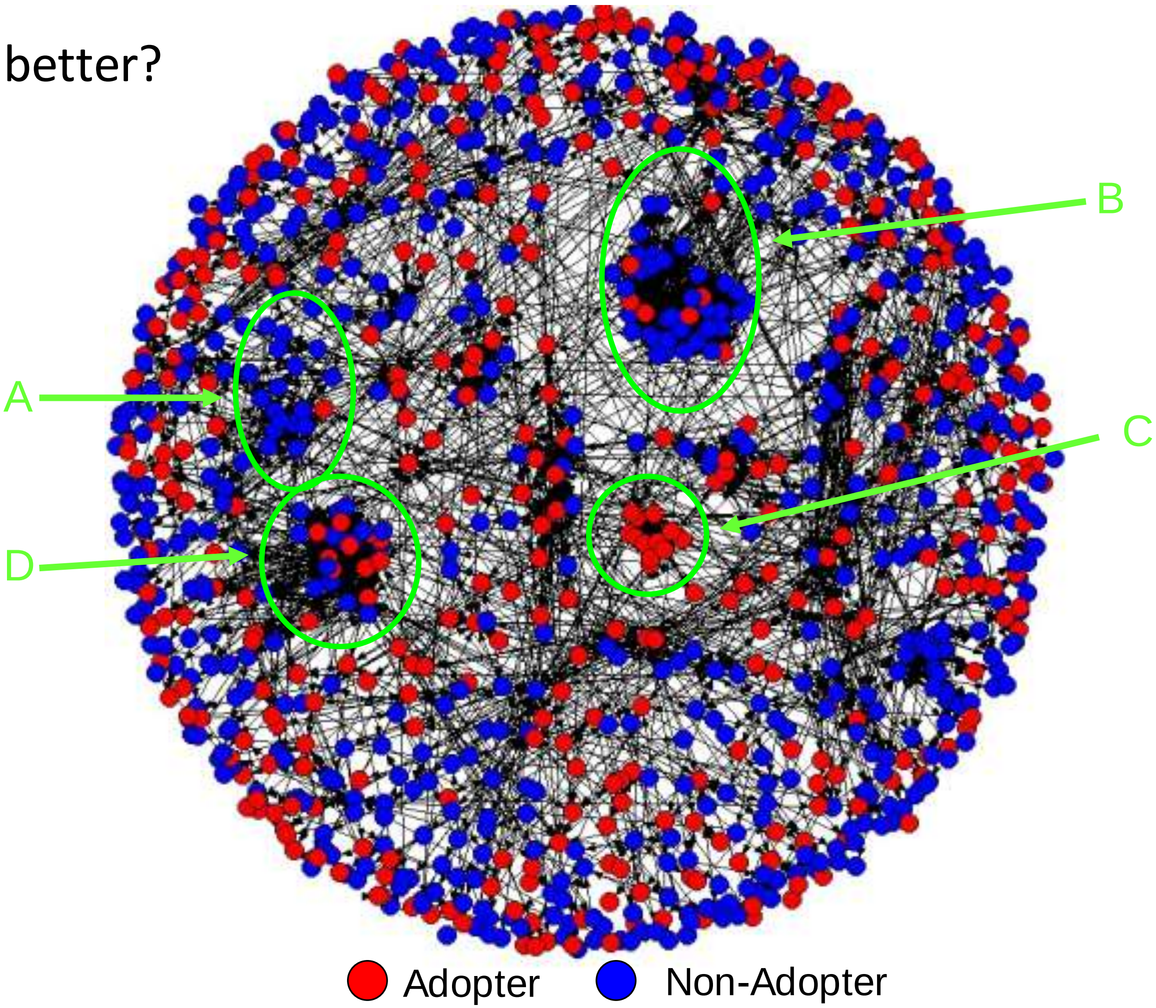
Data – Preliminary Analysis

Node degree helps a lot (need for social network)!



Data – Preliminary Analysis

Can we do better?



Maybe, but need the discipline of a model

Model

There are I consumers in a social network

Connection matrix: $C = [c_{ij}]$

$$c_{ij} = \begin{cases} 1 & \text{if consumers } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

Adoption decision: $D_i = \begin{cases} 1 & \text{if consumers } i \text{ adopts the product} \\ 0 & \text{otherwise} \end{cases}$

Adoption Probability

Binary Probit Model

$$\Pr(D_i = 1) = \Pr(U_i >= 0)$$

$$U_i = \alpha_i + \beta X_i + \varepsilon_i$$

$\varepsilon_i \sim N(0,1)$ Random disturbance

X_i Observed individual characteristic (gender, age, connection degree)

α_i Unobserved product taste

Modeled as a GMRF!

