Urban Segregation in Behaviour

A Data-Driven Approach

Xiaowen Dong

Department of Engineering Science

University of Oxford

Institute for Analytical Sociology

Norrköping, September 2019

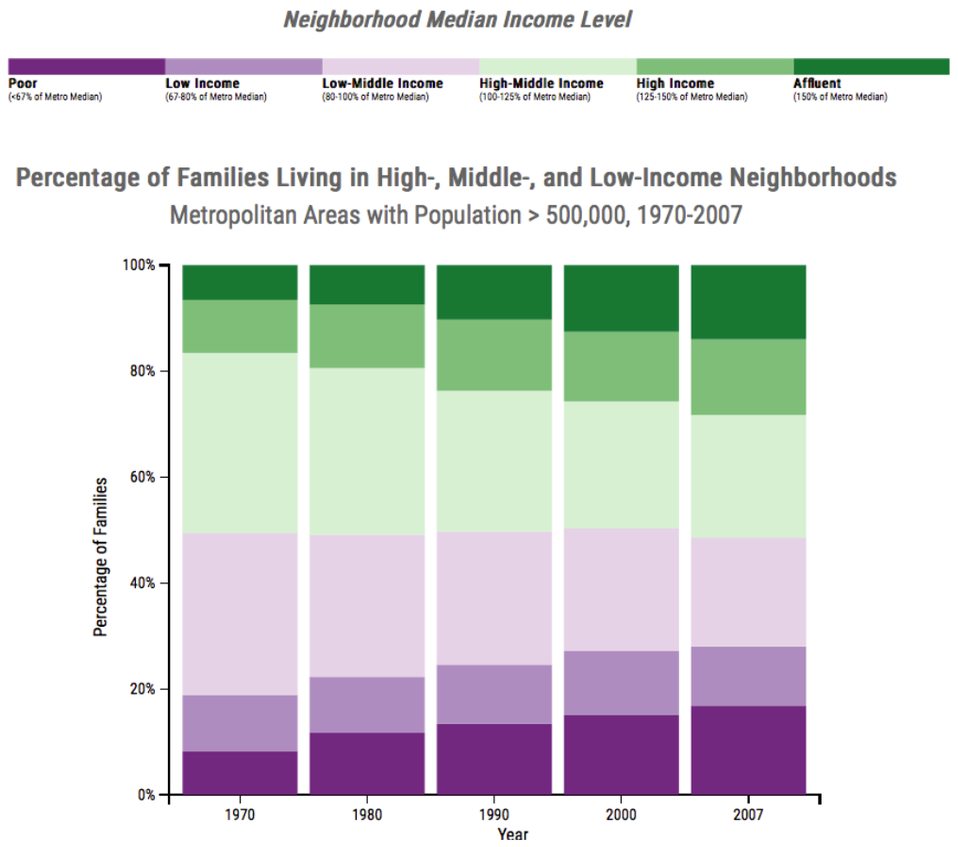
Background

- residential segregation challenges societies across the globe

Rosenfeld, "Some notes on the effects of residential segregation, and spatial isolation," 2013.

Background

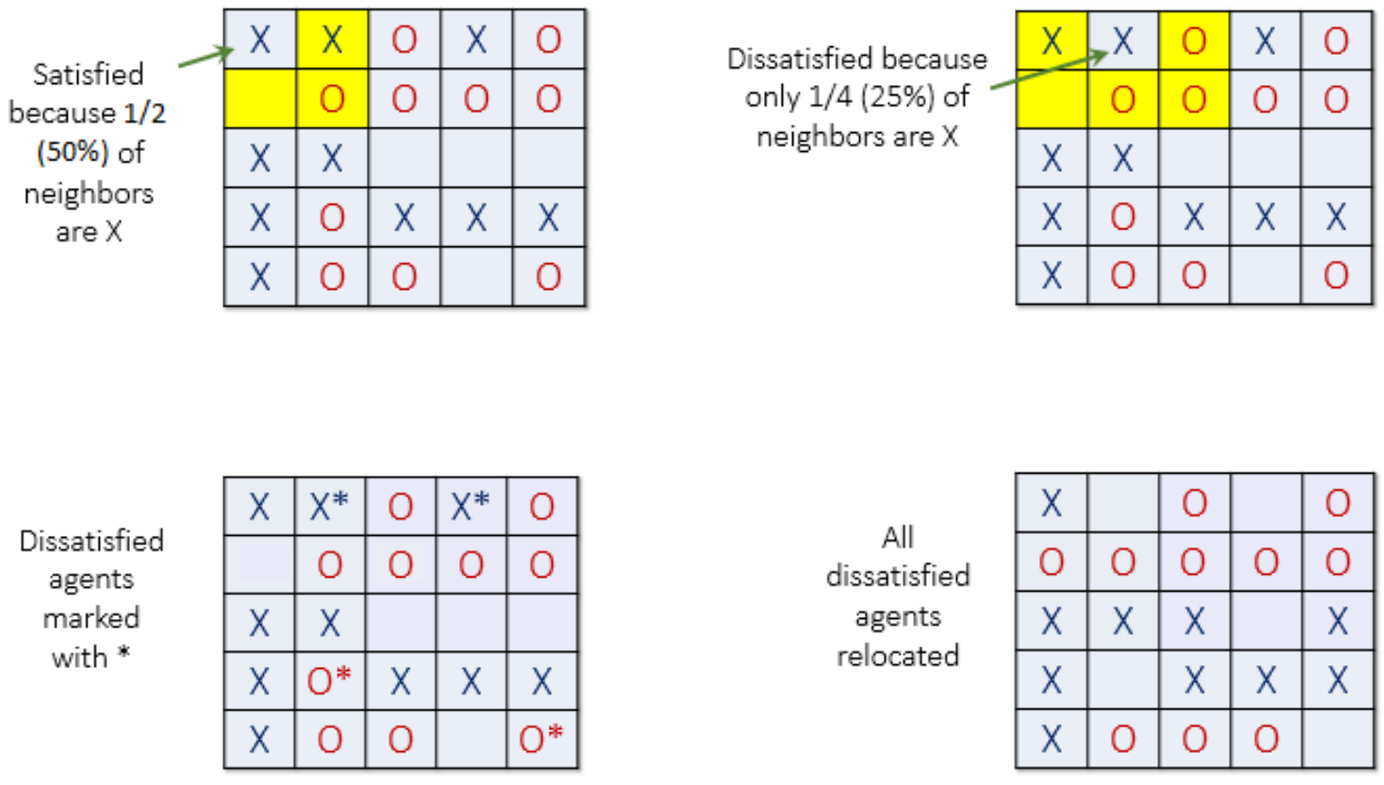
- residential segregation challenges societies across the globe
- income segregation has been rising in the US in the last 40-50 years



Rosenfeld, "Some notes on the effects of residential segregation, and spatial isolation," 2013.
Bischoff and Reardon, "Residential segregation by income, 1970-2009," US2010 Project, 2013.

Background

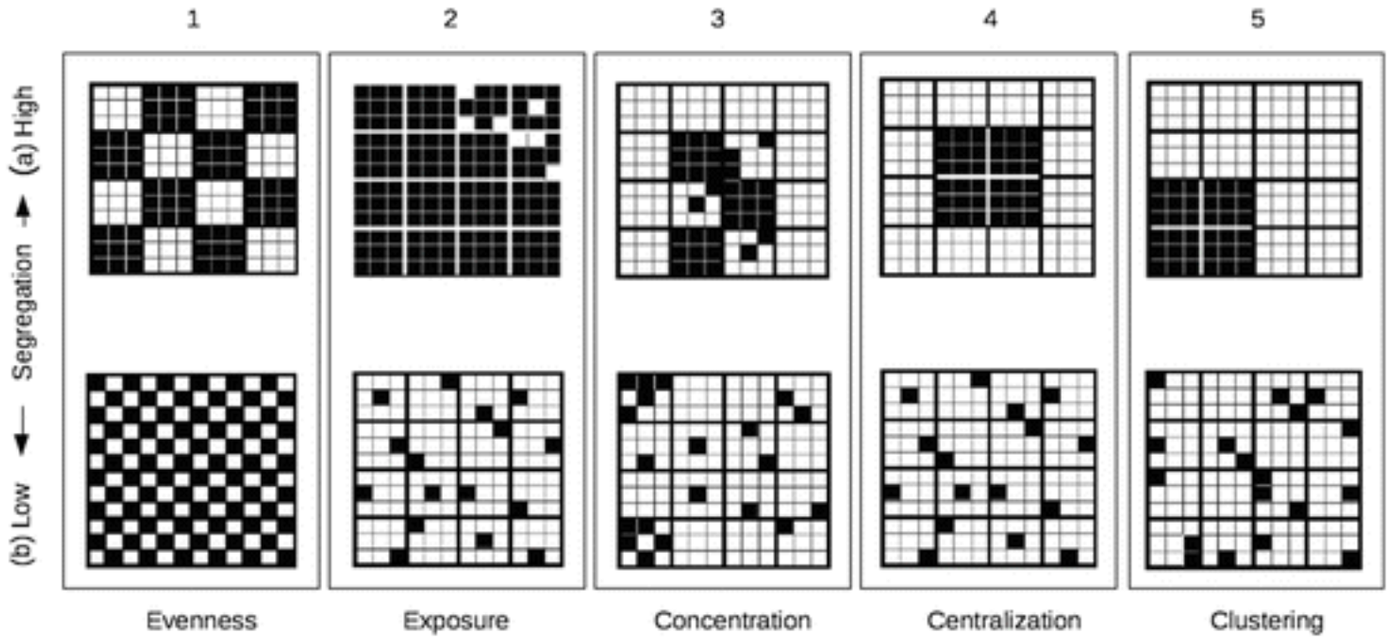
- Classical studies on residential segregation
  - Schelling's dynamic model of segregation [Schelling71]



Schelling, "Dynamic models of segregation," The Journal of Mathematical Sociology, 1971.
McCown, "Schelling's Model of Segregation," 2014.

