Social bridges in urban purchase behavior

Xiaowen Dong

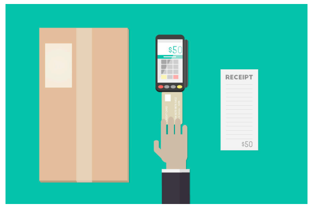
MIT Media Lab

with Yoshihiko Suhara, Burçin Bozkaya, Vivek K. Singh,
Bruno Lepri and Alex 'Sandy' Pentland

Cambridge, MA, June 2017

Introduction

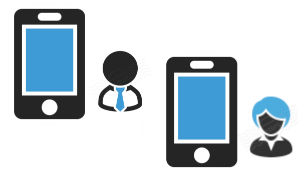
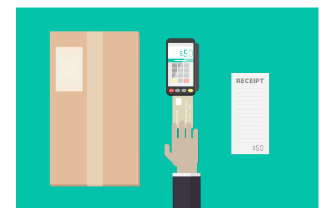
New data sources about human behavior are emerging

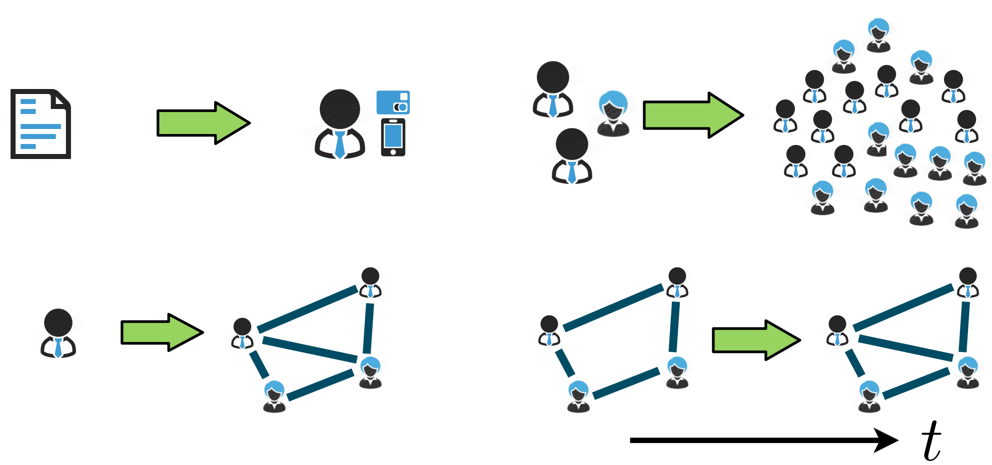
Introduction

New data sources about human behavior are emerging



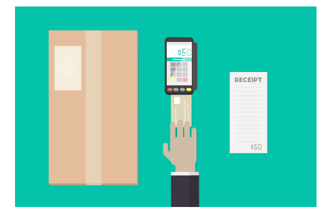
Computational social science (CSS):
A paradigm shift in social science



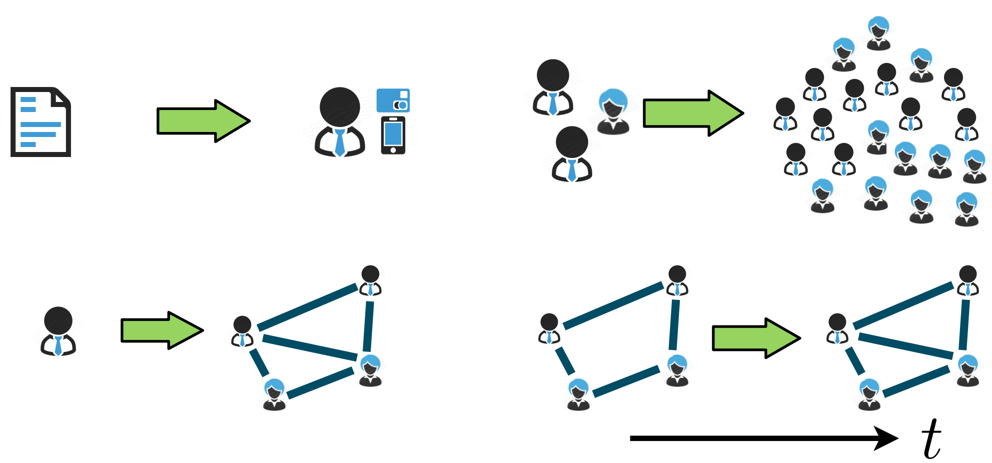
Introduction

New data sources about human behavior are emerging



Computational social science (CSS):
A paradigm shift in social science



Practical impact

Current population management:

- demographics
- individual records
- static information

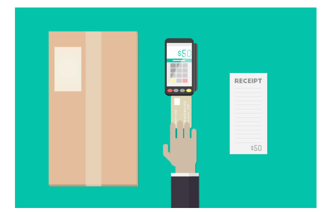
The new way:

- behavioral traits
- collective behavior
- dynamics

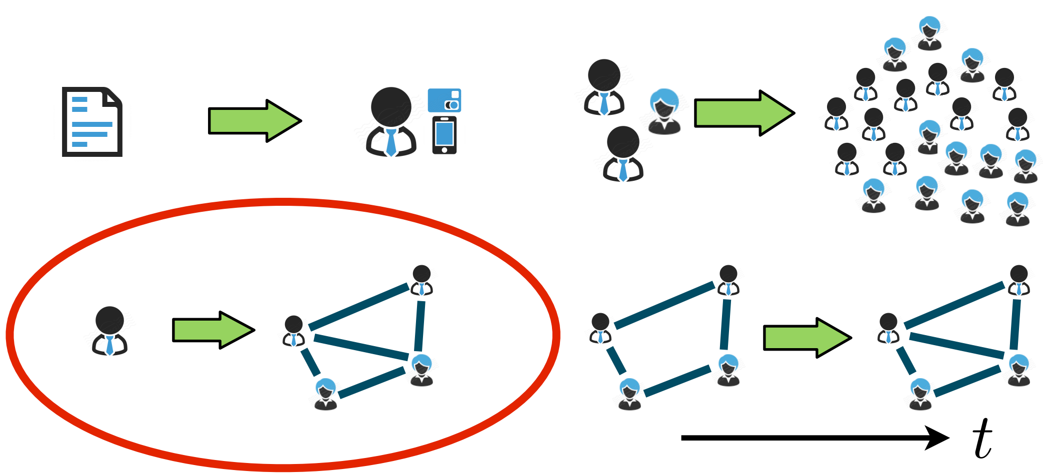
Introduction

New data sources about human behavior are emerging



Computational social science (CSS):
A paradigm shift in social science



Practical impact

Current population management:

- demographics
- individual records
- static information

The new way:

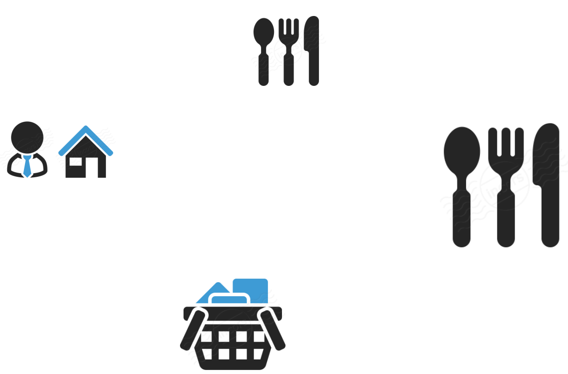
- behavioral traits
- collective behavior
- dynamics

How communication affects human decision-making?

Introduction

Introduction

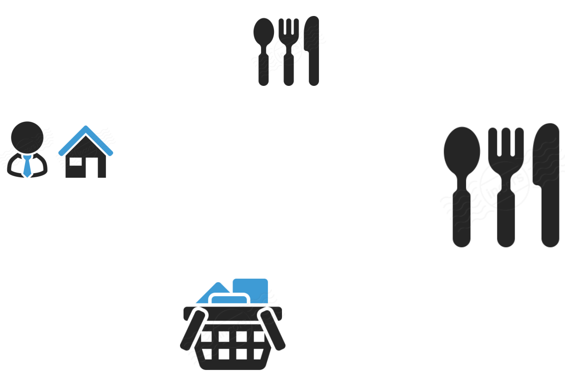
- Classical purchase behavior models treat individual purchases separately (Huff, 1964)



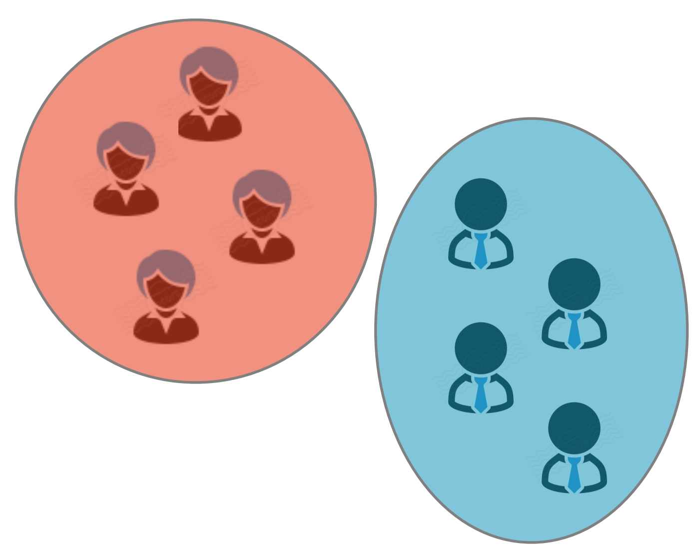
$$p_i \propto \frac{\mathrm{pop}_i}{d_i}$$

Introduction

- Classical purchase behavior models treat individual purchases separately (Huff, 1964)
- Study of purchase behavior influence is largely based on socio-demographics (Zeithaml, 1985)

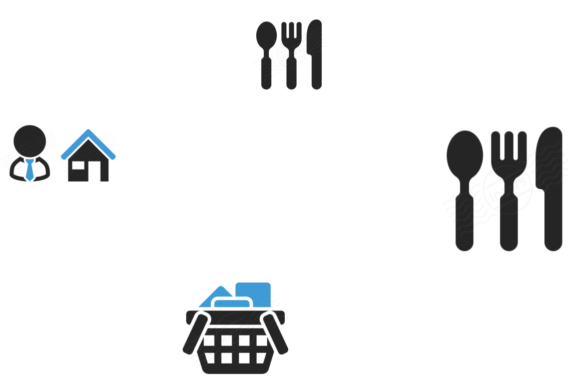


$$p_i \propto \frac{\mathrm{pop}_i}{d_i}$$

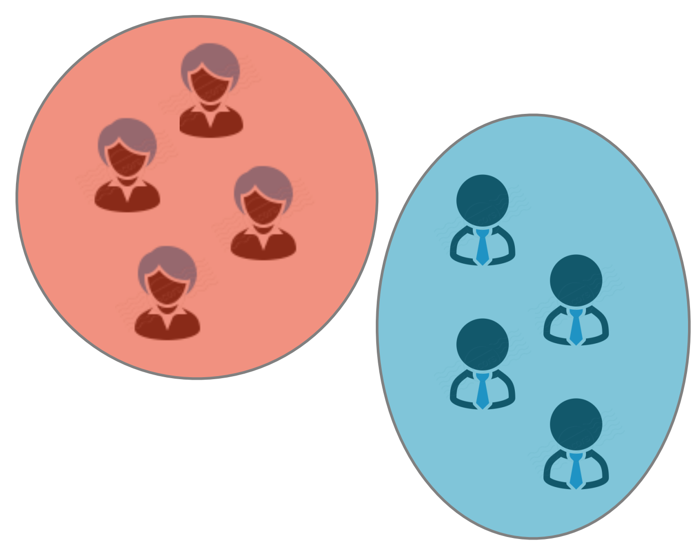


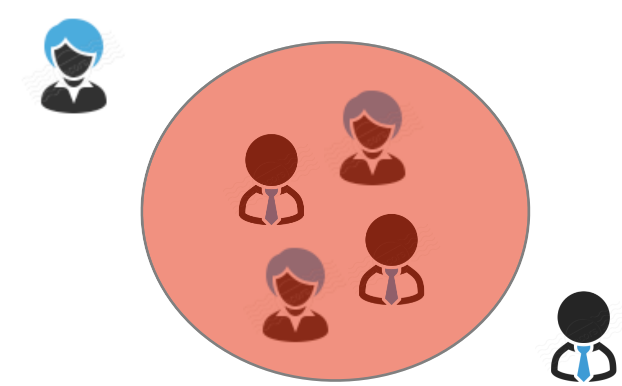
Introduction

- Classical purchase behavior models treat individual purchases separately (Huff, 1964)
- Study of purchase behavior influence is largely based on socio-demographics (Zeithaml, 1985)
- Word-of-mouth and physical exposure are powerful sources of behavioral propagation (Arndt, 1967; Bikhchandani, 1998; Algesheimer, 2005), but their effectiveness in modern city environment remains unknown