Gaussian Markov Random Field (GMRF)

Definition (GMRF): A random vector $\vec{x} = (x_1, \ldots x_n)^T$ is called GMRF w.r.t. the undirected graph $G = (V = \{1..n\}, E)$ with mean $\vec{\mu}$ and precision matrix $Q > 0$ if and only if its density has the form:

$$\pi(\vec{x}) = (2\pi)^{-n/2} \, | Q |^{1/2} \, \exp(-\frac{1}{2}(\vec{x} - \vec{\mu})^T Q(\vec{x} - \vec{\mu}))$$

And

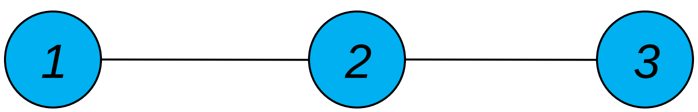
$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E, \forall i, j$$

- A multivariate normal vector
- Connection structure encoded in its precision matrix
- Non-zero off-diagonal elements correspond to connections

Properties of GMRF

- Can model connections of arbitrary topology
  - Better than using in-group correlation
- Encodes conditional independence

$$x_i \perp x_j \mid x_{-ij} \Leftrightarrow Q_{ij} = 0, \forall i, j$$

e.g. 

Consumers 1 and 3 should be correlated
But conditional on consumer 2, they should be independent

- Model parameters have intuitive explanations

Model Latent Product Taste Using GMRF

$$\begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_I \end{pmatrix} \sim N\left(\begin{pmatrix} \overline{\alpha} \\ \dots \\ \overline{\alpha} \end{pmatrix}, Q^{-1}\right) \qquad Q = [q_{ij}], \text{where } q_{ij} = 0 \text{ if } c_{ij} = 0$$

Straightforward Interpretation : $\text{Precision}(\alpha_i \mid \alpha_{-i}) = q_{ii}$

$$\text{Cor}(\alpha_i, \alpha_j \mid \alpha_{-ij}) = -q_{ij} / \sqrt{q_{ii} q_{jj}}$$

Parameterization (base model, **model B**):

$$Q = \begin{pmatrix} \kappa & -r\kappa & 0 & \dots & -r\kappa \\ -r\kappa & \kappa & 0 & \dots & 0 \\ 0 & 0 & \kappa & \dots & -r\kappa \\ \dots & \dots & \dots & \ddots & \dots \\ -r\kappa & 0 & -r\kappa & \dots & \kappa \end{pmatrix}$$

r Conditional correlation between connected consumers

κ Conditional precision

Model Extension

Model AI:

$$Q^{I} = \begin{pmatrix} \kappa_{d_1} & -r\sqrt{\kappa_{d_1}\kappa_{d_2}} & 0 & \cdots & -r\sqrt{\kappa_{d_1}\kappa_{d_I}} \\ -r\sqrt{\kappa_{d_1}\kappa_{d_2}} & \kappa_{d_2} & 0 & \cdots & 0 \\ 0 & 0 & \kappa_{d_3} & \cdots & -r\sqrt{\kappa_{d_3}\kappa_{d_I}} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ -r\sqrt{\kappa_{d_1}\kappa_{d_I}} & 0 & -r\sqrt{\kappa_{d_3}\kappa_{d_I}} & \cdots & \kappa_{d_I} \end{pmatrix}$$

$$\kappa_d = \kappa_0 + \kappa_1 \cdot \log(d+1)$$

The more we know about a consumer's connections, the more we should know about the consumer

Model AII:

$$Q^{II} = \begin{pmatrix} \kappa_{d_1} & -r_{21}\sqrt{\kappa_{d_1}\kappa_{d_2}} & 0 & \cdots & -r_{I1}\sqrt{\kappa_{d_1}\kappa_{d_I}} \\ -r_{21}\sqrt{\kappa_{d_1}\kappa_{d_2}} & \kappa_{d_2} & 0 & \cdots & 0 \\ 0 & 0 & \kappa_{d_3} & \cdots & -r_{I3}\sqrt{\kappa_{d_3}\kappa_{d_I}} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ -r_{I1}\sqrt{\kappa_{d_1}\kappa_{d_I}} & 0 & -r_{I3}\sqrt{\kappa_{d_3}\kappa_{d_I}} & \cdots & \kappa_{d_I} \end{pmatrix}$$

$$r_{ij} = r_0 + r_1 \cdot \log(Call_{ij})$$

The more communication between two consumers, the stronger the tie should be, and the stronger the correlation