Background

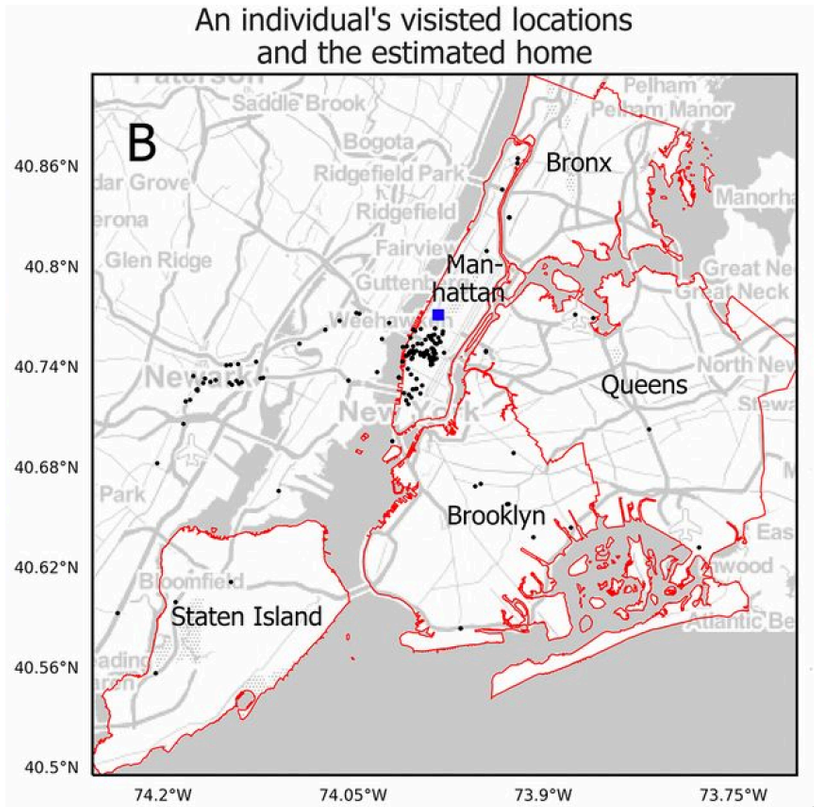
- Classical studies on residential segregation
  - Schelling's dynamic model of segregation [Schelling71]
  - five key dimensions and twenty indexes in American census reports [Iceland02]

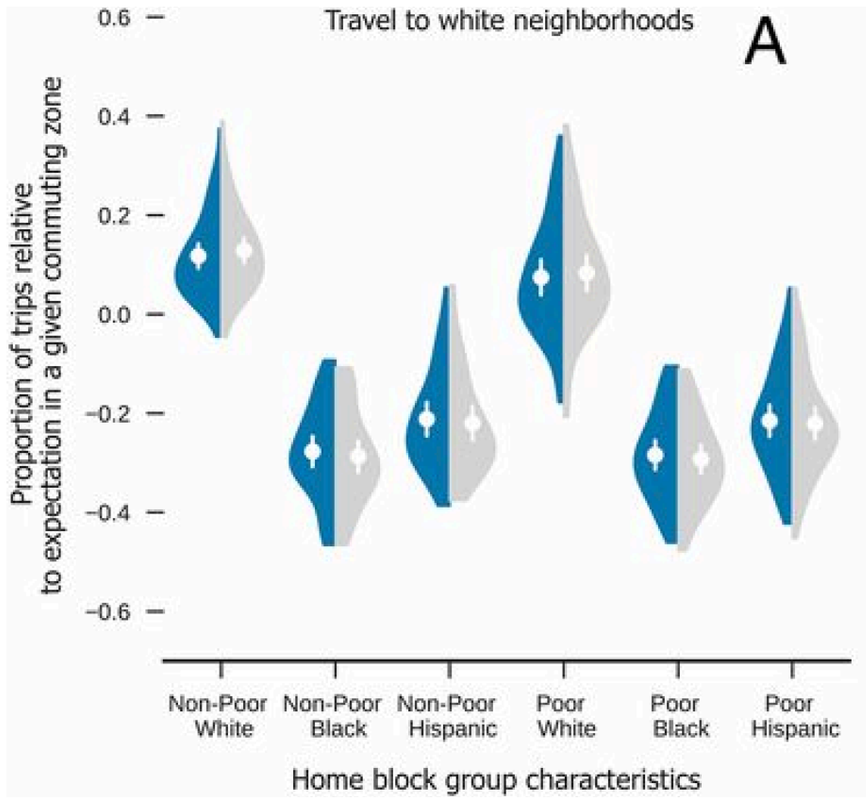


Iceland et al., "Racial and ethnic residential segregation in the United States: 1980-2000," 2002.
Rodriguez-Moral and Vorsatz, "An overview of the measurement of segregation," Complex Networks and Dynamics, 2016.

Background

- Recent studies on selective exposure in urban areas

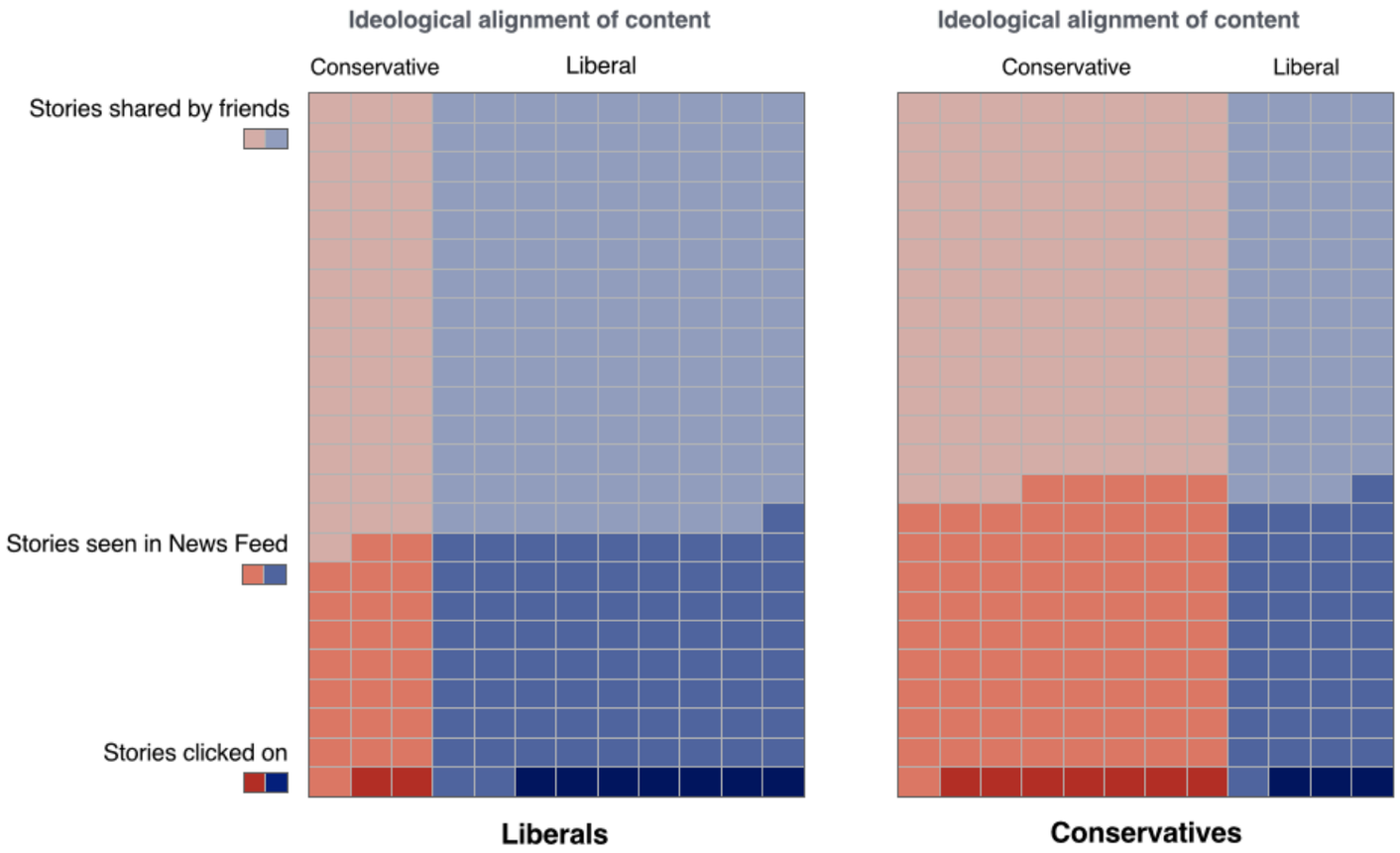




Wang et al., "Urban mobility and neighborhood isolation in America's 50 largest cities," PNAS, 2018.

Background

- Recent studies on selective exposure in social media



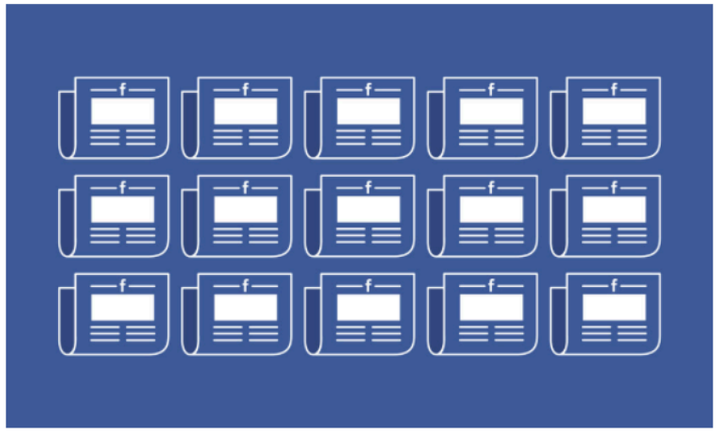
Messing, "Ideologically diverse news, an agenda for future research," 2015.

Background

- Recent studies on selective exposure in social media

BRIAN BARRETT CULTURE 06.29.16 5:50 PM

YOUR FACEBOOK ECHO CHAMBER JUST GOT A WHOLE LOT LOUDER



THEN ONE/WIRED

"We are updating News Feed over the coming weeks so that the things posted by the friends you care about are higher up in your News Feed," Facebook engineering director Lars Backstrom wrote. That sounds simple enough, but what it really means is the feed will promote content from your friends over content from publishers.

Barrett, "Your Facebook echo chamber just got a whole lot louder," 2016.

In this talk

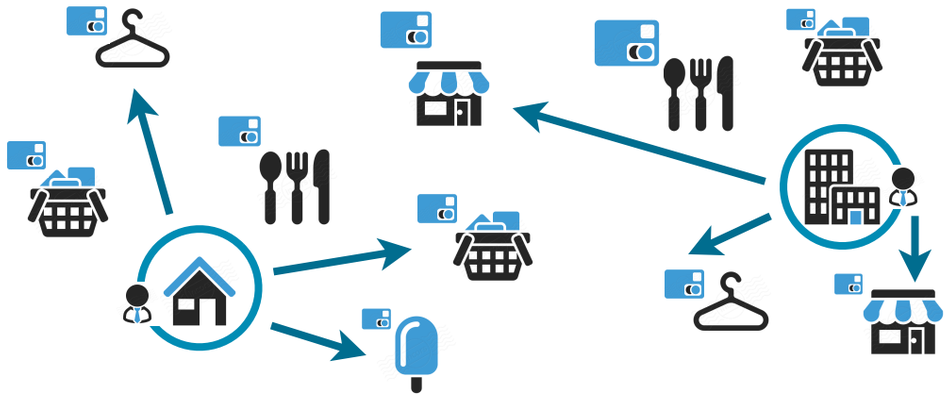
- Limitation of literature
  - focused on residential pattern using static census data
  - focused on macro-level (neighbourhood-level) segregation
  - focused on measurement or quantification