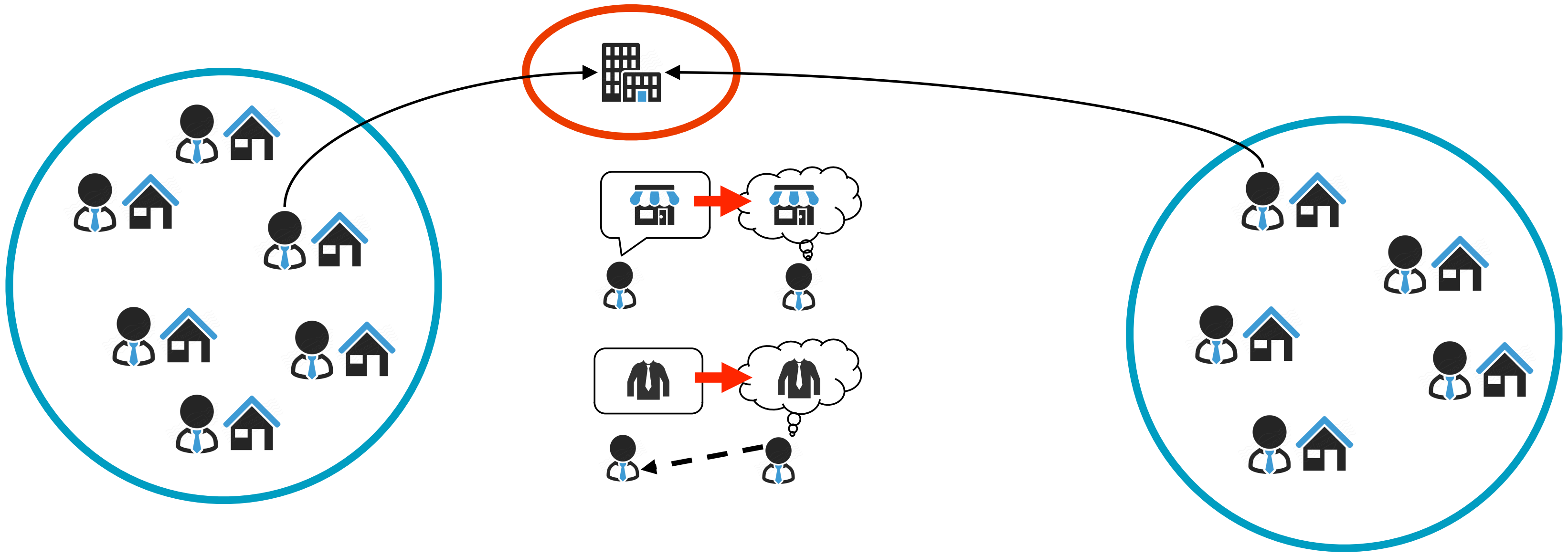
$$p_i \propto \frac{\mathrm{pop}_i}{d_i}$$





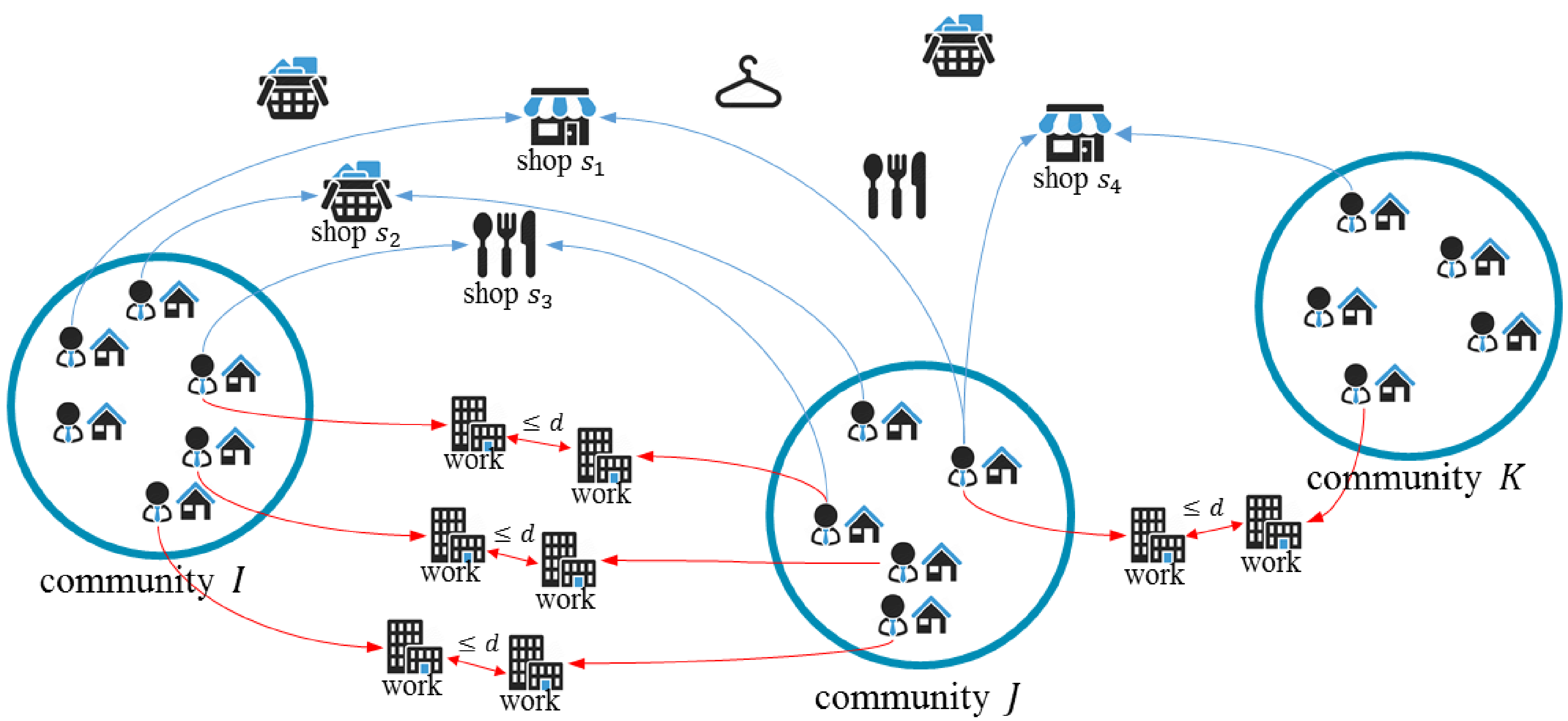
Introduction

- Hypothesis
  - Physical exposure at work environment promotes idea exchange



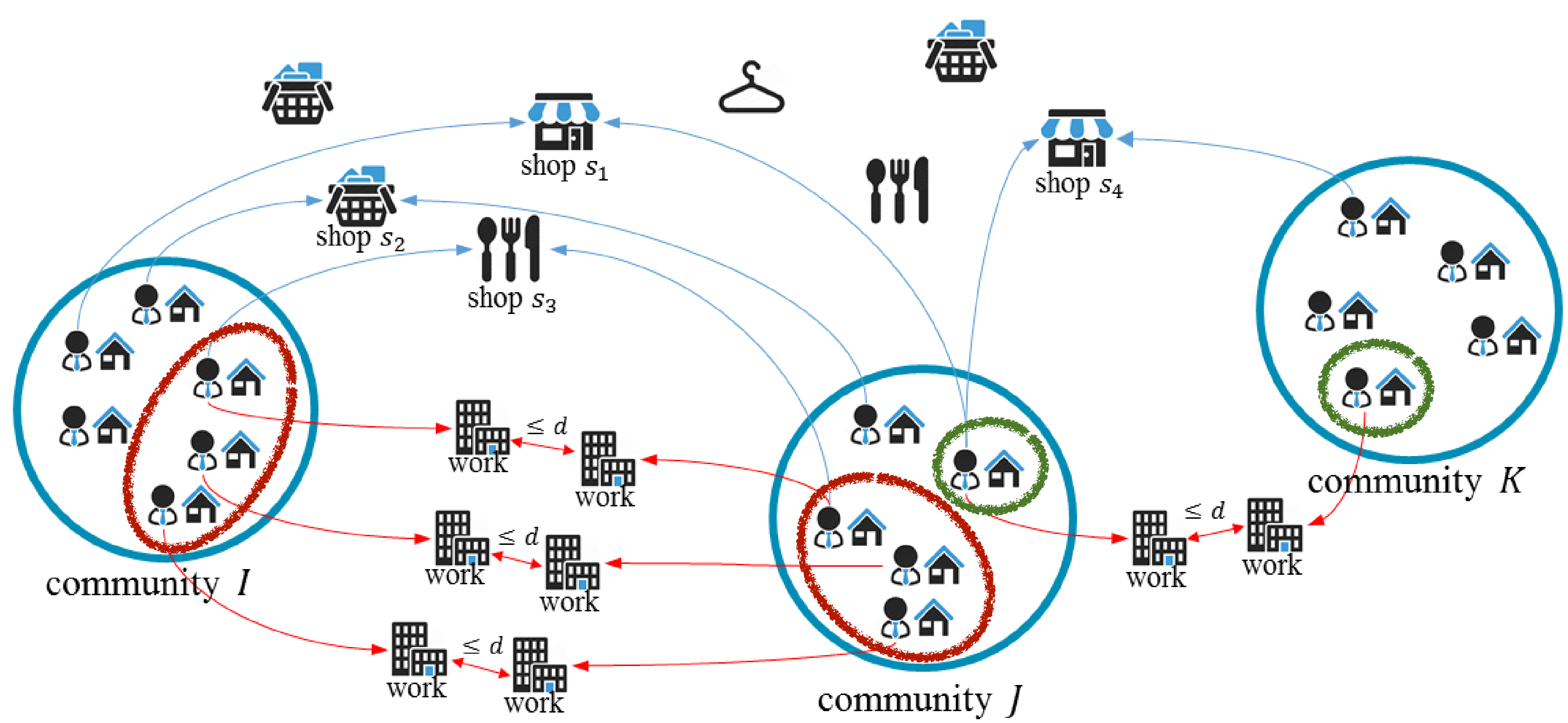
Introduction

- Hypothesis
  - Physical exposure at work environment promotes idea exchange
  - Individuals living in different communities but sharing similar work locations act as social bridges between communities



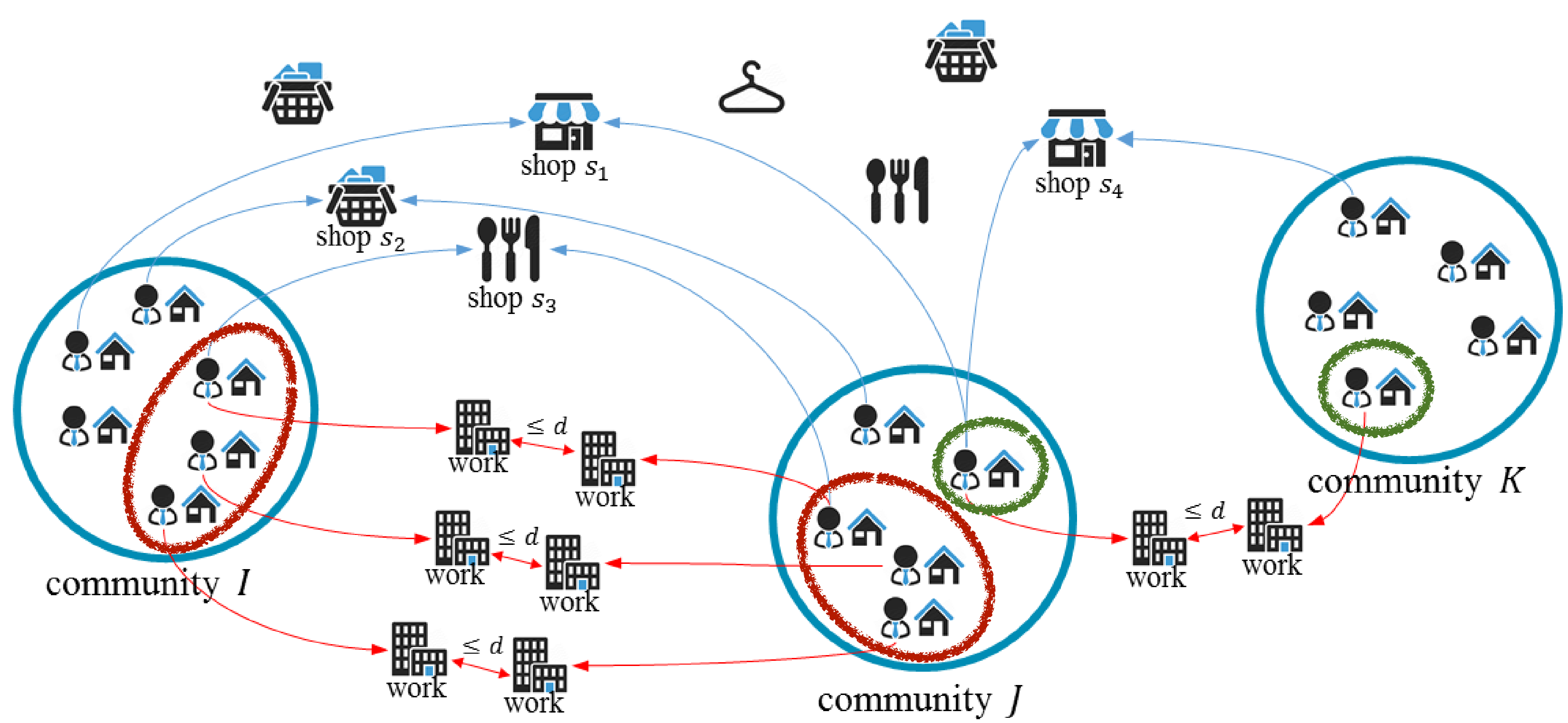
Introduction

- Hypothesis
  - Physical exposure at work environment promotes idea exchange
  - Individuals living in different communities but sharing similar work locations act as social bridges between communities