Estimation

- Hierarchical Bayesian approach
- MCMC draws with hybrid Metropolis-Gibbs fashion

$$f(\alpha_i \mid \alpha_{-i}, \beta, \overline{\alpha}, r, \kappa, X_i, D_i, C) \propto \varphi(\alpha_i \mid \alpha_{N(i)}, \overline{\alpha}, r, \kappa) L(D_i \mid \alpha_i, \beta, X_i, D_i)$$

$$f(\overline{\alpha} \mid \alpha_i : i = 1..I) \propto \phi((I + V_\alpha)^{-1}(\sum_{i=1}^{I} \alpha_i + V_\alpha \overline{\overline{\alpha}}), (I + V_\alpha)^{-1})$$

$$f(\beta \mid \alpha_i : i = 1..I, X_i, D_i) \propto \pi(\beta) \prod_{i=1}^{I} L(D_i \mid \alpha_i, \beta, X_i, D_i)$$

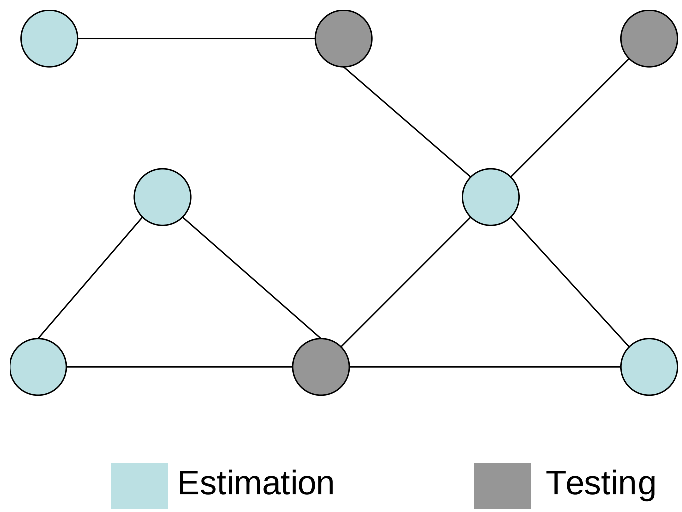
$$f(r \mid \alpha_i : i = 1..I, \beta, \overline{\alpha}, \kappa, C) \propto \pi(r) \prod_{i=1}^{I} \varphi(\alpha_i \mid \alpha_{N(i)}, \overline{\alpha}, r, \kappa)$$

$$f(\kappa \mid \alpha_i : i = 1..I, \beta, \overline{\alpha}, r, C) \propto \pi(\kappa) \prod_{i=1}^{I} \varphi(\alpha_i \mid \alpha_{N(i)}, \overline{\alpha}, r, \kappa)$$

Identifying Connections

- Based on phone call data
- Using a “threshold” method: two consumers are considered as connected if they made at least a certain number of phone calls
- Endogenizing network formation left for future extension
- Vary threshold value to ensure robustness

Dividing Training and Testing Data



- 80% of consumers for training, 20% for testing
- Each node (consumer) is individually randomly assigned ("flip-a-coin") to training or testing set.
- The sub-network consisting of training nodes is used for estimation
- Other division methods possible, for future extension
- Vary training dataset size for robustness check

Result: Parameter Estimation

Model B

Threshold	κ		r	
	Mean	SD	Mean	SD
1	0.0991	0.00036	0.0225	0.00012
3	0.0978	0.00064	0.0303	0.0004
5	0.0964	0.00044	0.0385	0.00072
8	0.0951	0.00059	0.0464	0.00075
10	0.0952	0.00074	0.0471	0.00088
20	0.0934	0.00051	0.0595	0.00104

Positive conditional correlation

Statistically significant

- The higher the threshold value, the higher the correlation
- Higher threshold filter out more "noise"

Result: Parameter Estimation

Model AI

Threshold	κ_0		κ_1		r	
	Mean	SD	Mean	SD	Mean	SD
1	0.129	0.0011	-0.013	0.00031	0.0227	0.00038
3	0.115	0.00093	-0.0097	0.00037	0.03487	0.0006
5	0.113	0.00153	-0.0094	0.00061	0.03912	0.00079
8	0.108	0.0011	-0.008	0.00075	0.0469	0.00088
10	0.1043	0.0015	-0.0063	0.00084	0.0536	0.00094
20	0.101	0.0016	-0.0054	0.00091	0.0607	0.0012