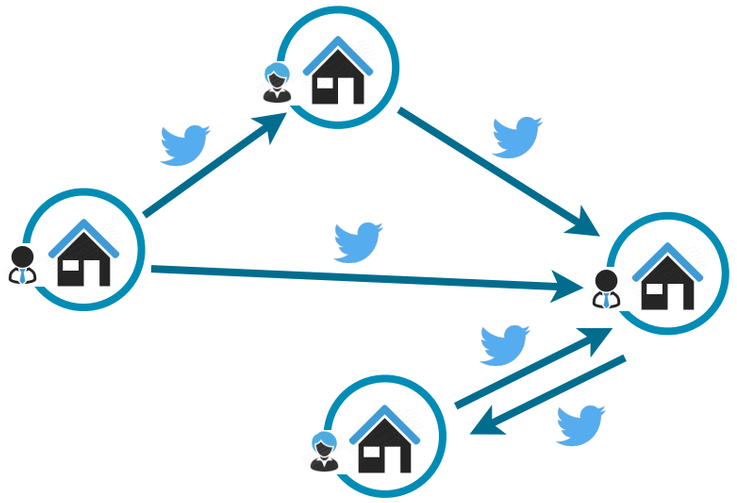
In this talk

- Limitation of literature
  - focused on residential pattern using static census data
  - focused on macro-level (neighbourhood-level) segregation
  - focused on measurement or quantification
- Our objective
  - from residential segregation to behaviour segregation
  - from macro-level to micro-level analysis
  - from quantifying to modelling segregation

Data sets

- Anonymised behavioural data of residents of three metropolitan areas in three months
  - credit card transactions (85K users in European, 200K in Latin American)
  - Twitter mentions (1M users in European, 260K in Latin American, 440K in Northern American)



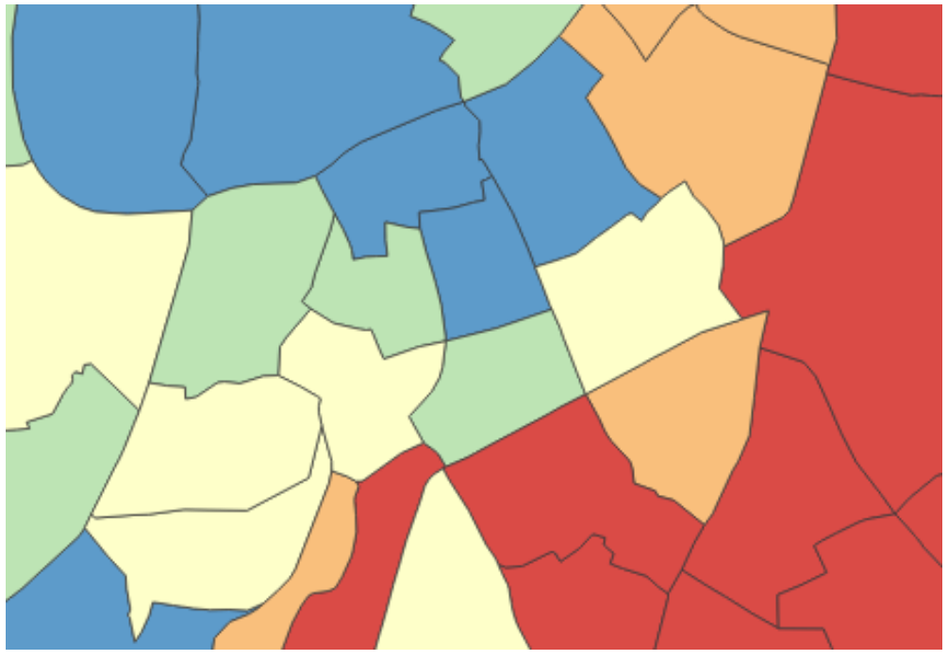


Data sets

- Analysis at the level of administrative neighbourhoods in the city

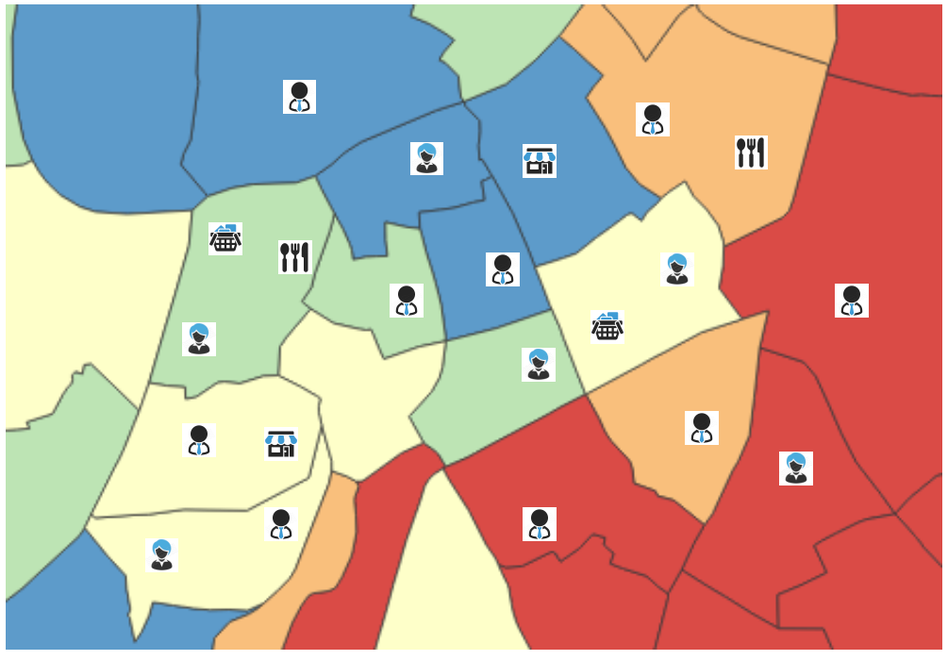
Data sets

- Analysis at the level of administrative neighbourhoods in the city
  - obtain neighbourhood-level socio-economic status from national census



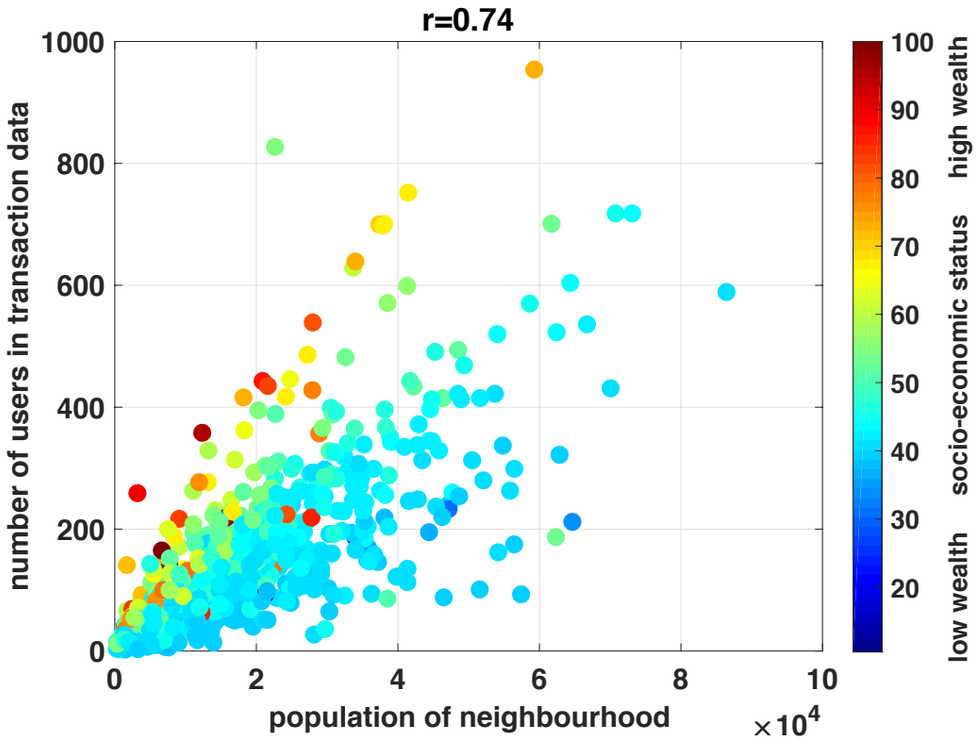
Data sets

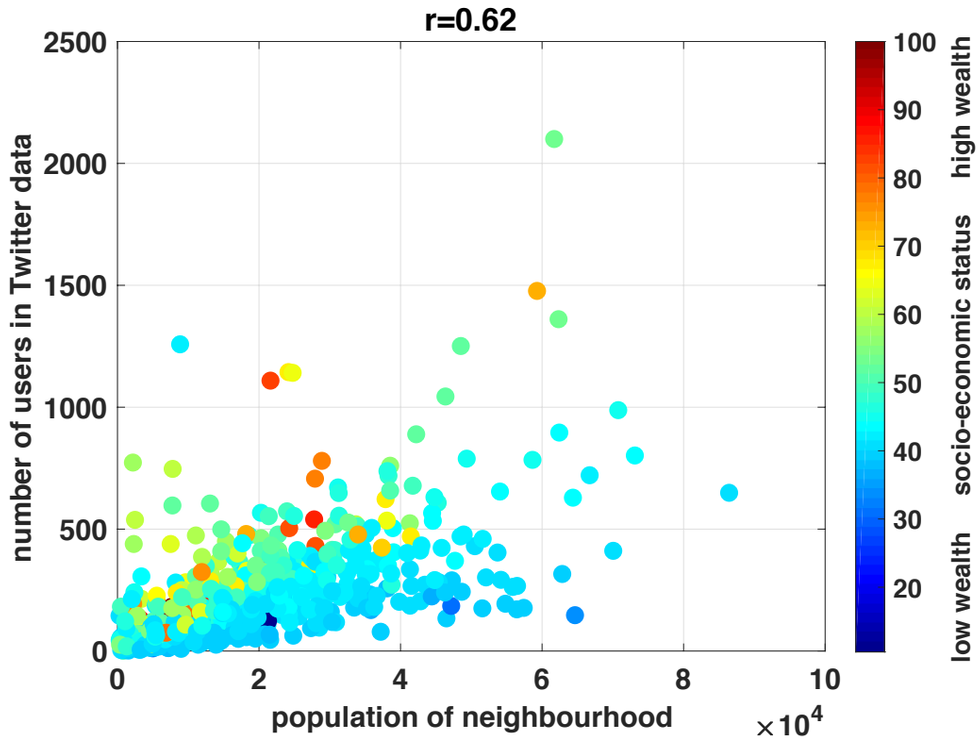
- Analysis at the level of administrative neighbourhoods in the city
  - obtain neighbourhood-level socio-economic status from national census
  - associate each user with a neighbourhood