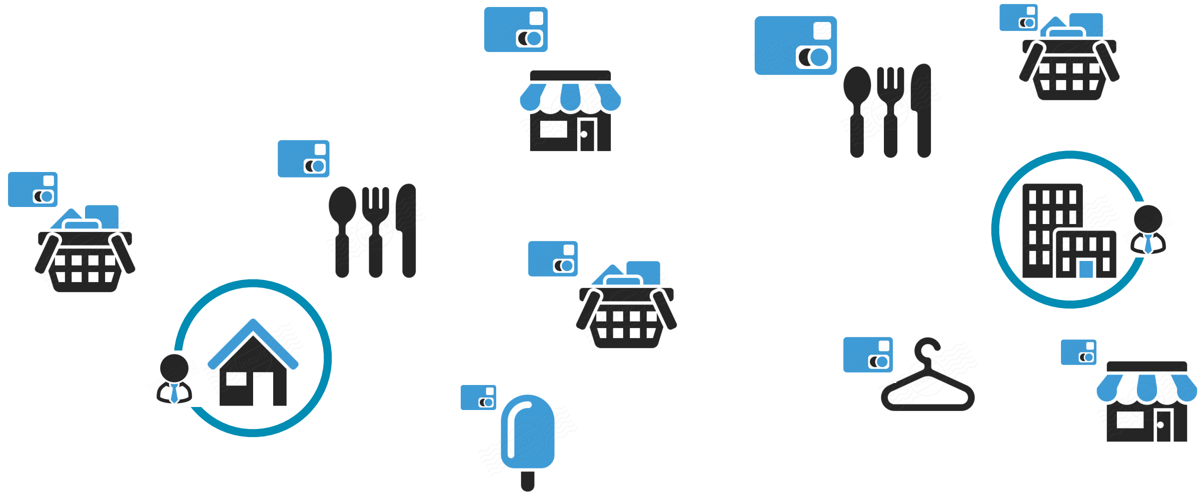
Introduction

- Hypothesis
  - Physical exposure at work environment promotes idea exchange
  - Individuals living in different communities but sharing similar work locations act as social bridges between communities
- Test at city scale



Data set

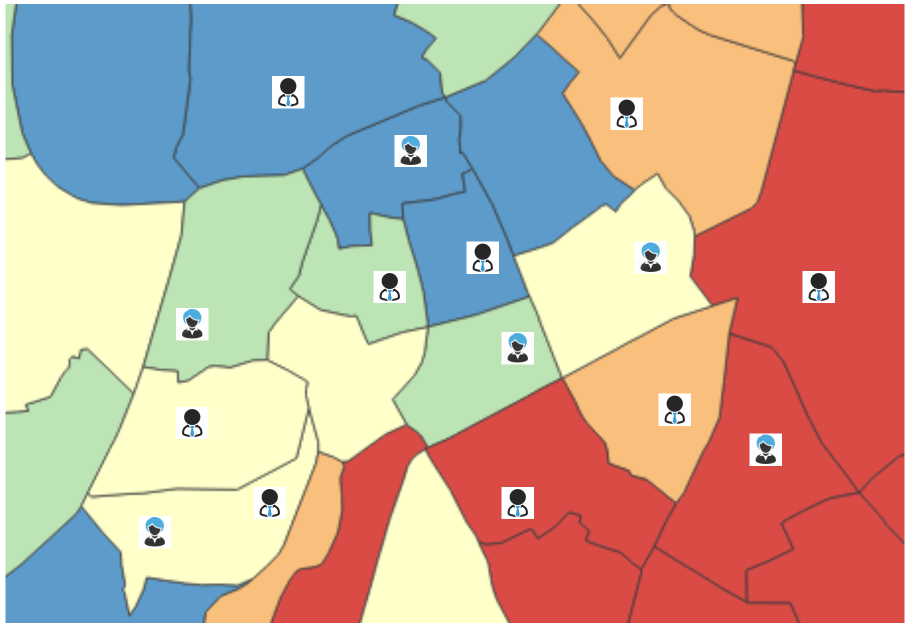
- A large-scale credit card transaction data set in two cities in an OECD country during 3 months



	City A	City B
# Customers	49K	9K
# Stores	110K	30K
# Transactions	2.3M	0.4M
% Female Customers	37.3%	31.9%
% Young (Below 30) Customers	20.5%	16.1%
% Single Customers	31.4%	22.7%
% College-Educated Customers	51.1%	47.5%
% Employed Customers	92.9%	92.1%
Median Income	2400	2100

Methods

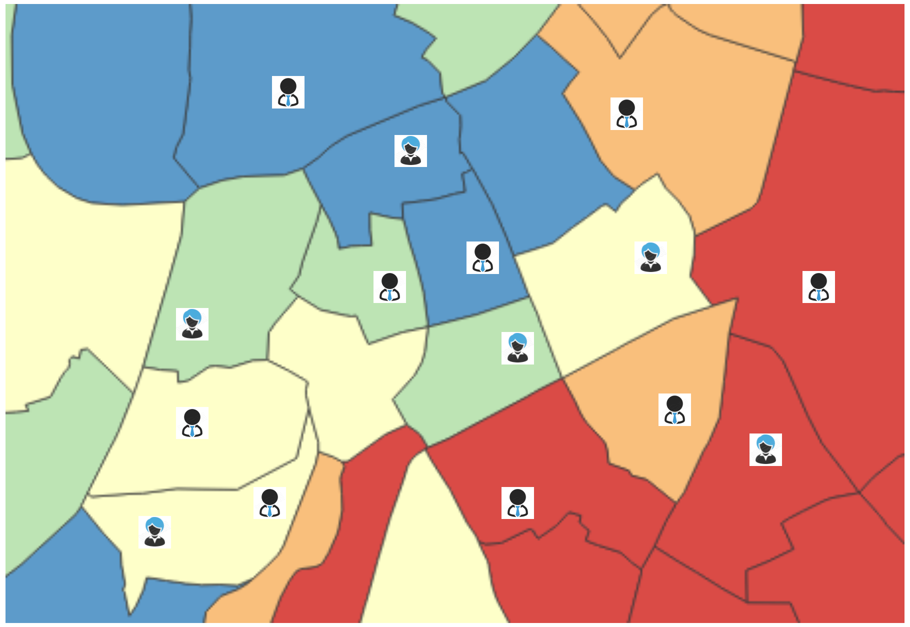
- Urban communities



Methods

- Urban communities
- Number of social bridges between communities

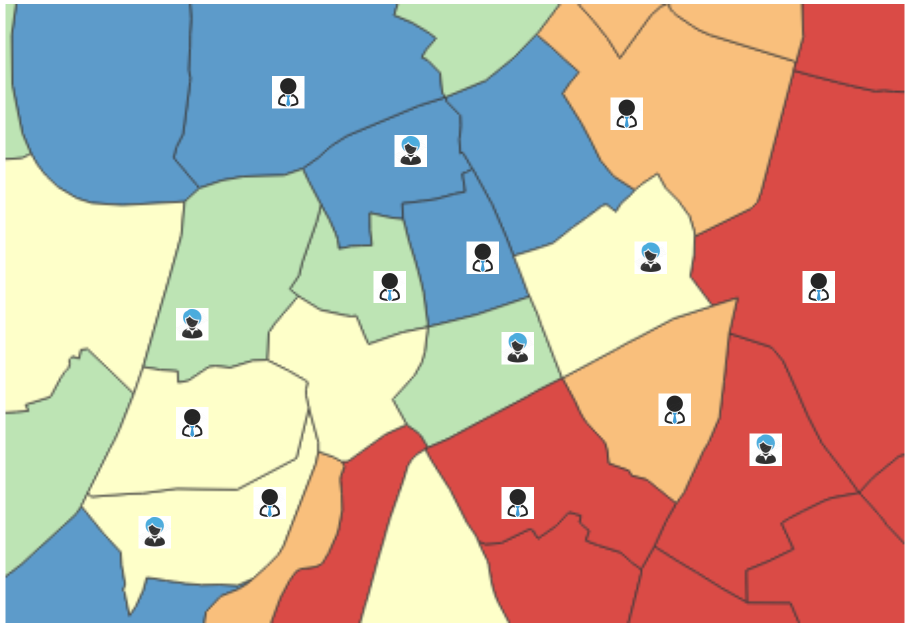
$$\mathrm{bdg}(I,J) = |\{i,j\}|$$
$$\text{s.t. } i \in I,\ j \in J,\ D(L_i, L_j) \leq d$$

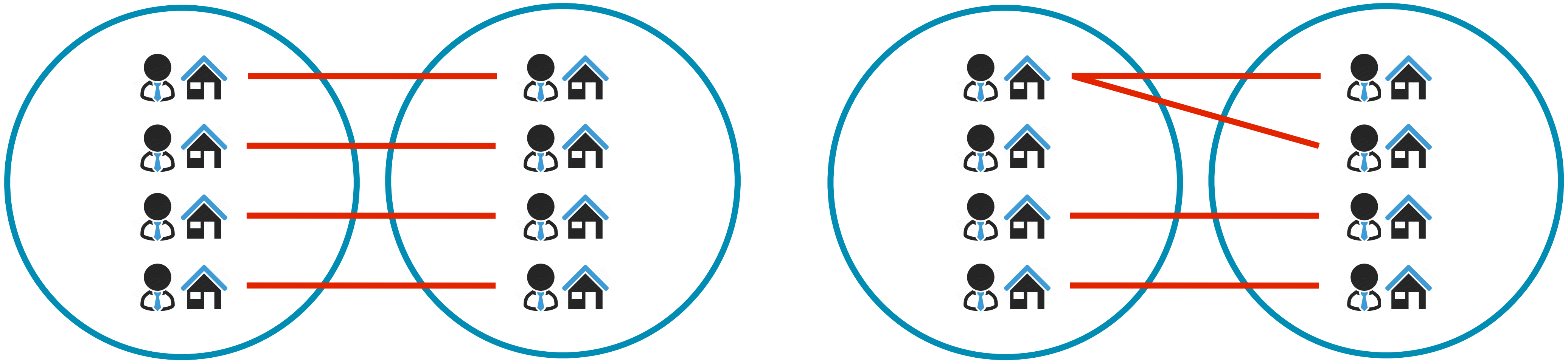


Methods

- Urban communities
- Number of social bridges between communities

$$\text{bdg}(I, J) = |\{i, j\}|$$
$$\text{s.t. } i \in I,\ j \in J,\ D(L_i, L_j) \leq d$$





bridges = 4

bridges = 4

Methods

- Three behavioral indexes
  - choice: number of co-visited stores
  - temporal: similarity between temporal distributions of purchases
  - spending: sum of differences in median spending amount of different categories