- Conditional precision is lower for nodes with higher degree
- Possibly explained by heterogeneity

Result: Parameter Estimation

Model All

Threshold	κ_0		κ_1		r_0		r_1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.129	0.0011	-0.0127	0.0004	-0.0013	0.000832	0.0128	0.0004
3	0.117	0.0008	-0.0099	0.0004	-0.021	0.0022	0.0183	0.0007
5	0.11	0.0012	-0.0078	0.0006	-0.025	0.0034	0.0199	0.001
8	0.1077	0.0016	-0.0074	0.0008	-0.0476	0.0036	0.0253	0.0009
10	0.1051	0.0011	-0.0063	0.0006	-0.0444	0.0047	0.0242	0.0012
20	0.0994	0.0014	-0.004	0.00087	-0.056	0.0061	0.0283	0.0014

- The more frequently the communication, the higher the conditional correlation!

- Not all connections are the same; strength matters.

Predictive Performance

- Prediction Approach:
  - "Individual-based": predict adoption when calculated probability is 0.5 or higher.
  - "Top-k": predict adoption for the k consumers with the highest calculated probabilities.

- Evaluation Approach:
  - Accuracy: percentage of correct predictions
  - Precision: percentage of correct predictions when the prediction is to adopt

Benchmark Models

Model	Explanatory Variables	Mechanism
BM1	Gender, Age	Logistic Regression
BM2	Gender, Age, Degree	Logistic Regression
BM3	Gender, Age, Degree, Percentage of Neighbors who Adopt	Logistic Regression
BM4	Gender, Age, Degree, Percentage of Neighbors who Adopt	Suppor Vector Machine, Linear Kernel
BM5	Gender, Age, Degree, Percentage of Neighbors who Adopt	Suppor Vector Machine, Polynomial Kernel

Accuracy – Individual Based

				Percent of Correct Prediction			
Threshold	Total Test Cases	Total Adoption	Adoption Percent	Mode B	Model AI	Model AII	"Naive" Model
1	46092	15752	34.18%	66.82%	66.71%	67.14%	65.82%
3	42675	15205	35.63%	65.93%	66.10%	66.52%	64.37%
5	39575	14234	35.97%	65.35%	65.24%	66.06%	64.03%
8	36715	13674	37.24%	64.52%	64.97%	65.49%	62.76%
10	35290	13103	37.13%	64.38%	63.84%	64.79%	62.87%
20	29846	11520	38.60%	63.11%	63.20%	63.74%	61.40%

- Better than naïve model (not by much)
- Higher threshold leads to lower accuracy
- But that's because "the problem gets harder"

Precision – Individual Based

	Model B		Model AI		Model AII	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	8385	52.88%	7671	52.76%	8129	53.72%
3	5658	55.07%	6439	55.71%	6752	56.80%
5	6609	54.18%	6359	55.56%	6672	56.01%
8	6707	54.96%	6333	55.35%	6700	57.48%
10	6182	55.26%	7344	54.10%	6242	55.43%
20	6213	54.45%	5977	55.19%	6693	55.22%

- Much better than naïve model
- Model AII is the best
- Performance best at medium threshold
- Balance between filtering out noise and retaining information

Benchmark Precision – Individual Based

	Model BM2		Model BM3	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	2006	56.23%	2089	59.89%
3	2060	54.13%	2226	57.77%
5	4142	56.78%	1951	58.89%
8	5475	55.87%	2015	60.10%
10	7124	52.91%	2176	59.93%
20	10939	48.43%	2289	62.69%

Slightly higher precision

On much fewer predictions!