Data sets

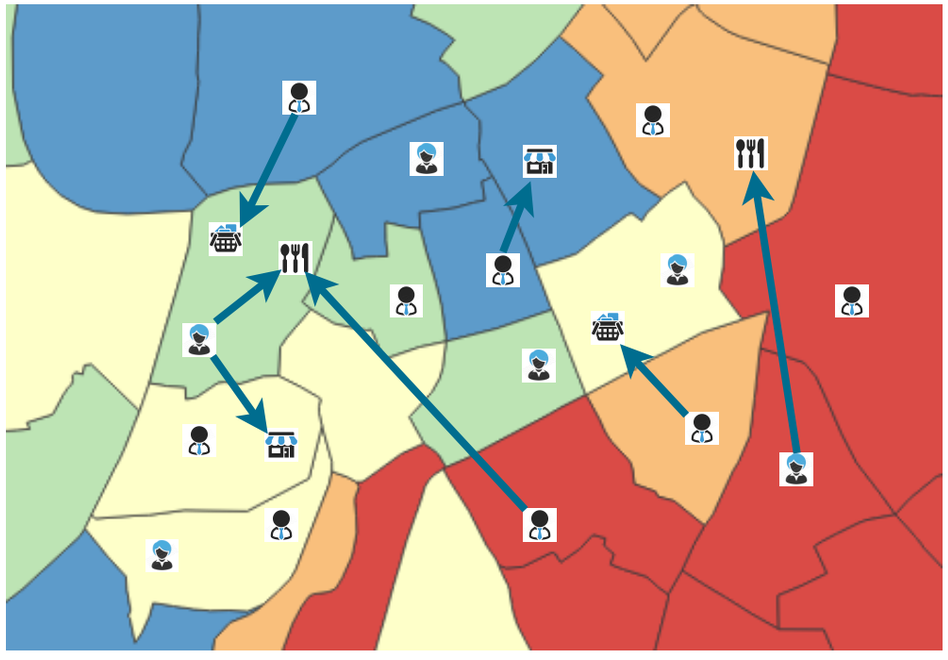
- Neighbourhood-level statistics in European metropolitan area

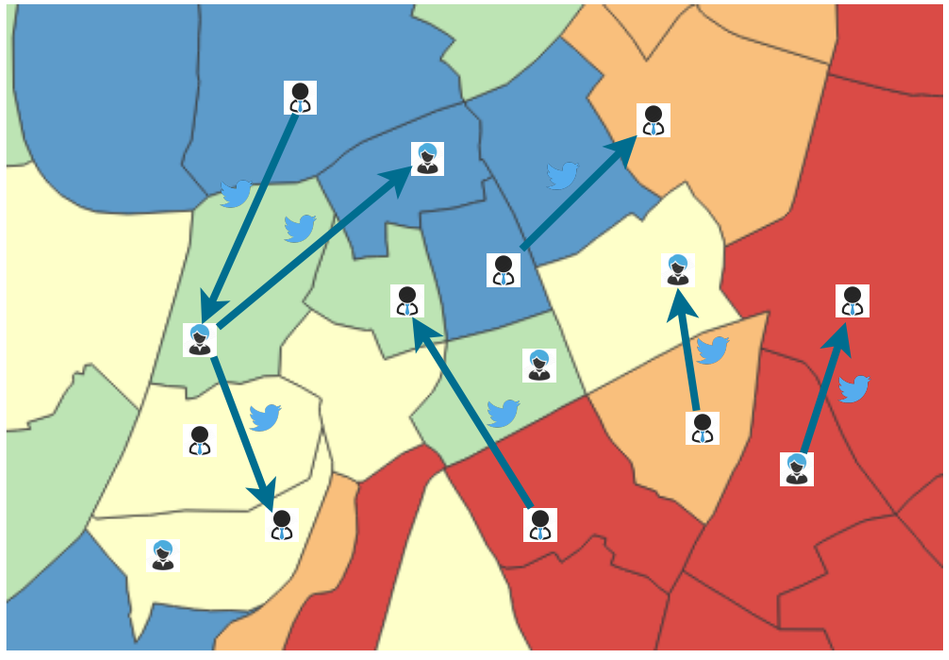




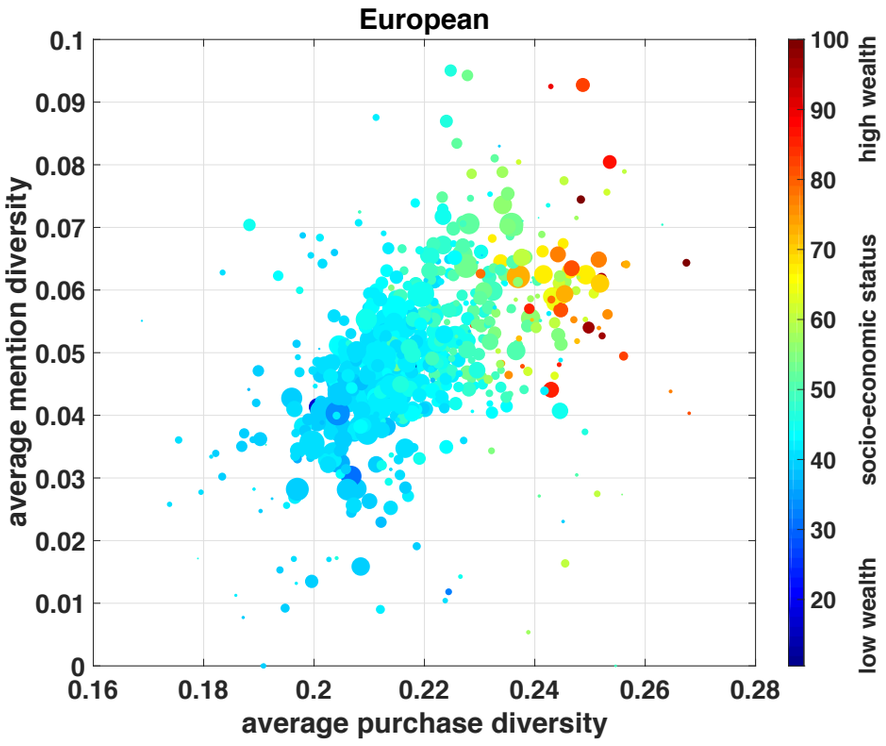
Interaction networks

- Nodes: neighbourhoods
- Edges: interaction
  - purchase: number of purchases made by users in neighbourhood i at stores in j
  - Twitter: number of mentions made by users in neighbourhood i to users in j





Behavioural diversity

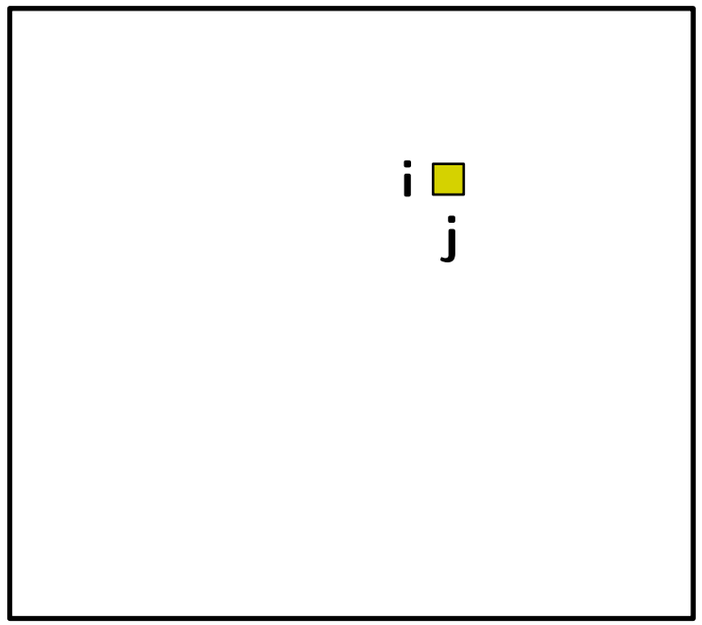


- exploration in physical space is associated with that in social space
- diversity is associated with neighbourhood social-economic status [Eagle10]

Eagle et al., "Network diversity and economic development," Science, 2010.