Methods

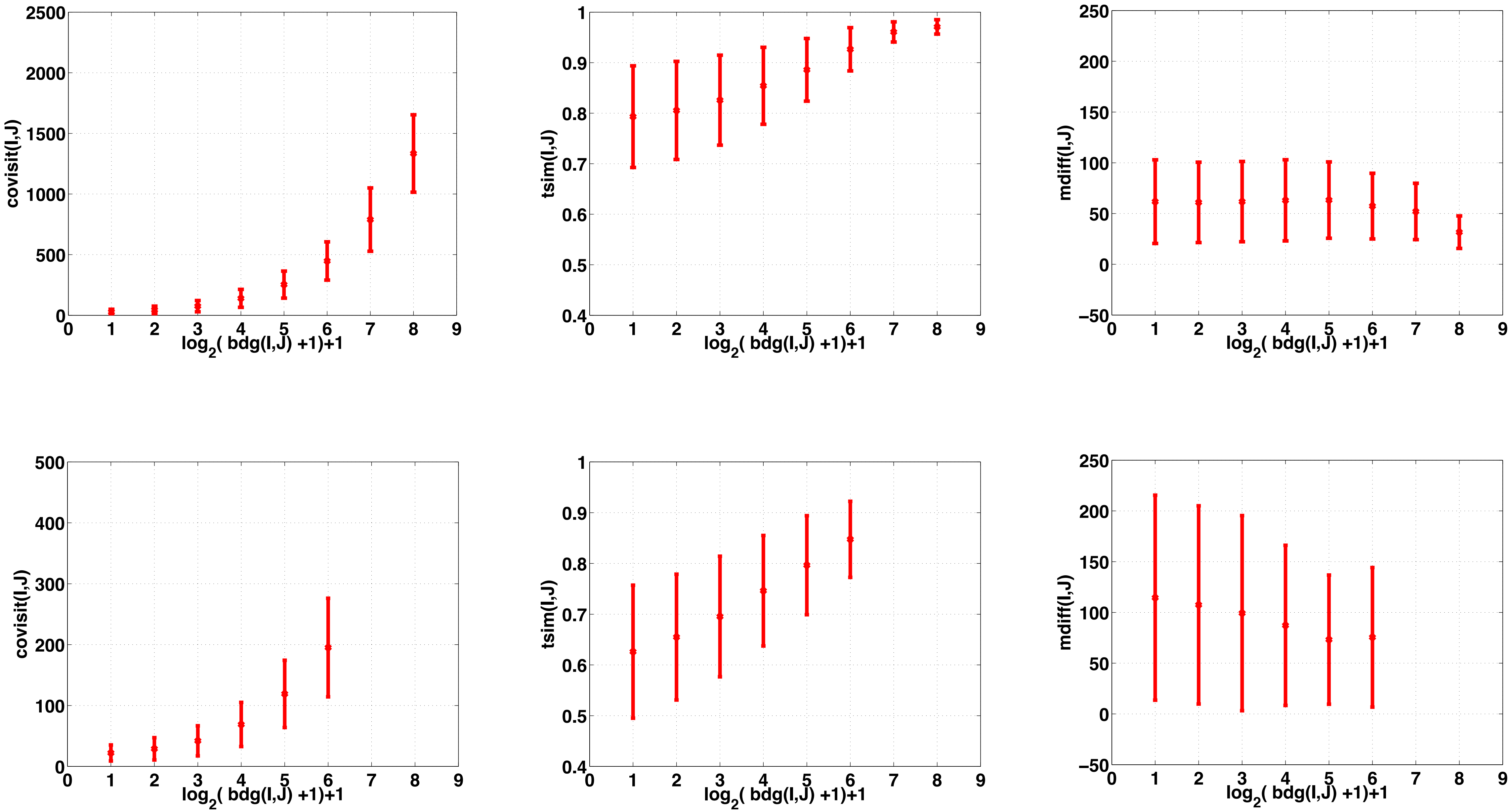
- Three behavioral indexes
  - choice: number of co-visited stores
  - temporal: similarity between temporal distributions of purchases
  - spending: sum of differences in median spending amount of different categories

- Remark
  - exclude transactions during working hours
  - exclude transactions at stores in home/work neighborhoods

Social bridge and behavioral indexes

City A

City B



Choice (co-visits)

Temporal

Spending

Social bridge and purchase similarity (co-visits)

- Multiple OLS regression analysis
  - dependent variable (DV): # co-visits (between community pair)
  - independent variables (IV): # social bridges
  - confounding variables: population, distance, demographics, income

Social bridge and purchase similarity (co-visits)

- Multiple OLS regression analysis
  - dependent variable (DV): # co-visits (between community pair)
  - independent variables (IV): # social bridges
  - confounding variables: population, distance, demographics, income
- Remark
  - entries are not independent in DV and IV
  - Quadratic Assignment Procedure (QAP) to test statistical significance
    - random shuffling of communities in DV
    - re-application of OLS

Social bridge and purchase similarity (co-visits)

- Regression coefficients

(a) City A

Indicator	β Coefficient	Confidence Interval
# Social Bridge	0.760 ***	[0.754, 0.766]
Population	0.102 ***	[0.095, 0.108]
Distance	0.094 ***	[0.090, 0.097]
Age	0.038 ***	[0.034, 0.042]
Gender	0.015 ***	[0.011, 0.019]
Marital Status	0.017 ***	[0.013, 0.021]
Education	0.046 ***	[0.042, 0.051]
Working Style	0.015 ***	[0.011, 0.019]
Income	0.034 ***	[0.030, 0.039]
Num. Obs.		61776
RMSE		0.465
Adj. R^2		0.784

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

(b) City B

Indicator	β Coefficient	Confidence Interval
# Social Bridge	0.410 ***	[0.393, 0.426]
Population	0.288 ***	[0.272, 0.305]
Distance	0.167 ***	[0.156, 0.179]
Age	0.060 ***	[0.048, 0.072]
Gender	0.155 ***	[0.143, 0.167]
Marital Status	0.023 ***	[0.011, 0.035]
Education	-0.008	[-0.021, 0.005]
Working Style	0.031 ***	[0.019, 0.043]
Income	0.085 ***	[0.072, 0.099]
Num. Obs.		12403
RMSE		0.643
Adj. R^2		0.586

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Social bridge and purchase similarity (co-visits)

- Regression coefficients

(a) City A

Indicator	β Coefficient	Confidence Interval
# Social Bridge	0.760 ***	[0.754, 0.766]
Population	0.102 ***	[0.095, 0.108]
Distance	0.094 ***	[0.090, 0.097]
Age	0.038 ***	[0.034, 0.042]
Gender	0.015 ***	[0.011, 0.019]
Marital Status	0.017 ***	[0.013, 0.021]
Education	0.046 ***	[0.042, 0.051]
Working Style	0.015 ***	[0.011, 0.019]
Income	0.034 ***	[0.030, 0.039]
Num. Obs.		61776
RMSE		0.465
Adj. R^2		0.784

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(b) City B

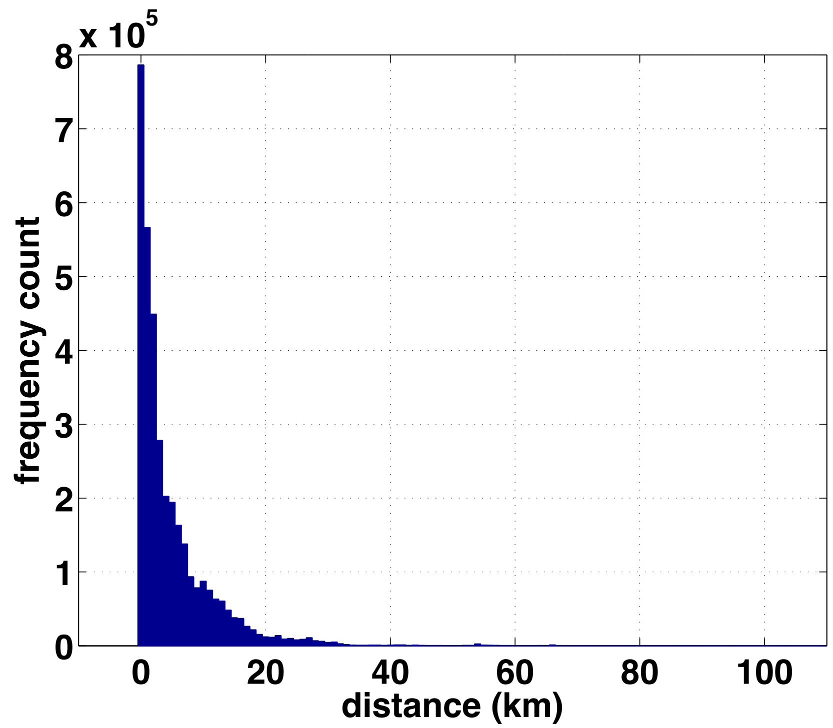
Indicator	β Coefficient	Confidence Interval
# Social Bridge	0.410 ***	[0.393, 0.426]
Population	0.288 ***	[0.272, 0.305]
Distance	0.167 ***	[0.156, 0.179]
Age	0.060 ***	[0.048, 0.072]
Gender	0.155 ***	[0.143, 0.167]
Marital Status	0.023 ***	[0.011, 0.035]
Education	-0.008	[-0.021, 0.005]
Working Style	0.031 ***	[0.019, 0.043]
Income	0.085 ***	[0.072, 0.099]
Num. Obs.		12403
RMSE		0.643
Adj. R^2		0.586

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

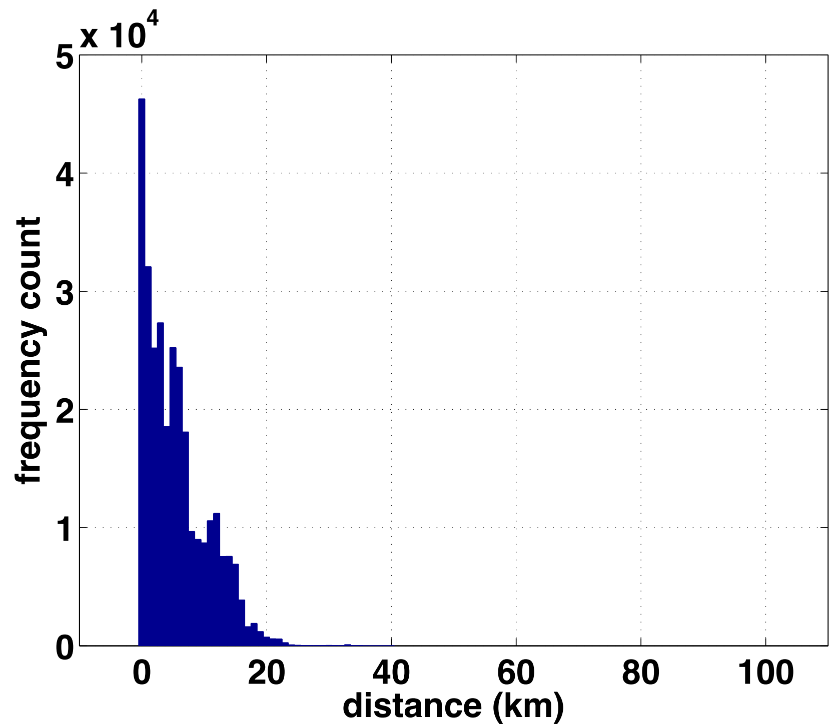
Social bridge is a stronger indicator of similar purchase behavior

Social bridge and purchase similarity (co-visits)

- Histogram of distance between co-visited store and co-working location



City A (62% > 2km)

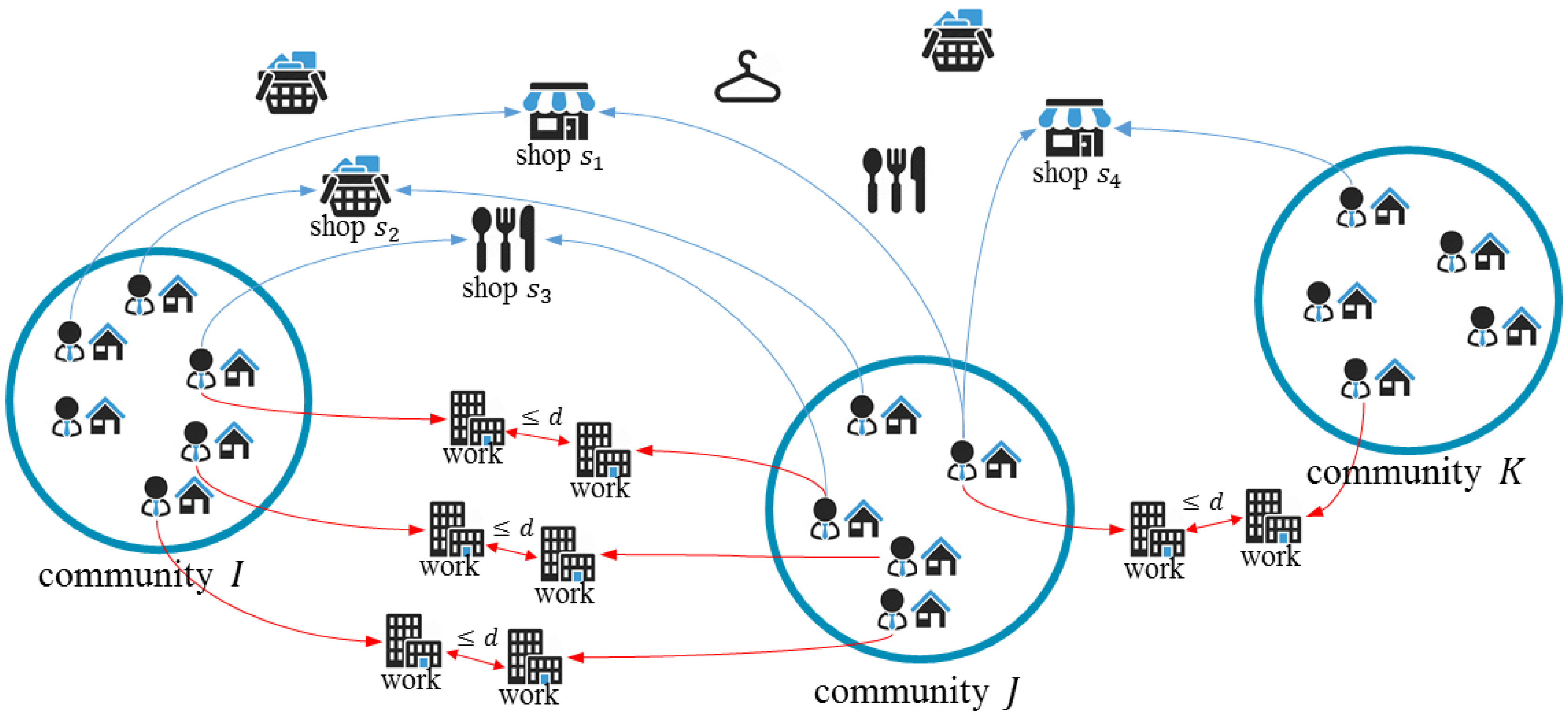


City B (74% > 2km)