	Model B		Model AI		Model AII	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	8385	52.88%	7671	52.76%	8129	53.72%
3	5658	55.07%	6439	55.71%	6752	56.80%
5	6609	54.18%	6359	55.56%	6672	56.01%
8	6707	54.96%	6333	55.35%	6700	57.48%
10	6182	55.26%	7344	54.10%	6242	55.43%
20	6213	54.45%	5977	55.19%	6693	55.22%

Benchmark Precision – Individual Based

	Model BM4		Model BM5	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	3470	62.07%	1654	68.50%
3	3718	61.97%	1946	65.83%
5	3371	62.06%	2529	64.41%
8	4383	62.03%	2977	65.10%
10	4712	60.36%	3474	63.27%
20	4688	60.30%	3403	62.83%

Same story here

	Model B		Model AI		Model AII	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	8385	52.88%	7671	52.76%	8129	53.72%
3	5658	55.07%	6439	55.71%	6752	56.80%
5	6609	54.18%	6359	55.56%	6672	56.01%
8	6707	54.96%	6333	55.35%	6700	57.48%
10	6182	55.26%	7344	54.10%	6242	55.43%
20	6213	54.45%	5977	55.19%	6693	55.22%

Precision – Top-K

	Model B		Model AI		Model AII	
Threshold	Top 1000	Top 2000	Top 1000	Top 2000	Top 1000	Top 2000
1	66.00%	65.80%	65.90%	62.25%	66.30%	65.35%
3	69.80%	64.60%	68.60%	64.90%	72.00%	68.00%
5	69.80%	67.00%	69.60%	65.10%	73.10%	68.75%
8	71.10%	67.05%	67.50%	64.65%	73.80%	68.55%
10	71.40%	65.55%	68.70%	65.25%	71.70%	67.40%
20	70.50%	66.40%	73.50%	66.90%	72.40%	67.10%

- Much higher precision than individual-based predictions
- Model AII is still the best
- Almost twice the accuracy of a naïve model
- Performance again the best for medium threshold values

Benchmark Precision – Top-K

	Model BM1		Model BM2		Model BM3	
Threshold	Top 1000	Top 2000	Top 1000	Top 2000	Top 1000	Top 2000
1	34.20%	34.05%	59.60%	56.25%	62.20%	60.25%
3	36.10%	35.90%	55.70%	53.90%	60.50%	57.90%
5	35.80%	35.80%	54.50%	52.45%	61.50%	59.00%
8	35.70%	37.75%	55.50%	53.90%	61.40%	60.00%
10	36.00%	38.70%	54.10%	53.25%	60.50%	59.45%
20	36.80%	38.15%	54.90%	52.15%	63.60%	62.85%

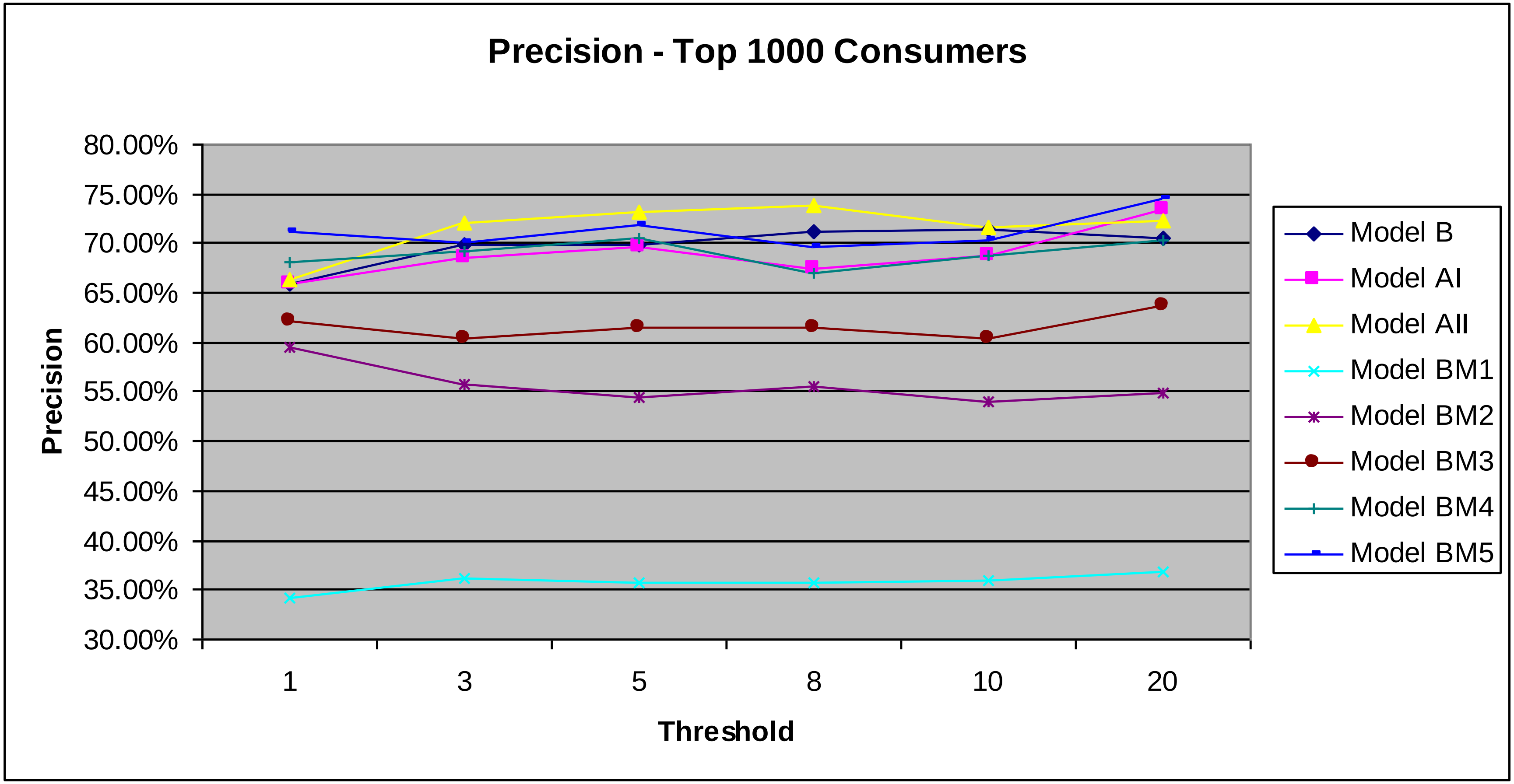
- Logistic-regression based models not nearly as good