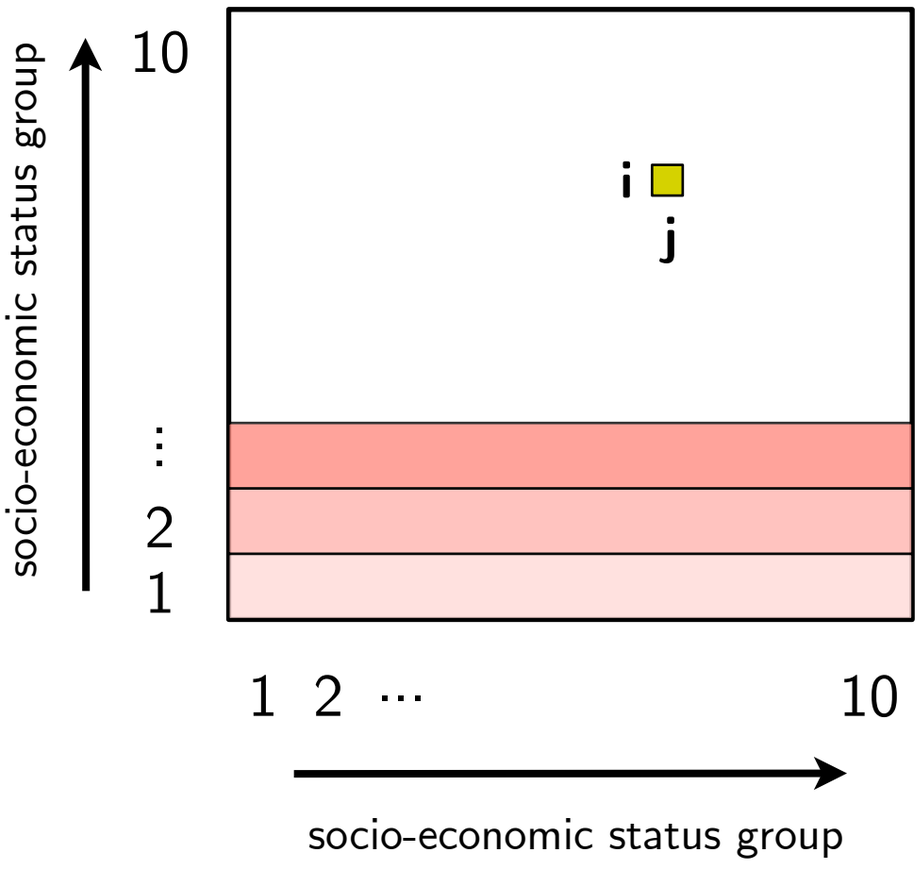
Measuring segregation

- Construct mixing matrices



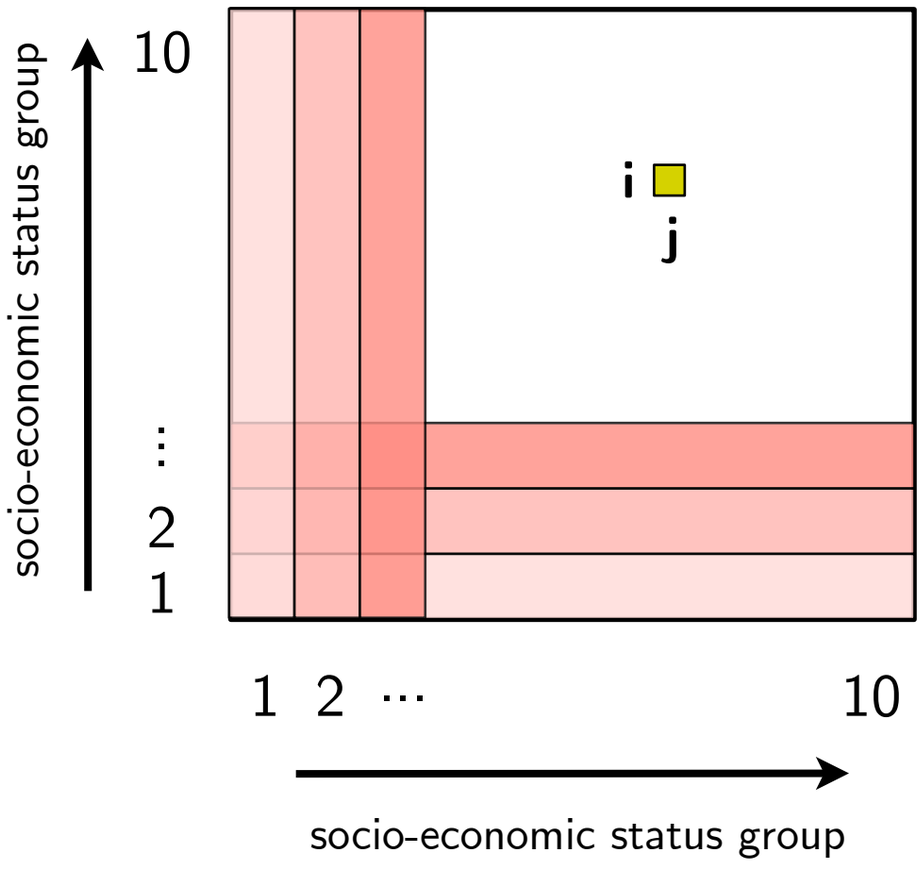
Measuring segregation

- Construct mixing matrices



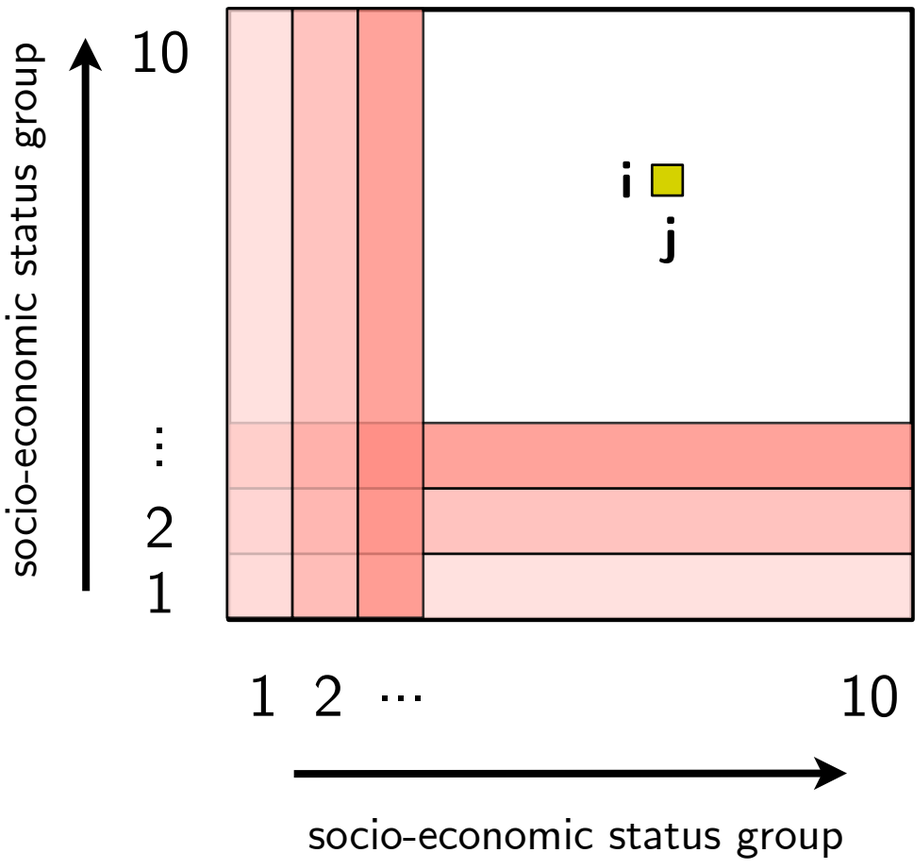
Measuring segregation

- Construct mixing matrices

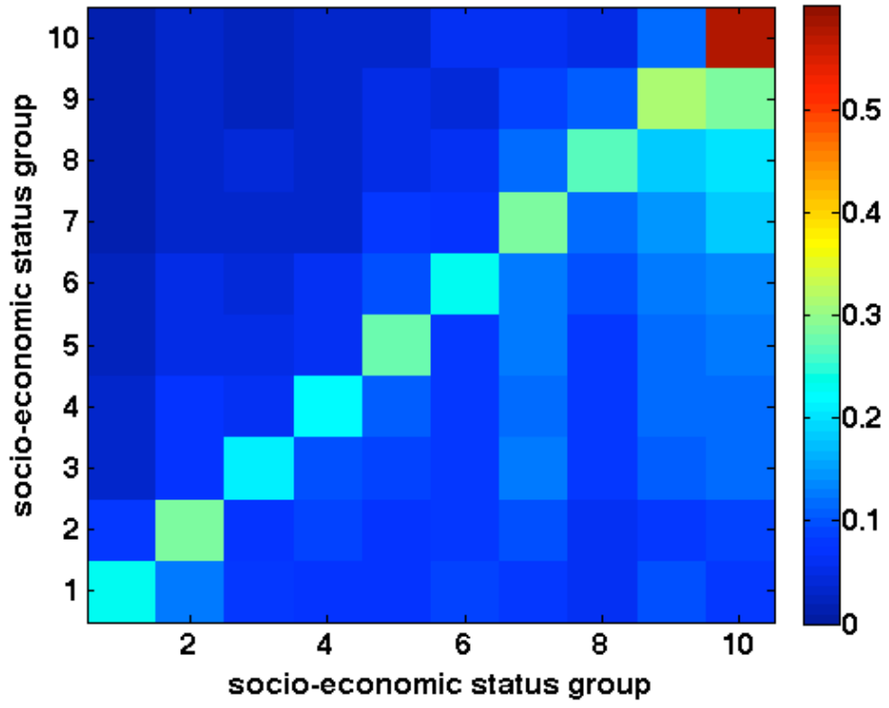


Measuring segregation

- Construct mixing matrices



aggregate



Measuring segregation

group 3	0.15	0.05	0.1
group 2	0.05	0.15	0.08
group 1	0.2	0.1	0.12
	group 1	group 2	group 3

Newman, "Mixing patterns in networks," Physical Review E, 2003.

Measuring segregation

b_y	0.4	0.3	0.3	
group 3	0.15	0.05	0.1	0.3
group 2	0.05	0.15	0.08	0.28
group 1	0.2	0.1	0.12	0.42
	group 1	group 2	group 3	a_x

$$\sum_{xy} e_{xy} = 1$$

$$\sum_{y} e_{xy} = a_x$$

$$\sum_{x} e_{xy} = b_y$$

Newman, "Mixing patterns in networks," Physical Review E, 2003.

Measuring segregation

b_y	0.4	0.3	0.3	
group 3	0.15	0.05	0.1	0.3
group 2	0.05	0.15	0.08	0.28
group 1	0.2	0.1	0.12	0.42
	group 1	group 2	group 3	a_x

$$\sum_{xy} e_{xy} = 1$$

$$\sum_{y} e_{xy} = a_x$$

$$\sum_{x} e_{xy} = b_y$$

segregation as mixing coefficient

$$m = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_x \sigma_y} = 0.03$$

$$\sigma_x^2 = \sum_x x^2 a_x - (\sum_x x a_x)^2$$

Newman, "Mixing patterns in networks," Physical Review E, 2003.

Measuring segregation

b_y	0.4	0.3	0.3	
group 3	0.15	0.05	0.1	0.3
group 2	0.05	0.15	0.08	0.28
group 1	0.2	0.1	0.12	0.42
	group 1	group 2	group 3	a_x

$$\sum_{xy} e_{xy} = 1$$

$$\sum_{y} e_{xy} = a_x$$

$$\sum_{x} e_{xy} = b_y$$

segregation as mixing coefficient

$$m = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_x \sigma_y} = 0.03$$

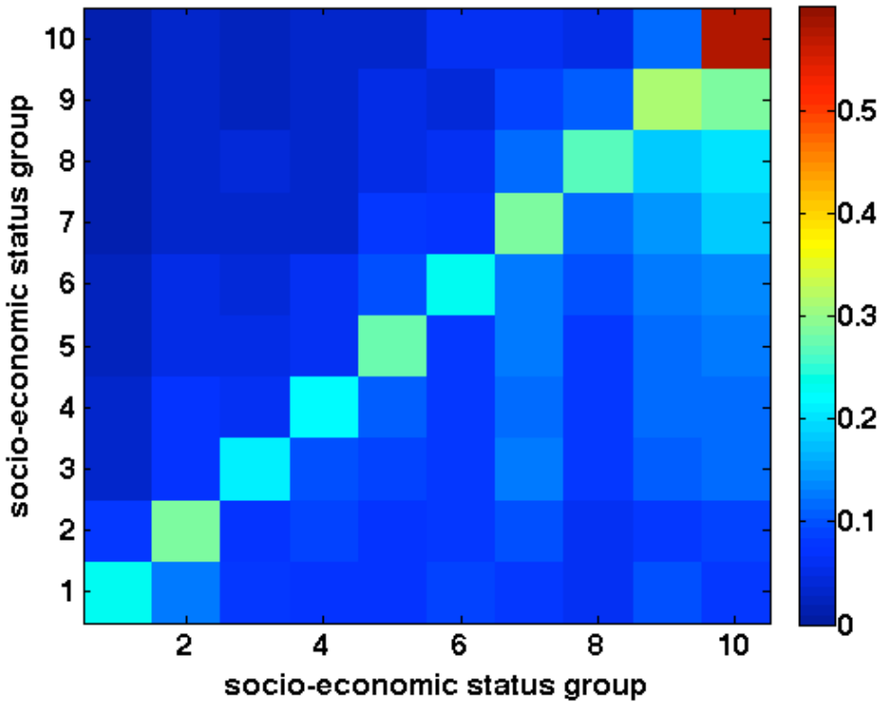
$$\sigma_x^2 = \sum_x x^2 a_x - (\sum_x x a_x)^2$$

remarks

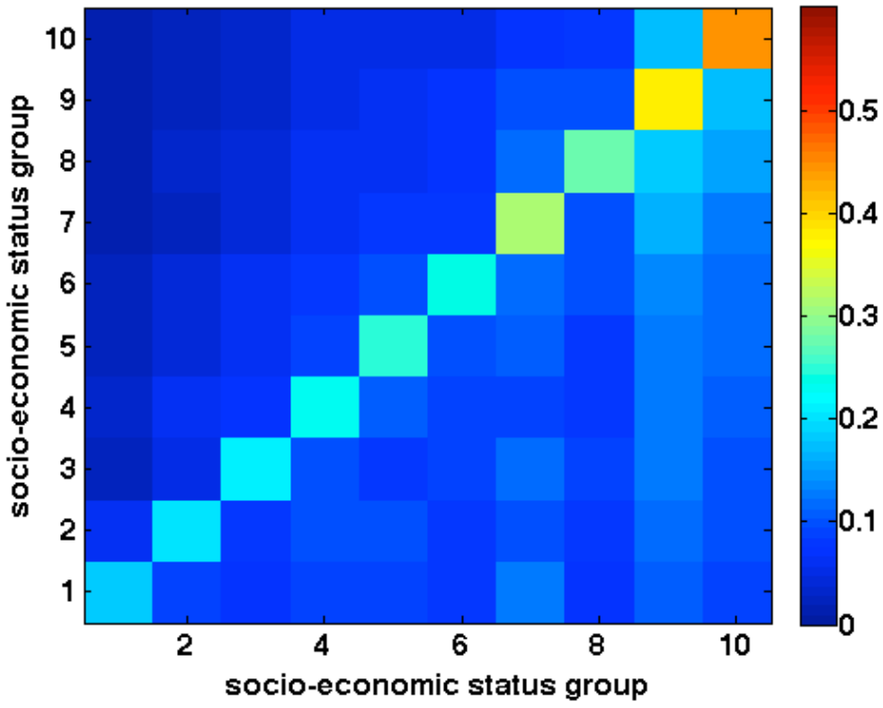
- sample Pearson correlation coefficient
- $-1 \leq m \leq 1$ (from perfect disassortativity to perfect assortativity)

Newman, "Mixing patterns in networks," Physical Review E, 2003.