Co-visitation is not simply due to proximity between co-visited store and co-working location

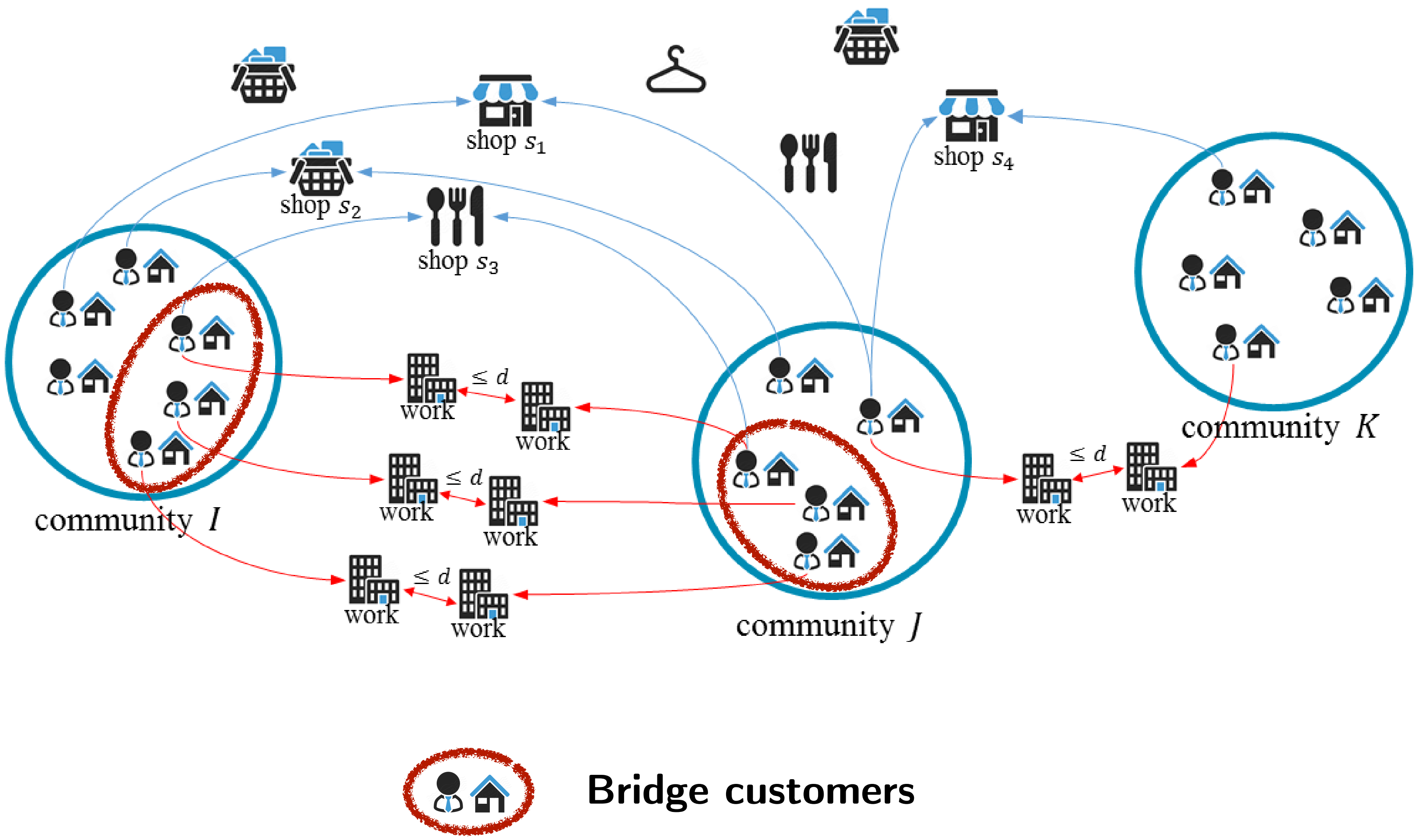
Co-visits by two types of customers

- Bridge customers vs. Non-bridge customers



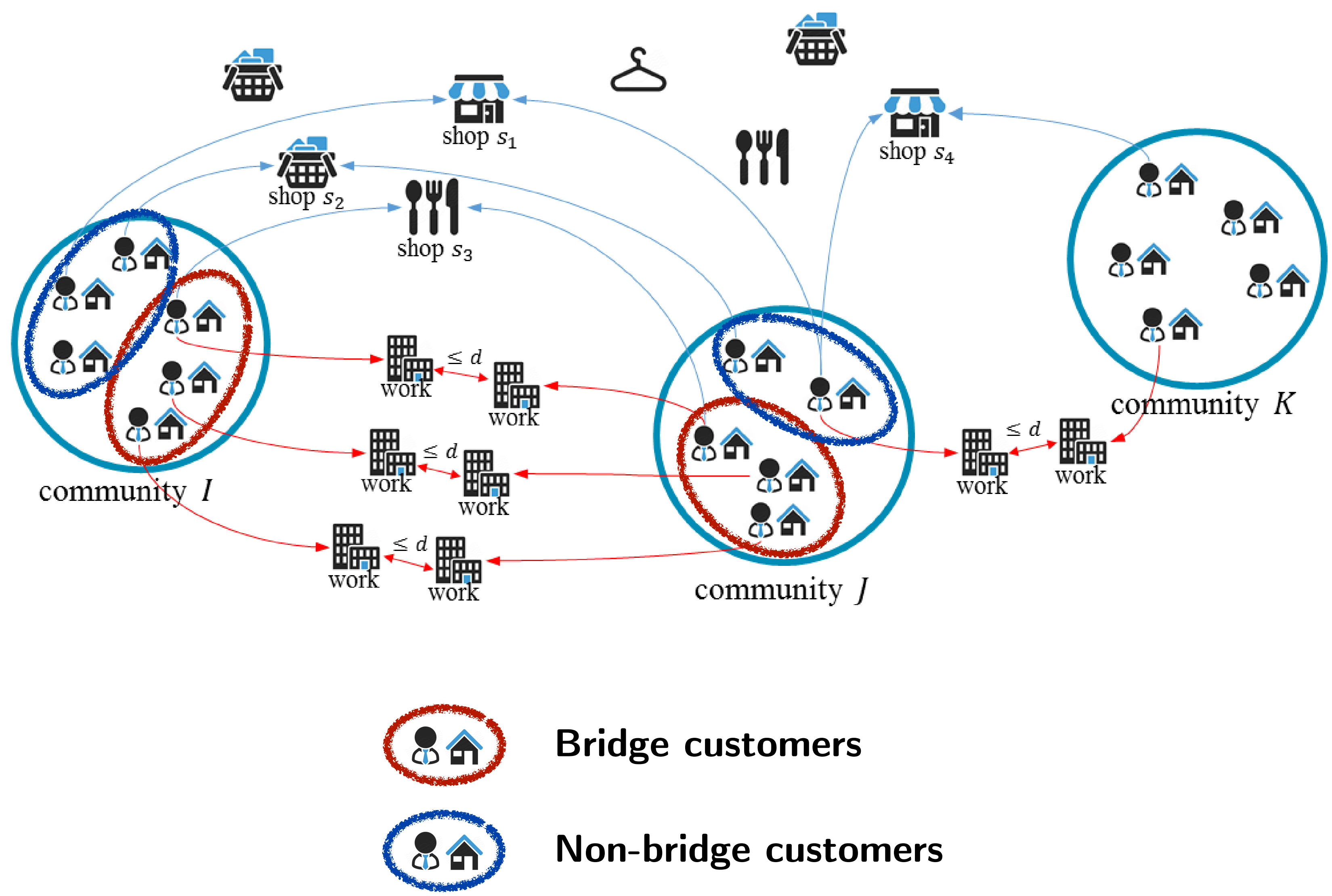
Co-visits by two types of customers

- Bridge customers vs. Non-bridge customers



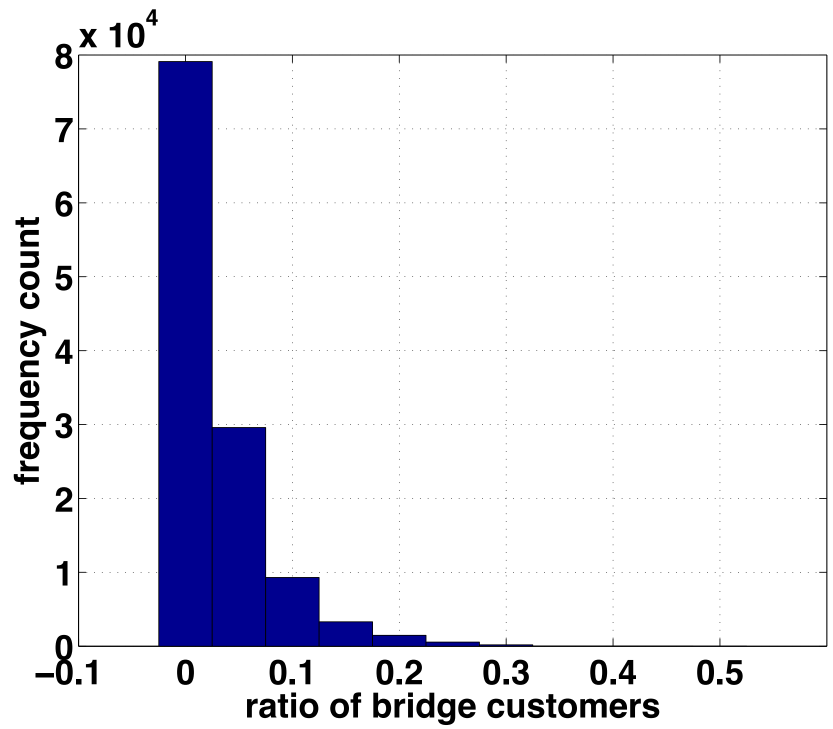
Co-visits by two types of customers

- Bridge customers vs. Non-bridge customers

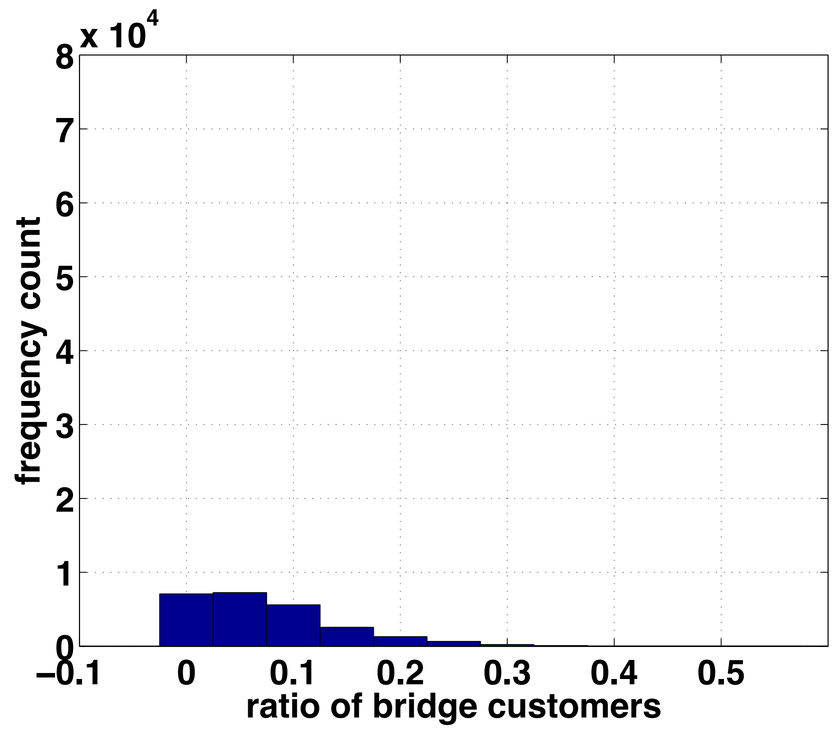


Co-visits by two types of customers

- Histogram of ratio of bridge customers



City A

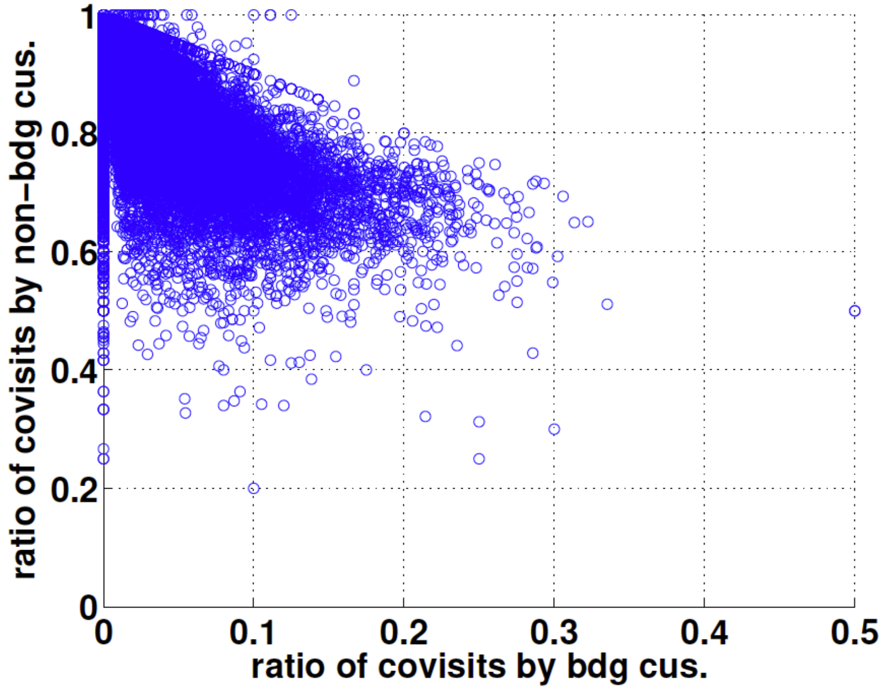


City B

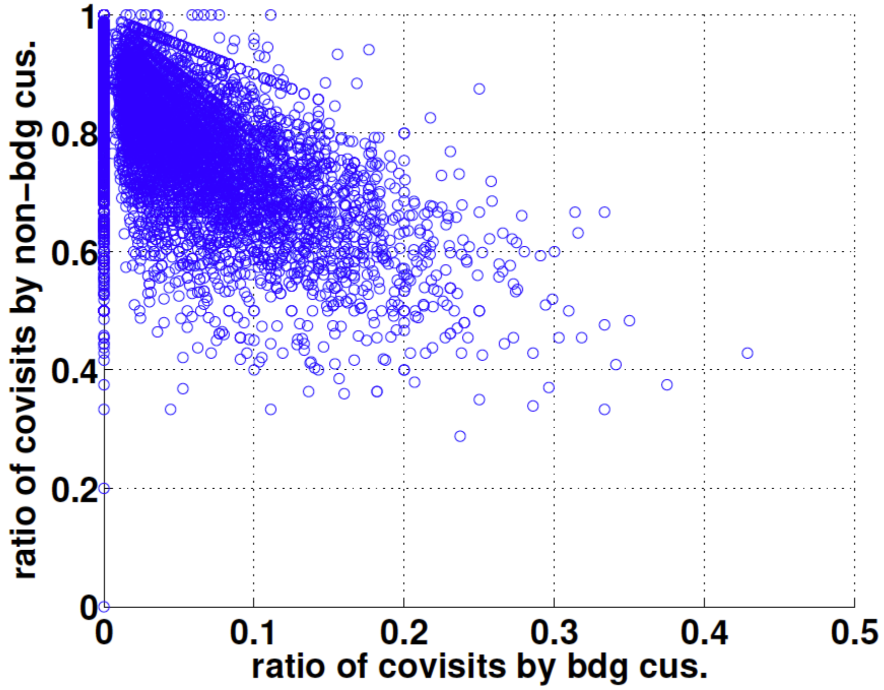
Ratio of bridge customers are relatively small

Co-visits by two types of customers

- Percentage of co-visits by bridge customers



City A



City B

A large portion of co-visits are by non-bridge customers

Co-visits by two types of customers

- Regression coefficients

(a) City A

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
By All	0.760 ***	[0.754, 0.766]	0.784
By Bridge Cus.	1.005 ***	[0.999, 1.011]	0.766
By Non-Bridge Cus.	0.653 ***	[0.646, 0.660]	0.705

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(b) City B

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
By All	0.410 ***	[0.393, 0.426]	0.586
By Bridge Cus.	0.717 ***	[0.700, 0.734]	0.558
By Non-Bridge Cus.	0.238 ***	[0.220, 0.256]	0.490

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Co-visits by two types of customers

- Regression coefficients

(a) City A

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
By All	0.760 ***	[0.754, 0.766]	0.784
By Bridge Cus.	1.005 ***	[0.999, 1.011]	0.766
By Non-Bridge Cus.	0.653 ***	[0.646, 0.660]	0.705

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(b) City B

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
By All	0.410 ***	[0.393, 0.426]	0.586
By Bridge Cus.	0.717 ***	[0.700, 0.734]	0.558
By Non-Bridge Cus.	0.238 ***	[0.220, 0.256]	0.490

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Social bridge is a indicator of similar purchase behavior even for non-bridge customers

Co-visits in three merchant categories

- Regression coefficients

(a) City A

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
Supermarkets	0.610 ***	[0.603, 0.618]	0.693
Restaurants	0.812 ***	[0.805, 0.818]	0.776
Clothing Stores	0.623 ***	[0.615, 0.631]	0.631

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(b) City B

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
Supermarkets	0.291 ***	[0.274, 0.309]	0.537
Restaurants	0.445 ***	[0.426, 0.465]	0.399