Benchmark Precision – Top-K

	Model BM4		Model BM5	
Threshold	Top 1000	Top 2000	Top 1000	Top 2000
1	68.10%	66.25%	71.10%	67.05%
3	69.30%	65.25%	70.10%	65.90%
5	70.50%	65.70%	71.80%	66.70%
8	67.10%	66.80%	69.70%	67.50%
10	68.80%	65.60%	70.40%	66.80%
20	70.30%	68.25%	74.60%	67.40%

- SVM-based models almost as good, but still lower

In Pictures…


Precision - Top 1000 Consumers
Precision
80.00%
75.00%
70.00%
65.00%
60.00%
55.00%
50.00%
45.00%
40.00%
35.00%
30.00%
1
3
5
8
10
20
Threshold
Model B
Model AI
Model AII
Model BM1
Model BM2
Model BM3
Model BM4
Model BM5

Varying Training Dataset Size

	Model AII			Model BM5		
TrainingPortion	Individual	Top 1000	Top 2000	Individual	Top 1000	Top 2000
90%	56.85%	69.40%	62.20%	64.55%	66.10%	61.55%
80%	56.17%	71.60%	68.05%	66.11%	73.70%	67.55%
70%	55.30%	73.10%	69.25%	65.03%	72.10%	68.60%
60%	54.83%	74.90%	70.30%	63.46%	71.80%	68.55%
50%	53.86%	74.60%	71.85%	63.14%	73.90%	69.55%
40%	54.32%	76.50%	73.80%	61.31%	74.20%	70.90%
30%	53.64%	73.60%	69.75%	61.74%	74.40%	70.35%
20%	52.86%	72.30%	69.70%	61.92%	72.80%	69.25%
10%	52.74%	69.70%	68.40%	56.17%	69.30%	64.80%

- Result and comparison both stable
- Precision has an “inverted-U” shape w.r.t. training data size
- Fewer good candidates when test dataset is smaller

Future Extensions

- Dynamic Model
  - Repeat purchase decisions
  - Product choice decisions

- Incorporate Influence
  - We have communication data!

- Endogenize network formation

Key take aways

- Modeling the correlation of latent product tastes
  - In a large-scale social network
  - Using Gaussian Markov Random Field (GMRF)

- Estimation confirms positive correlation among connected consumers
  - We have communication data! Higher correlation for stronger ties

- Predictive precision better than logistic regression based and SVM based benchmark models