Segregation in offline & online interaction

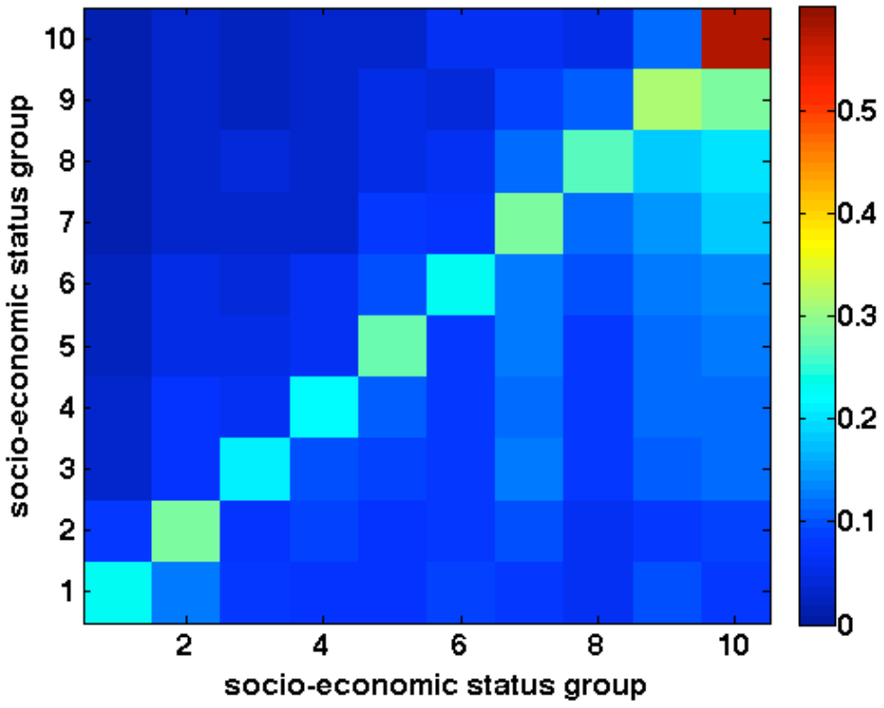


Purchase ($m = 0.43$ **)**

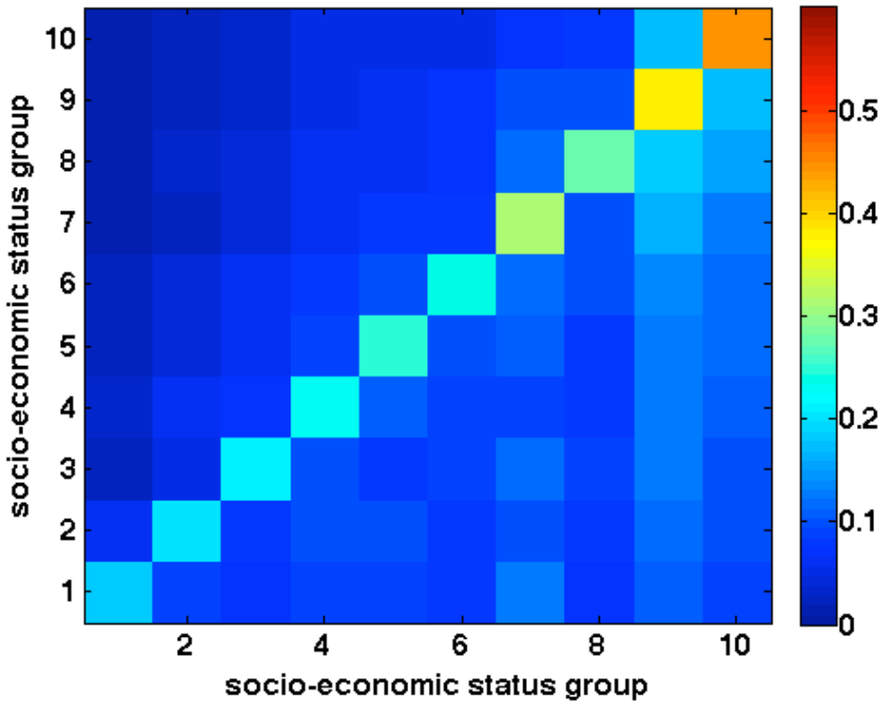


Twitter ($m = 0.35$ **)**

Segregation in offline & online interaction



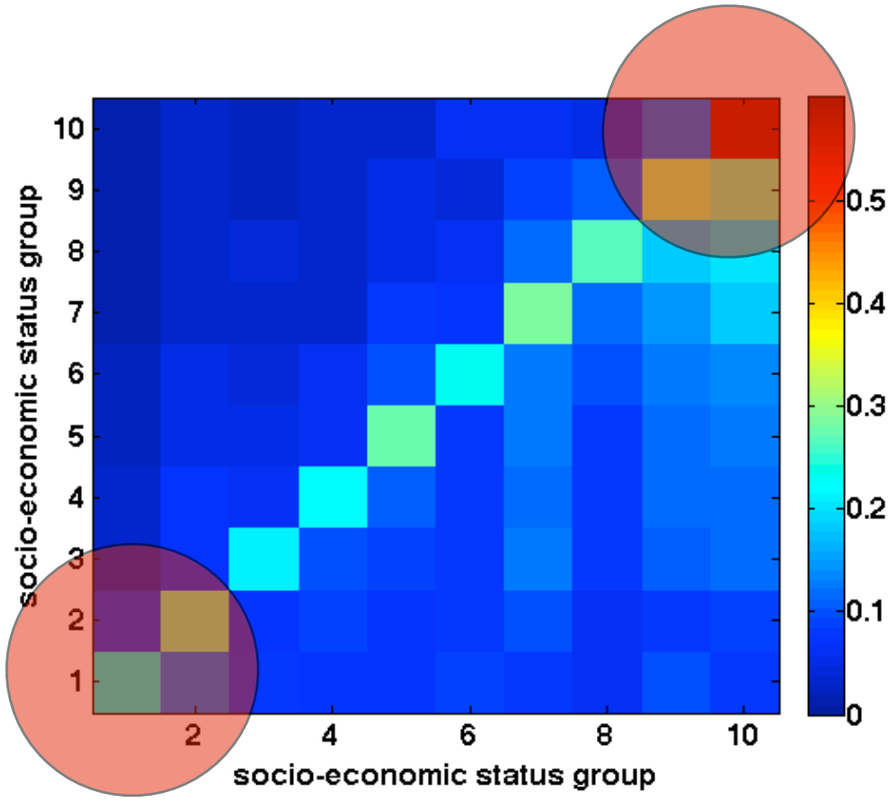
Purchase ($m = 0.43$)



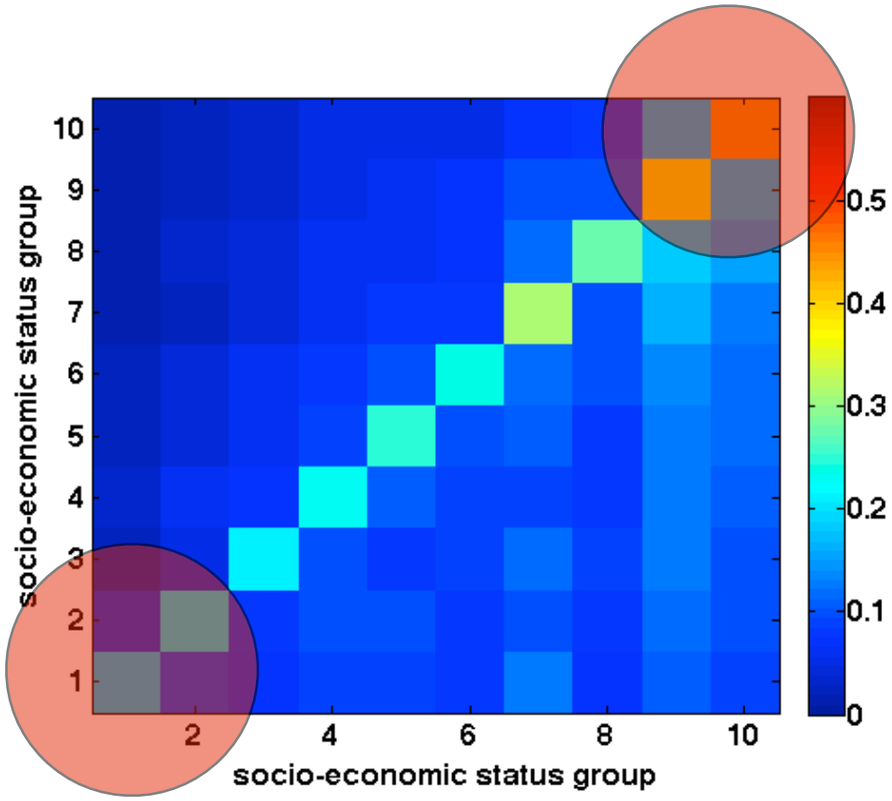
Twitter ($m = 0.35$)

- is it due to expected interaction or wealth distribution imposed by geography?