Clothing Stores	0.330 ***	[0.312, 0.347]	0.539

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Co-visits in three merchant categories

- Regression coefficients

(a) City A

Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
Supermarkets	0.610 ***	[0.603, 0.618]	0.693
Restaurants	0.812 ***	[0.805, 0.818]	0.776
Clothing Stores	0.623 ***	[0.615, 0.631]	0.631

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(b) City B

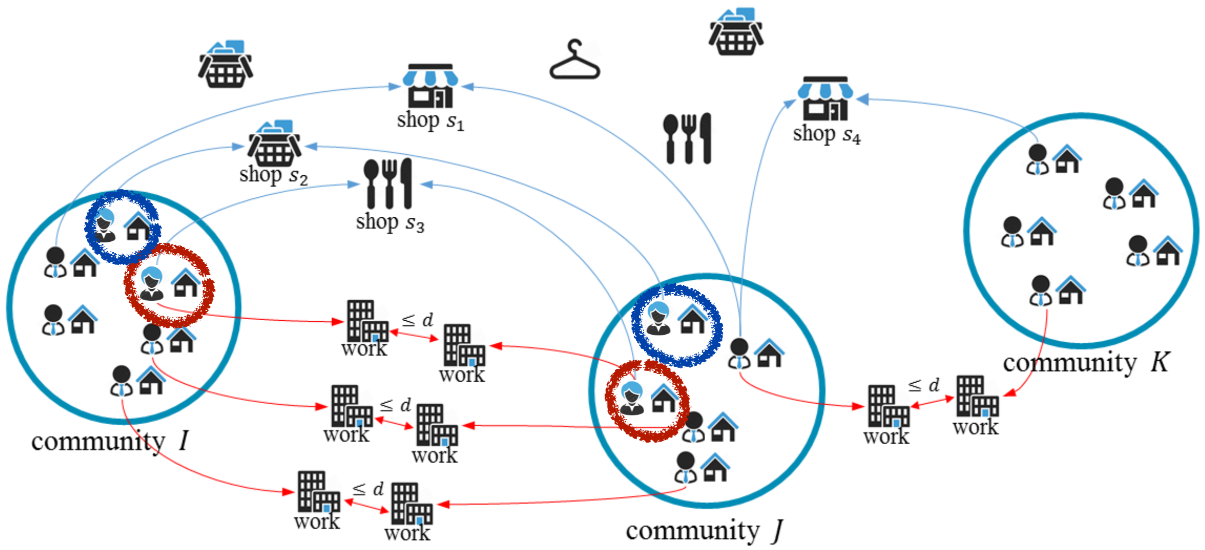
Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
Supermarkets	0.291 ***	[0.274, 0.309]	0.537
Restaurants	0.445 ***	[0.426, 0.465]	0.399
Clothing Stores	0.330 ***	[0.312, 0.347]	0.539

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Effect of social bridge is stronger for restaurants but weaker for supermarkets

Gender difference in social bridge

- Regression coefficients



(a) City A

Bridge Types	Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
Female-Female	**By Non-Bridge Female**	0.527 ***	[0.520, 0.533]	0.625
Female-Female	**By Non-Bridge Male**	0.404 ***	[0.398, 0.411]	0.615
Male-Male	**By Non-Bridge Female**	0.360 ***	[0.352, 0.368]	0.543
Male-Male	**By Non-Bridge Male**	0.393 ***	[0.385, 0.400]	0.604

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

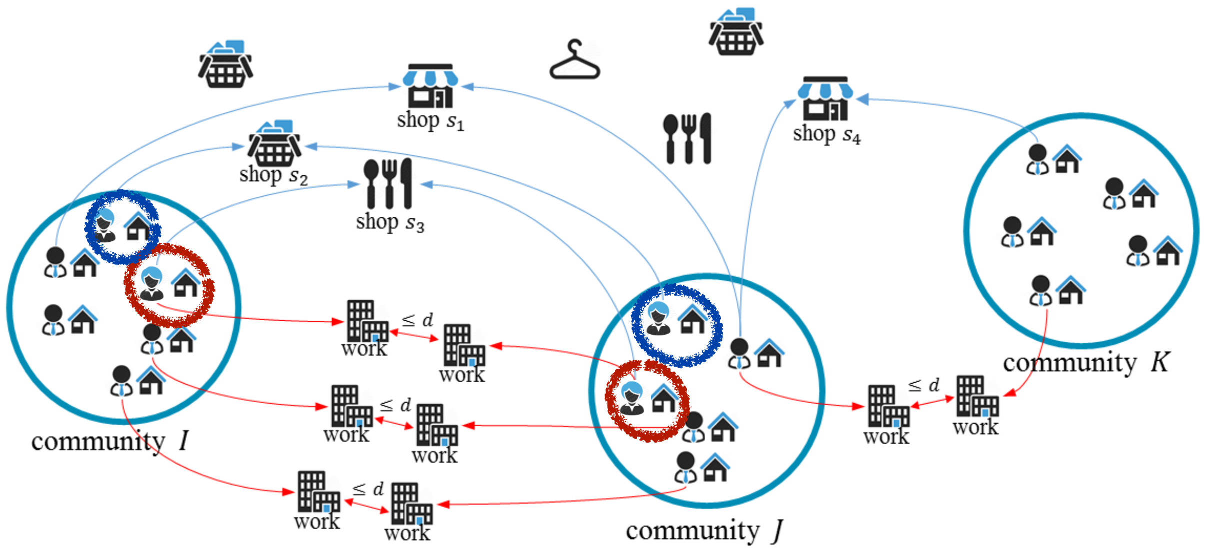
(b) City B

Bridge Types	Co-Visits Types	β Coefficient	Confidence Interval	**Adj. R^2**
Female-Female	**By Non-Bridge Female**	0.327 ***	[0.311, 0.343]	0.340
Female-Female	**By Non-Bridge Male**	0.106 ***	[0.091, 0.120]	0.468
Male-Male	**By Non-Bridge Female**	-0.073 ***	[-0.092, -0.055]	0.261
Male-Male	**By Non-Bridge Male**	0.044 ***	[0.028, 0.060]	0.460

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Gender difference in social bridge

- Regression coefficients



(a) City A

Bridge Types	Co-Visits Types	β Coefficient	Confidence Interval	Adj. R^2
Female-Female	**By Non-Bridge Female**	0.527 ***	[0.520, 0.533]	0.625
Female-Female	**By Non-Bridge Male**	0.404 ***	[0.398, 0.411]	0.615
Male-Male	**By Non-Bridge Female**	0.360 ***	[0.352, 0.368]	0.543
Male-Male	**By Non-Bridge Male**	0.393 ***	[0.385, 0.400]	0.604

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(b) City B

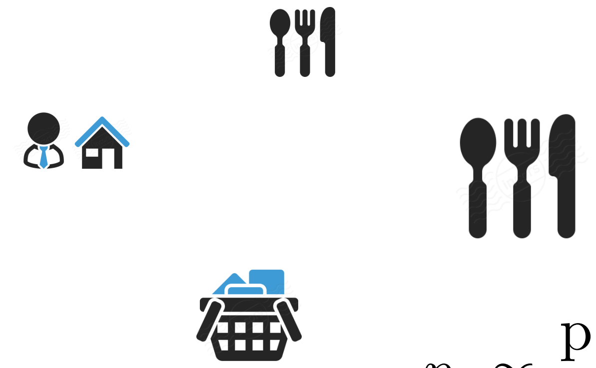
Bridge Types	Co-Visits Types	β Coefficient	Confidence Interval	Adj. R^2
Female-Female	**By Non-Bridge Female**	0.327 ***	[0.311, 0.343]	0.340
Female-Female	**By Non-Bridge Male**	0.106 ***	[0.091, 0.120]	0.468
Male-Male	**By Non-Bridge Female**	-0.073 ***	[-0.092, -0.055]	0.261
Male-Male	**By Non-Bridge Male**	0.044 ***	[0.028, 0.060]	0.460