SMU SINGAPORE MANAGEMENT UNIVERSITY

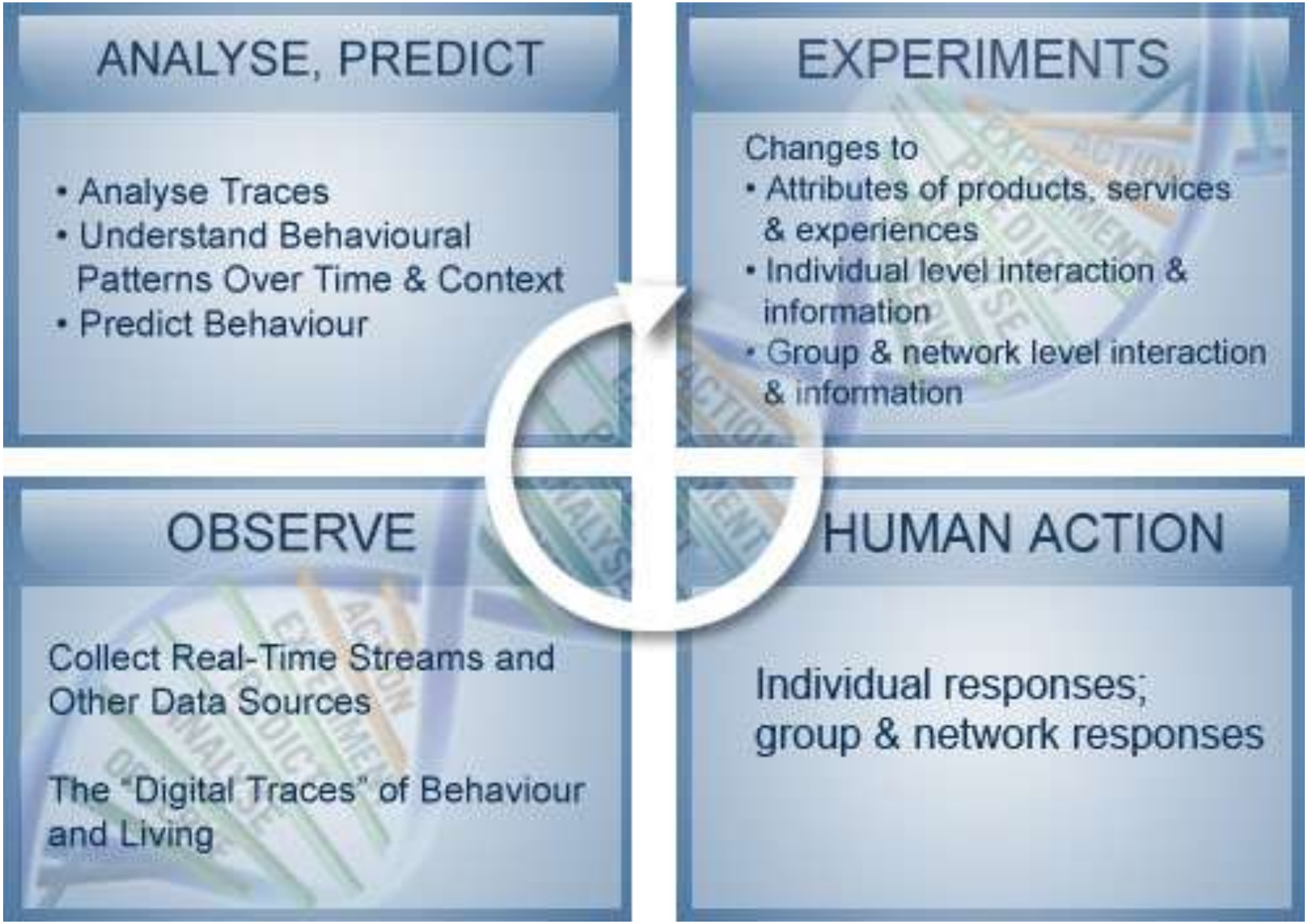
Living Analytics Research Centre

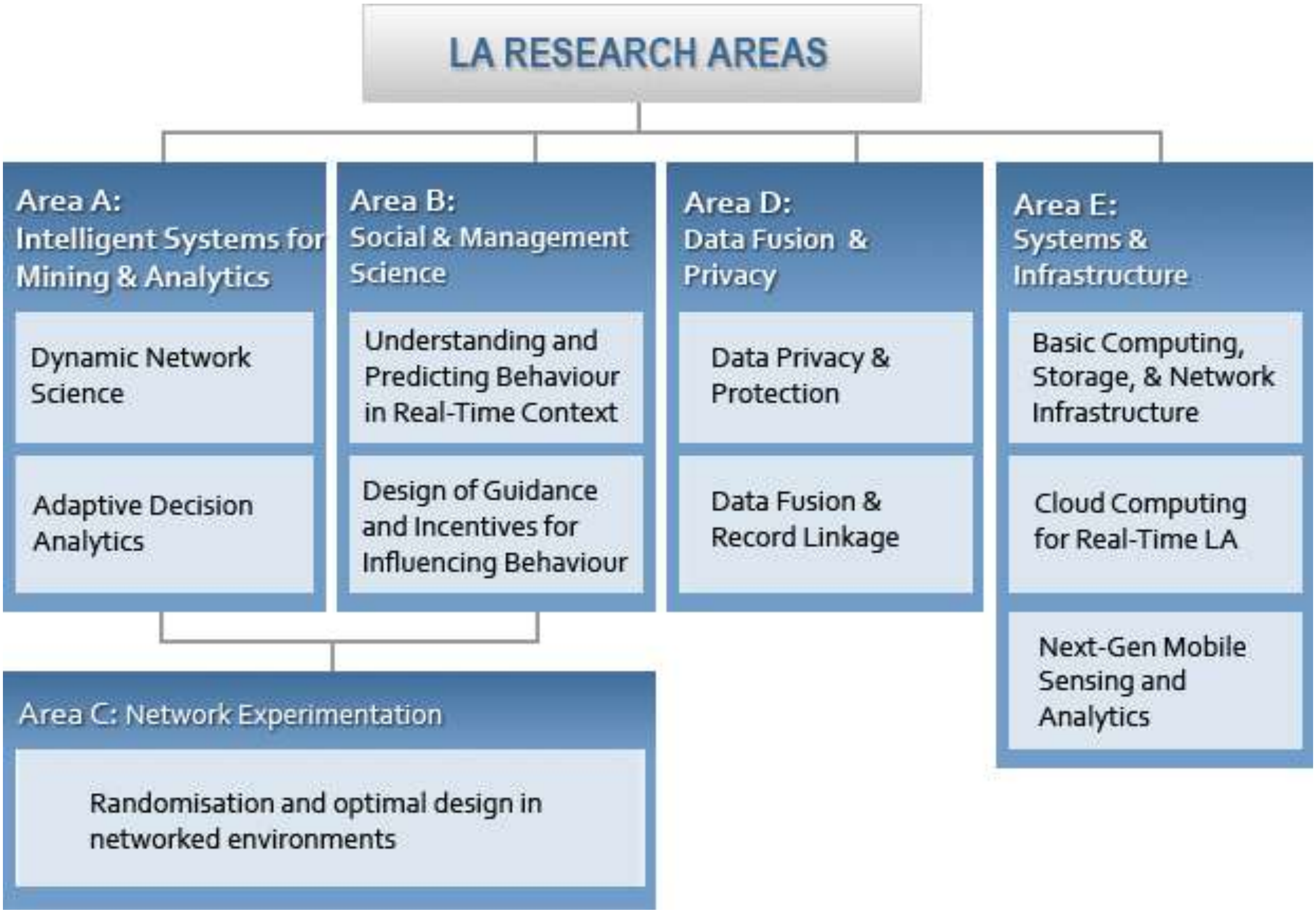
Carnegie Mellon University

Home | About Us | Research | PhD Study | News | Events | Career | Industry Partners

People

Research


ANALYSE, PREDICT
• Analyse Traces
• Understand Behavioural Patterns Over Time & Context
• Predict Behaviour
EXPERIMENTS
Changes to
• Attributes of products, services & experiences
• Individual level interaction & information
• Group & network level interaction & information
OBSERVE
Collect Real-Time Streams and Other Data Sources
The "Digital Traces" of Behaviour and Living
HUMAN ACTION
Individual responses; group & network responses


LA RESEARCH AREAS
Area A:
Intelligent Systems for Mining & Analytics
Dynamic Network Science
Adaptive Decision Analytics
Area B:
Social & Management Science
Understanding and Predicting Behaviour in Real-Time Context
Design of Guidance and Incentives for Influencing Behaviour
Area D:
Data Fusion & Privacy
Data Privacy & Protection
Data Fusion & Record Linkage
Area E:
Systems & Infrastructure
Basic Computing, Storage, & Network Infrastructure
Cloud Computing for Real-Time LA
Next-Gen Mobile Sensing and Analytics
Area C: Network Experimentation
Randomisation and optimal design in networked environments

Experiments with network data

- Statistical theory of design of experiments assumes independence between test and control
- This independence is violated in network settings since observations are affected by network interaction and influences
- This is work to be done and one of the key areas of focus of the Living Analytics Center