Segregation in offline & online interaction



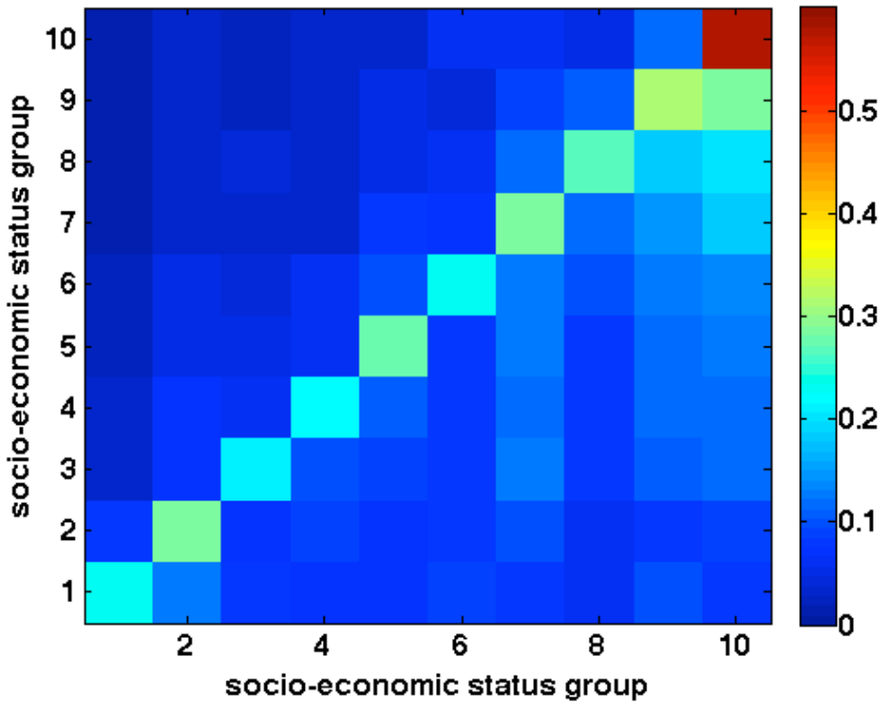
Purchase ($m = 0.43$)



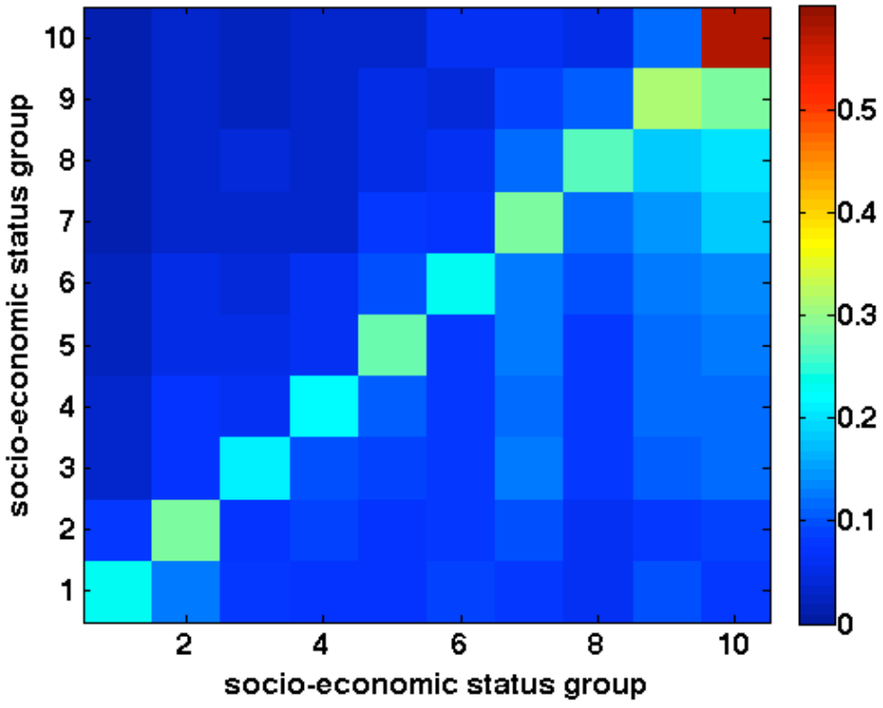
Twitter ($m = 0.35$)

- is it due to expected interaction or wealth distribution imposed by geography?
- does segregation vary across socio-economic status groups?

Two null models of interaction



Two null models of interaction



simulation via gravity model

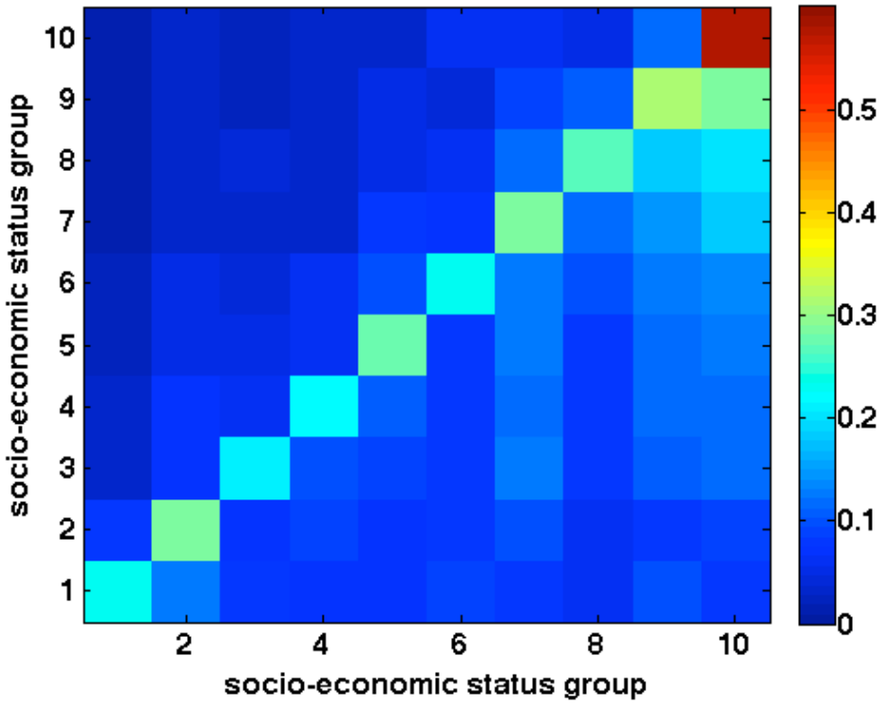
- fit gravity model to empirical weights [Krings09]

$$\hat{w}_{ij} \approx \frac{{M_1}^{\beta_1} {M_2}^{\beta_2}}{(d_{ij} + \epsilon)^{\alpha}}$$

- expected interaction due to segregated distribution of residential households

Krings et al., "Urban gravity: A model for inter-city telecommunication flows," JSTAT, 2009.

Two null models of interaction



simulation via gravity model

- fit gravity model to empirical weights [Krings09]

$$\hat{w}_{ij} \approx \frac{{M_1}^{\beta_1} {M_2}^{\beta_2}}{(d_{ij} + \epsilon)^{\alpha}}$$

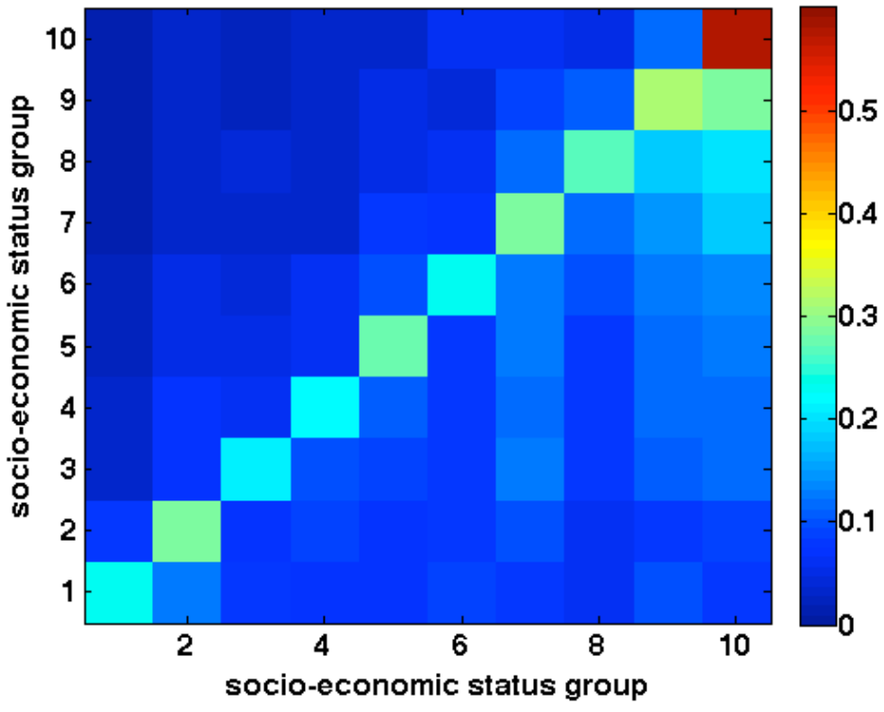
- expected interaction due to segregated distribution of residential households

shuffling socio-economic status

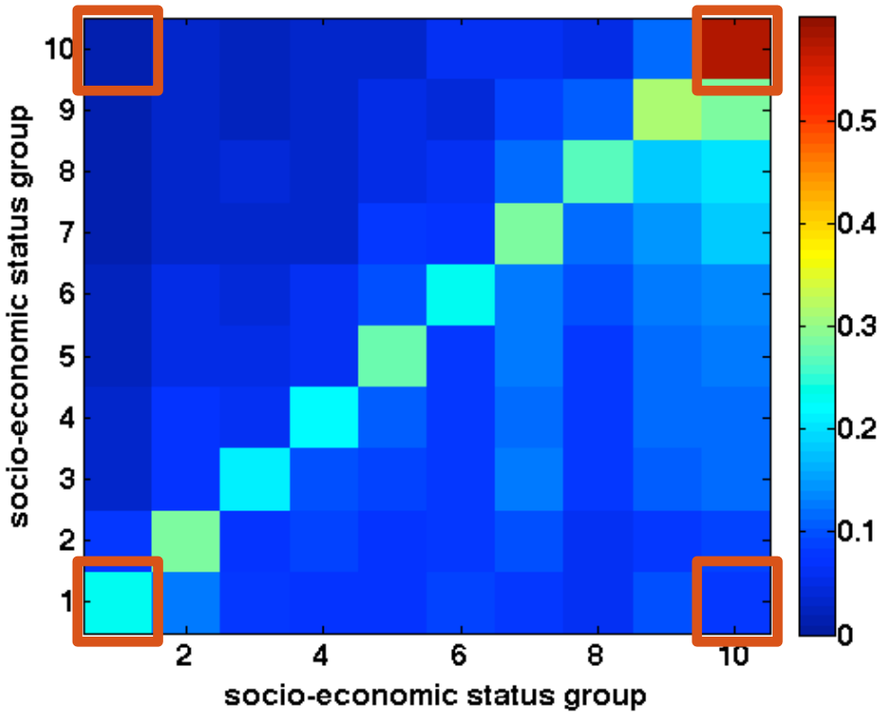
- randomly shuffle socio-economic status of neighbourhoods
- interaction due to shopping or talking to people in own neighbourhoods

Krings et al., "Urban gravity: A model for inter-city telecommunication flows," JSTAT, 2009.