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Female-female bridges show a stronger effect

Comparison with a null model

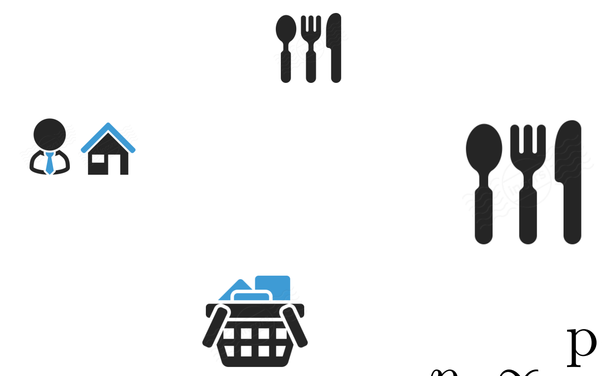
- Purchase choices are influenced by merchant popularity and location (Huff, 1964)



$$p_i \propto \frac{\mathrm{pop}_i}{d_i}$$

Comparison with a null model

- Purchase choices are influenced by merchant popularity and location (Huff, 1964)



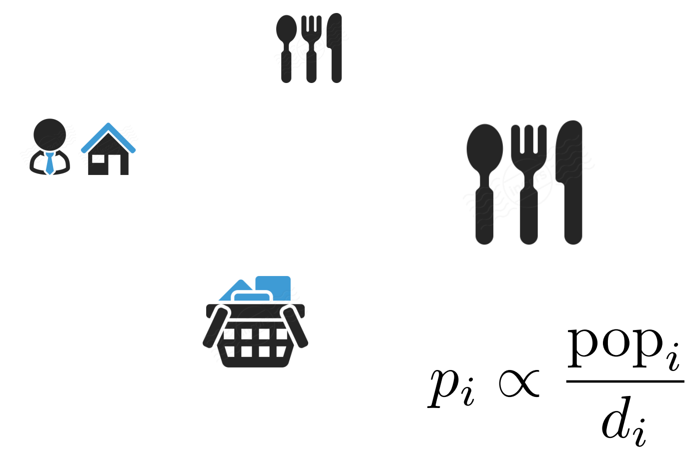
$$p_i \propto \frac{\mathrm{pop}_i}{d_i}$$

probability customer **i** visits store **s**

$$p_{is} = \frac{u_{is}}{\sum_{s \in S} u_{is}} = \frac{A_s^{\alpha_1} / D_{is}^{\alpha_2}}{\sum_{s \in S}(A_s^{\alpha_1} / D_{is}^{\alpha_2})}$$

Comparison with a null model

- Purchase choices are influenced by merchant popularity and location (Huff, 1964)



probability customer **i** visits store **s**

popularity of store **s**

distance between customer **i** and store **s**

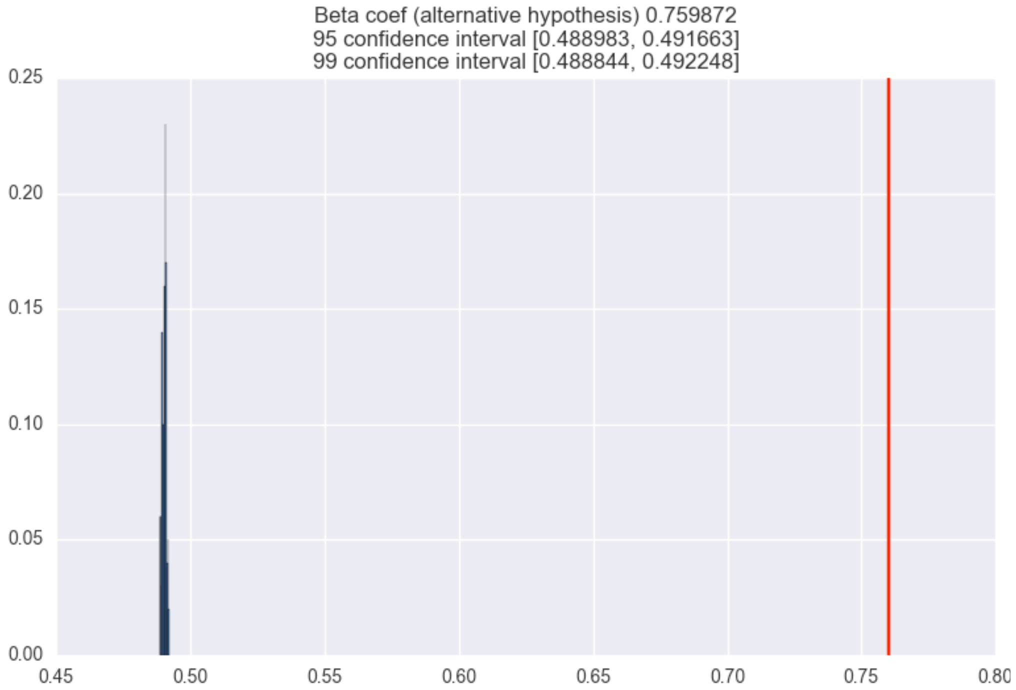
$$p_{is} = \frac{u_{is}}{\sum_{s \in S} u_{is}} = \frac{A_s^{\alpha_1} / D_{is}^{\alpha_2}}{\sum_{s \in S} (A_s^{\alpha_1} / D_{is}^{\alpha_2})}$$

Comparison with a null model

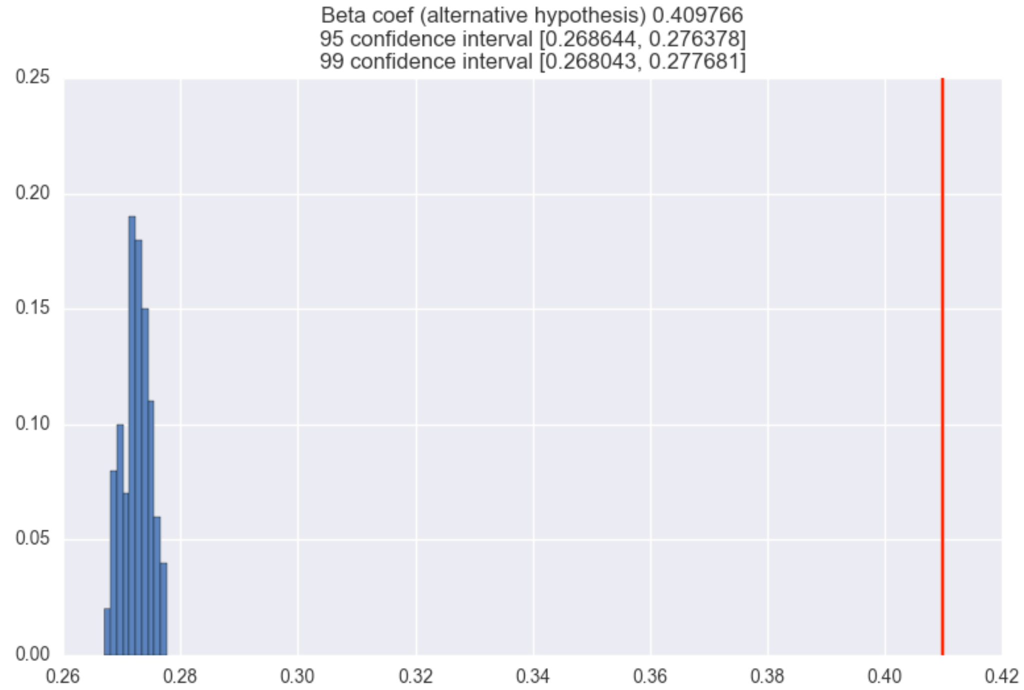
- Simulate individual purchases and co-visitation between communities
- Compare the regression coefficient with the empirical one

Comparison with a null model

- Simulate individual purchases and co-visitation between communities
- Compare the regression coefficient with the empirical one



City A



City B

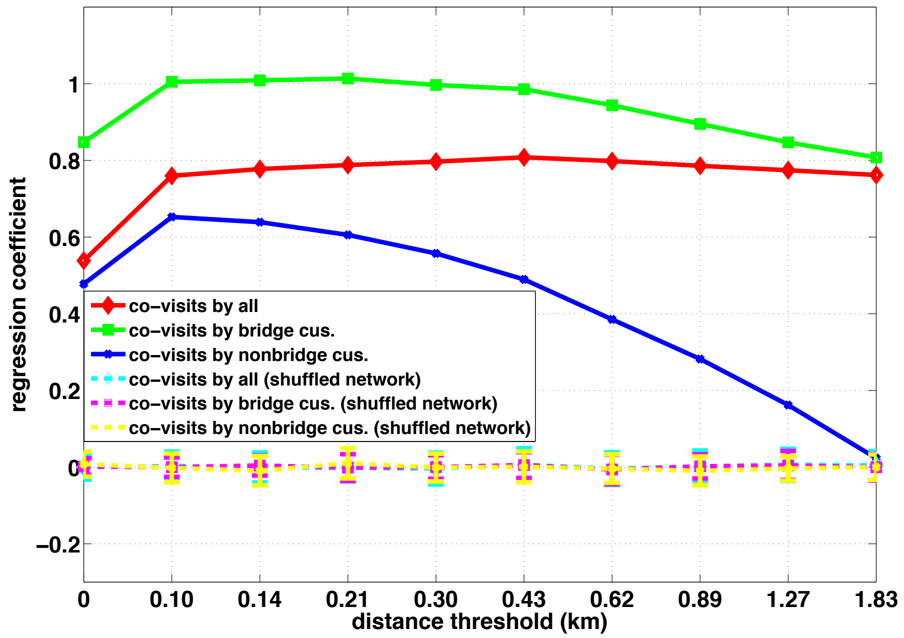
Effect of social bridge is not simply due to merchant popularity and location

Influence of distance threshold

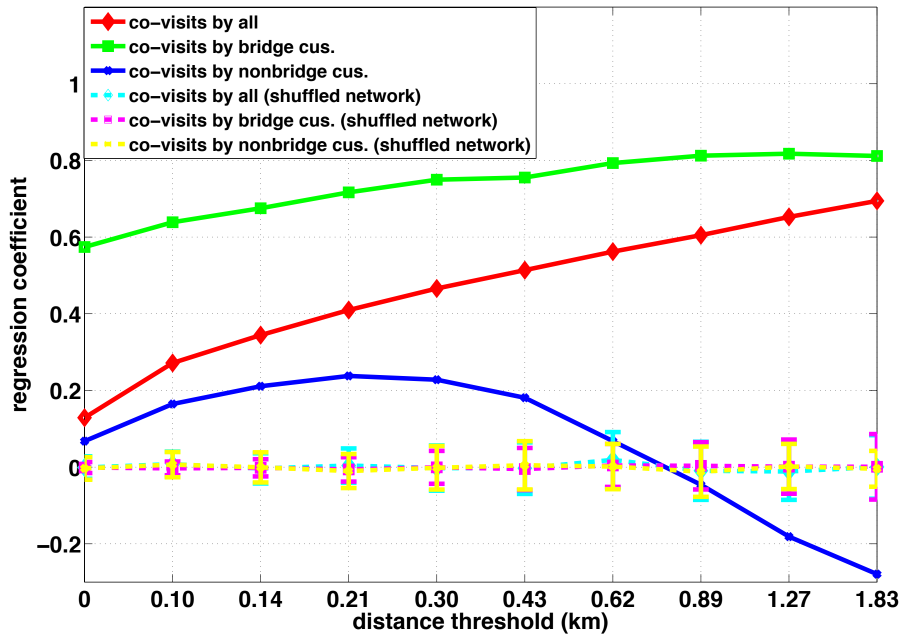
- Regression coefficient as a function of distance **d**

Influence of distance threshold

- Regression coefficient as a function of distance **d**



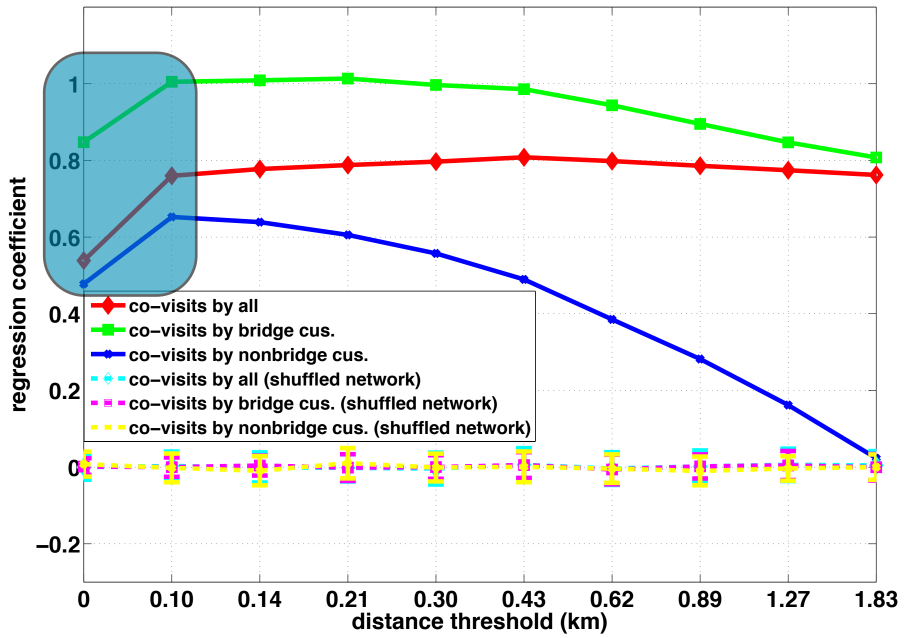
City A



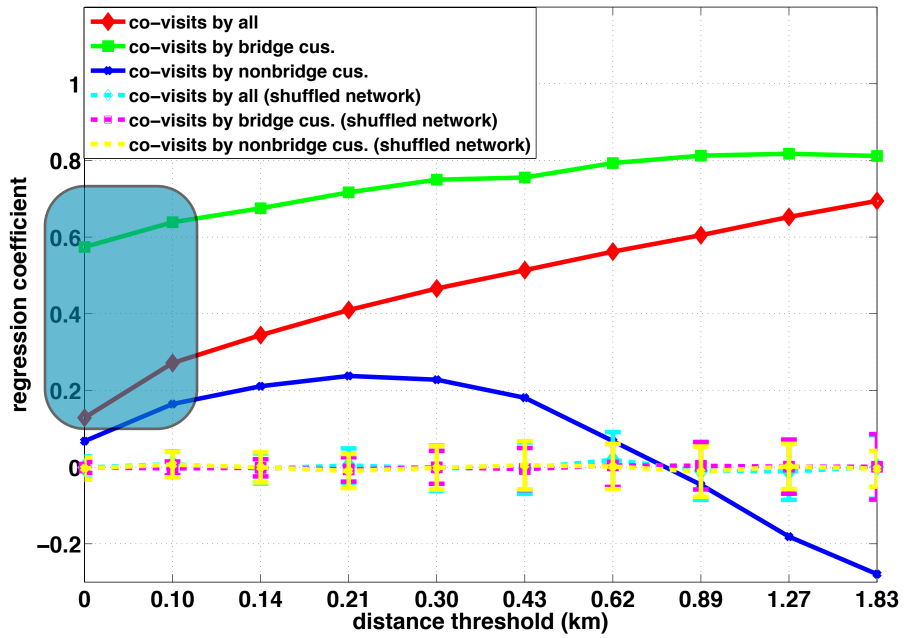
City B

Influence of distance threshold

- Regression coefficient as a function of distance **d**



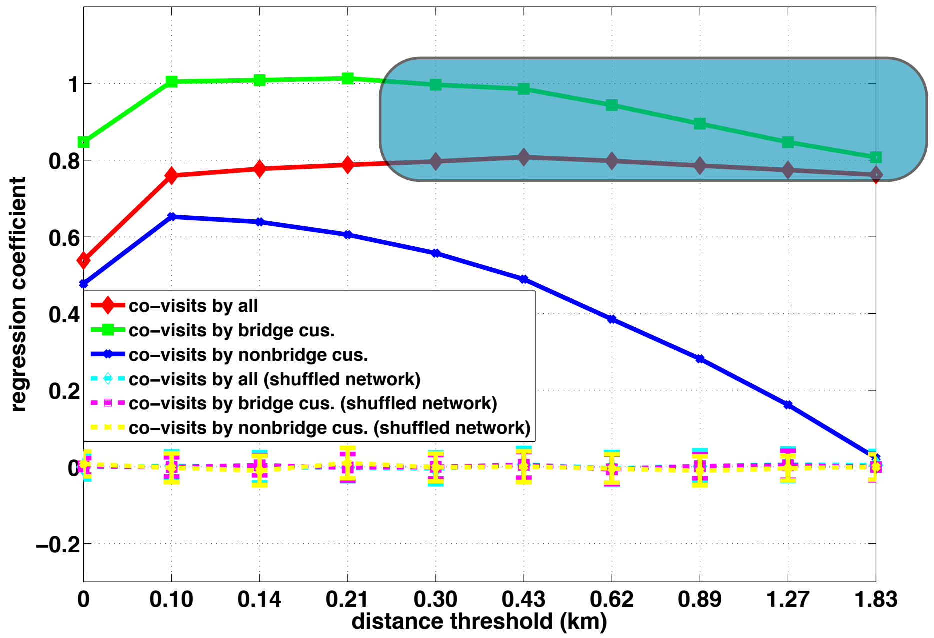
City A



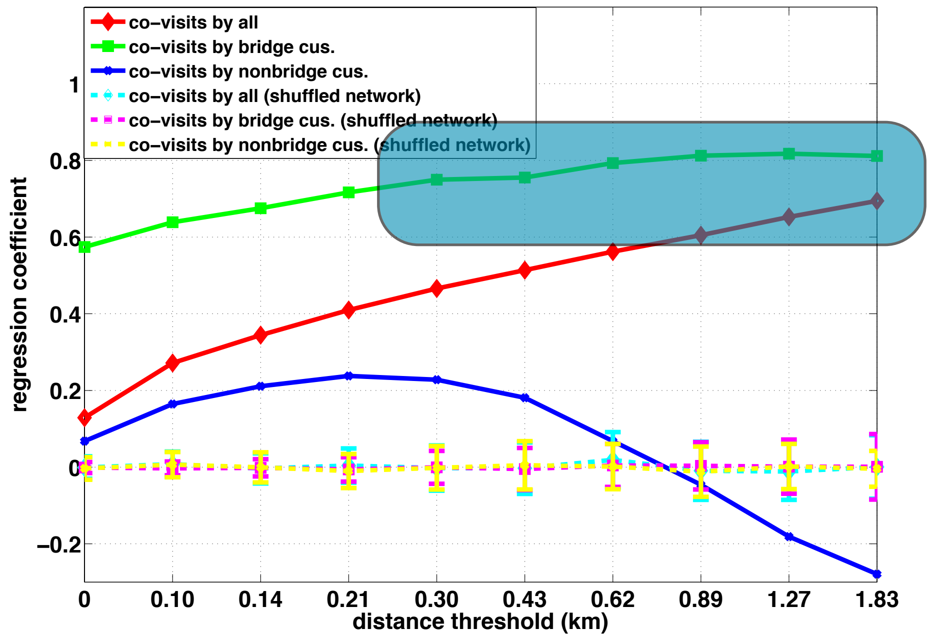
City B

Influence of distance threshold

- Regression coefficient as a function of distance **d**



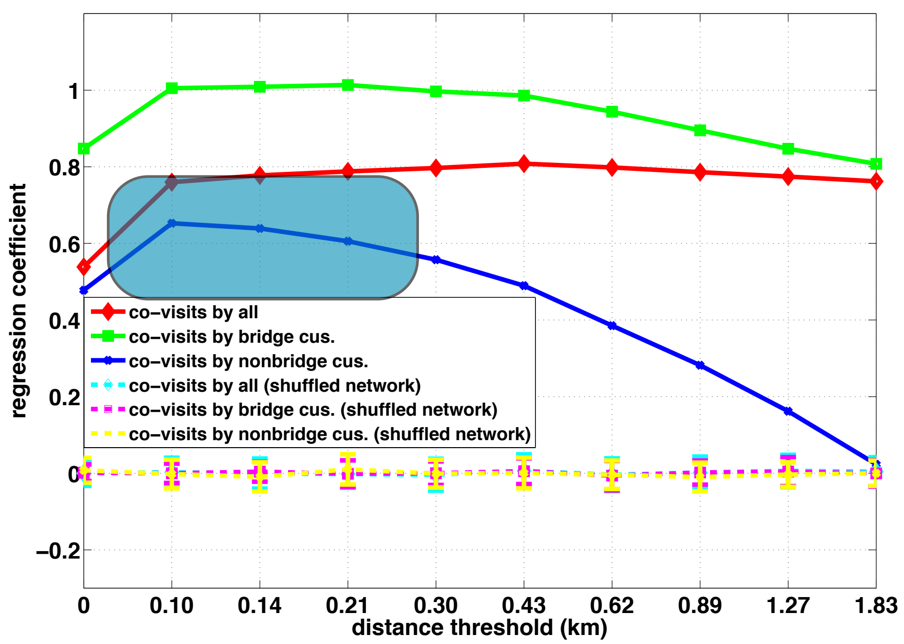
City A



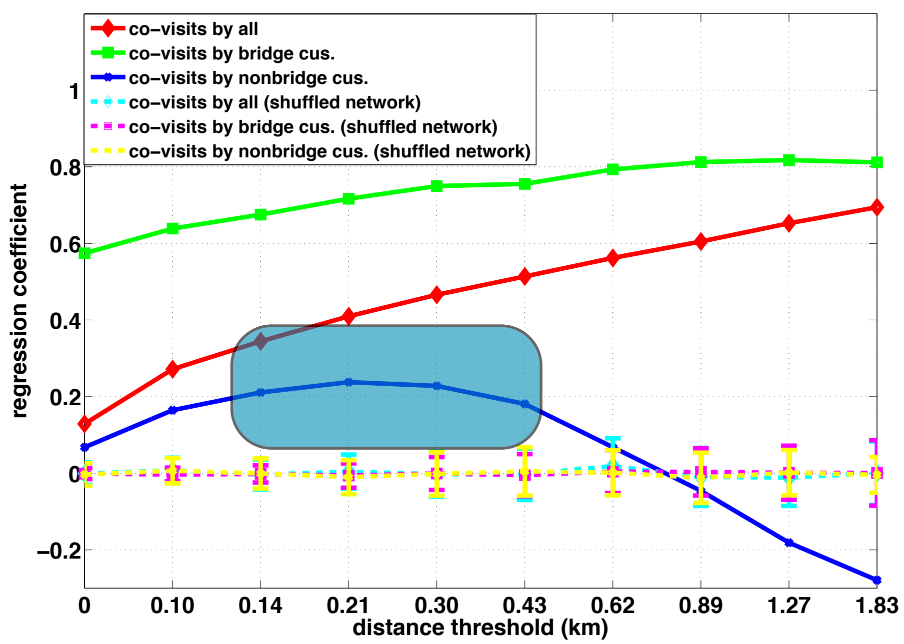
City B

Influence of distance threshold

- Regression coefficient as a function of distance **d**



City A



City B

Peak region of blue curve (co-visits by non-bridge customers) suggests geographical constraint for social bridge effect

Application: Prediction of co-visits

- Three-class classification: small, medium, large amount of co-visitation
- For each IV (feature), train on 20% of communities and test on the rest 80%, using LIBSVM (Chang, 2011)

Application: Prediction of co-visits

- Three-class classification: small, medium, large amount of co-visitation
- For each IV (feature), train on 20% of communities and test on the rest 80%, using LIBSVM (Chang, 2011)

Indicator	City A	City B
# Social Bridge	72.5%	53.6%
Population	55.0%	50.4%
Distance	49.9%	45.4%
Age	40.0%	36.6%
Gender	35.1%	40.9%
Marital Status	35.3%	34.7%
Education	38.1%	39.0%
Working Style	38.3%	38.0%
Income	33.6%	36.1%

Social bridge is more efficient in predicting co-visitation than traditional factors

Discussion

- Social bridge captures a form of social learning due to physical exposure: similar to "the familiar stranger" (Milgram, 1977)

Discussion

- Social bridge captures a form of social learning due to physical exposure: similar to "the familiar stranger" (Milgram, 1977)
- Bridge customers are conceptually similar to "structural hole spanners" (Lou, 2013)

Discussion

- Social bridge captures a form of social learning due to physical exposure: similar to "the familiar stranger" (Milgram, 1977)
- Bridge customers are conceptually similar to "structural hole spanners" (Lou, 2013)
- Easy to compute: as long as location information (social media, etc.) is available

Discussion

- Social bridge captures a form of social learning due to physical exposure: similar to "the familiar stranger" (Milgram, 1977)
- Bridge customers are conceptually similar to "structural hole spanners" (Lou, 2013)
- Easy to compute: as long as location information (social media, etc.) is available
- No causal relation, but tested against demographics and null model based on popularity and distance (Huff, 1964)

Discussion

- Social bridge captures a form of social learning due to physical exposure: similar to "the familiar stranger" (Milgram, 1977)
- Bridge customers are conceptually similar to "structural hole spanners" (Lou, 2013)
- Easy to compute: as long as location information (social media, etc.) is available
- No causal relation, but tested against demographics and null model based on popularity and distance (Huff, 1964)
- Strong correlation can lead to applications such as behavior prediction and stratification, campaign targeting, and resource allocation