Segregation w.r.t. socio-economic “distance”



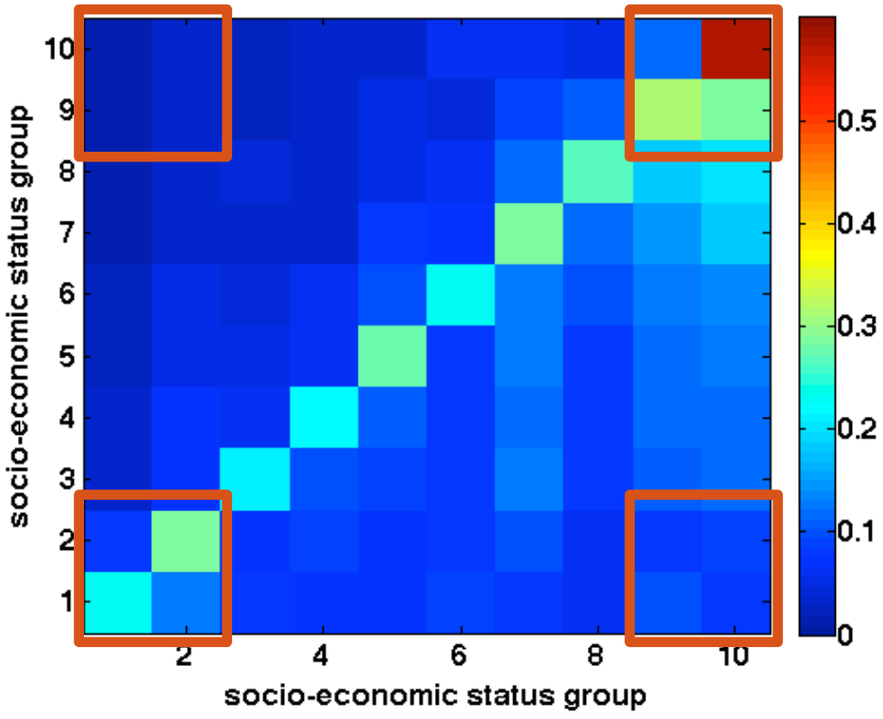
Segregation w.r.t. socio-economic "distance"



10% extreme each side

- construct sub-network from extreme socio-economic groups
- calculate assortative mixing for sub-network

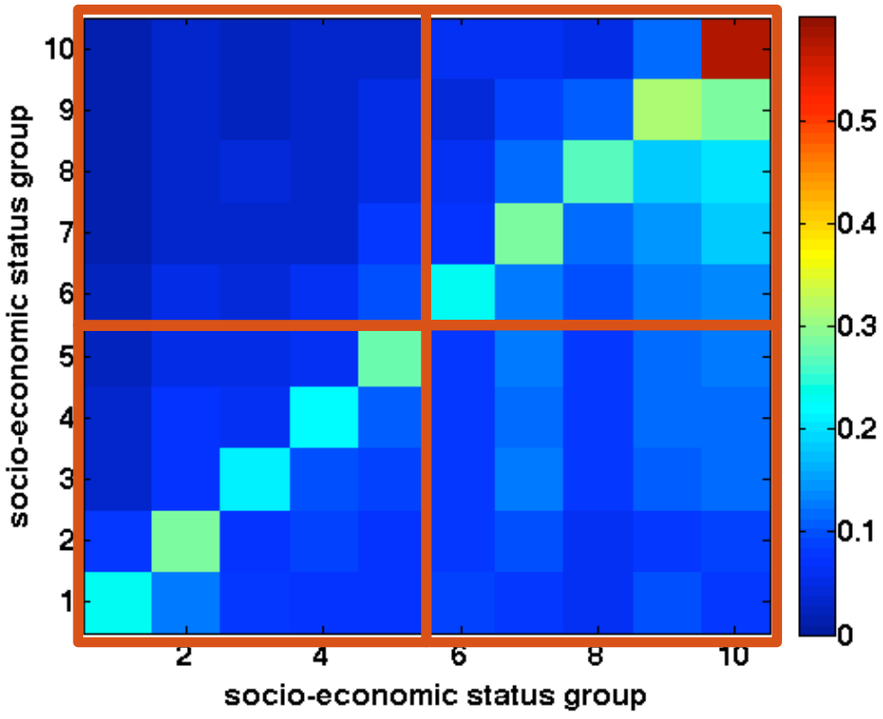
Segregation w.r.t. socio-economic "distance"



20% extreme each side

- construct sub-network from extreme socio-economic groups
- calculate assortative mixing for sub-network
- repeat the calculation with more groups

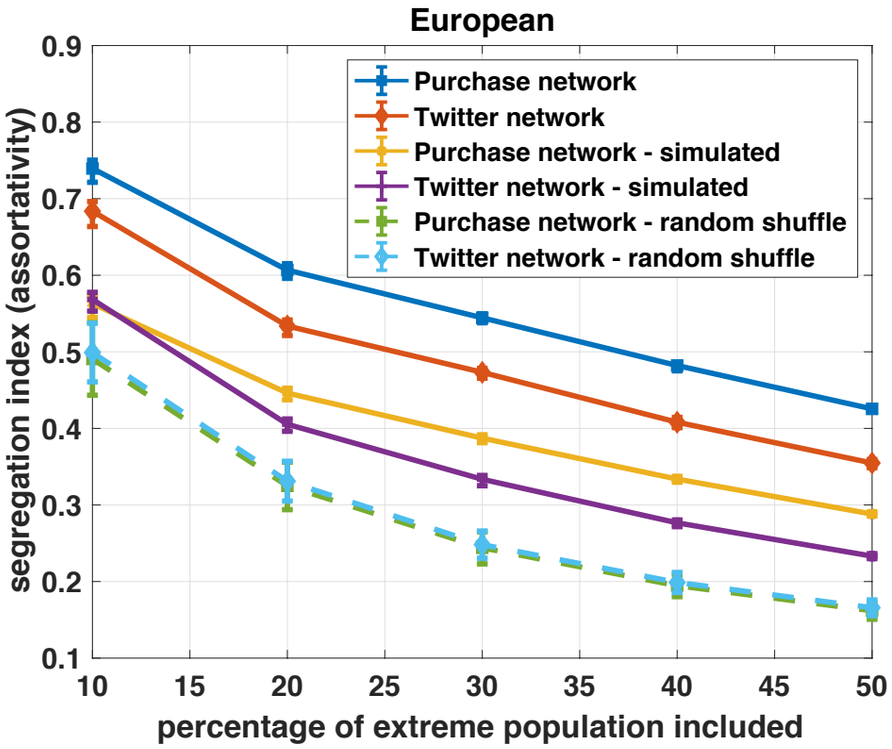
Segregation w.r.t. socio-economic "distance"



50% extreme each side
(complete data set)

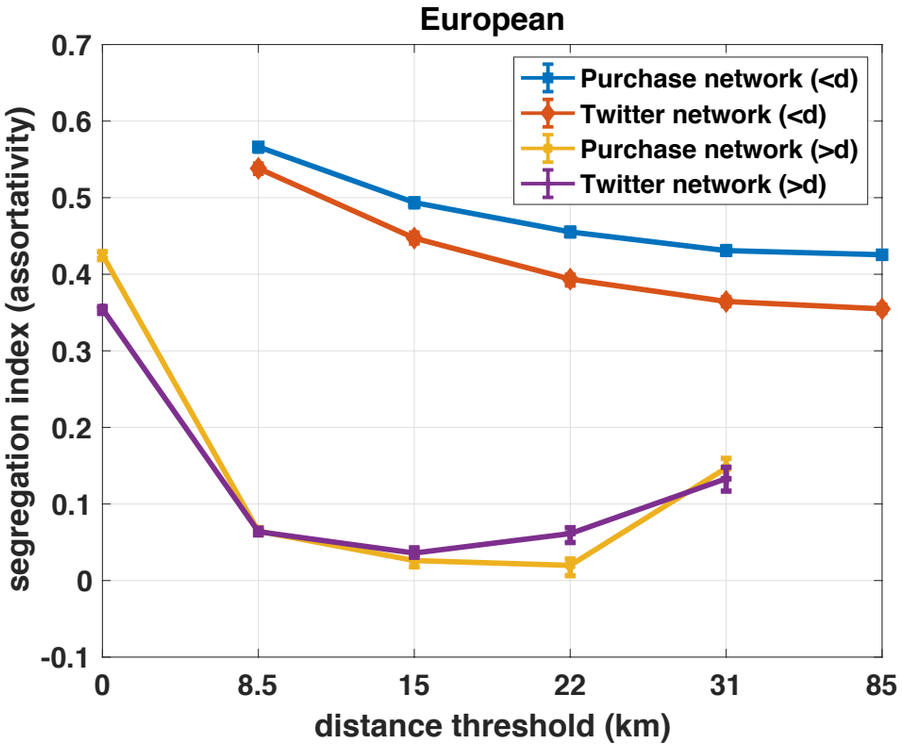
- construct sub-network from extreme socio-economic groups
- calculate assortative mixing for sub-network
- repeat the calculation with more groups

Segregation w.r.t. socio-economic "distance"



- segregation in interaction is stronger than that imposed by geography
- it is also stronger than that due to within-neighbourhood interaction
- segregation is most pronounced between highest/lowest SES groups
- middle SES groups serve as "social bridges"

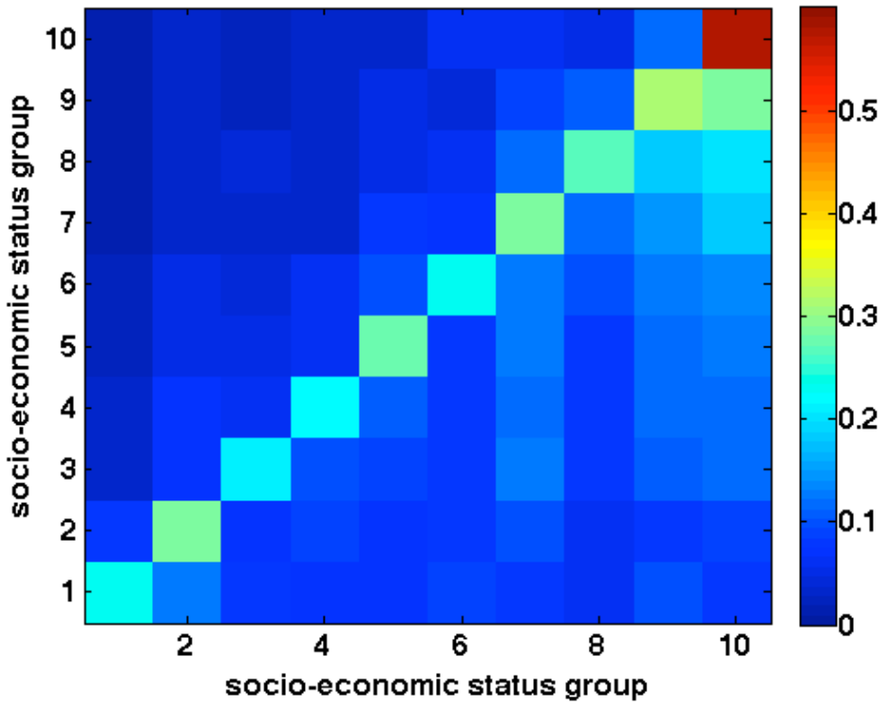
Segregation w.r.t. geographical distance



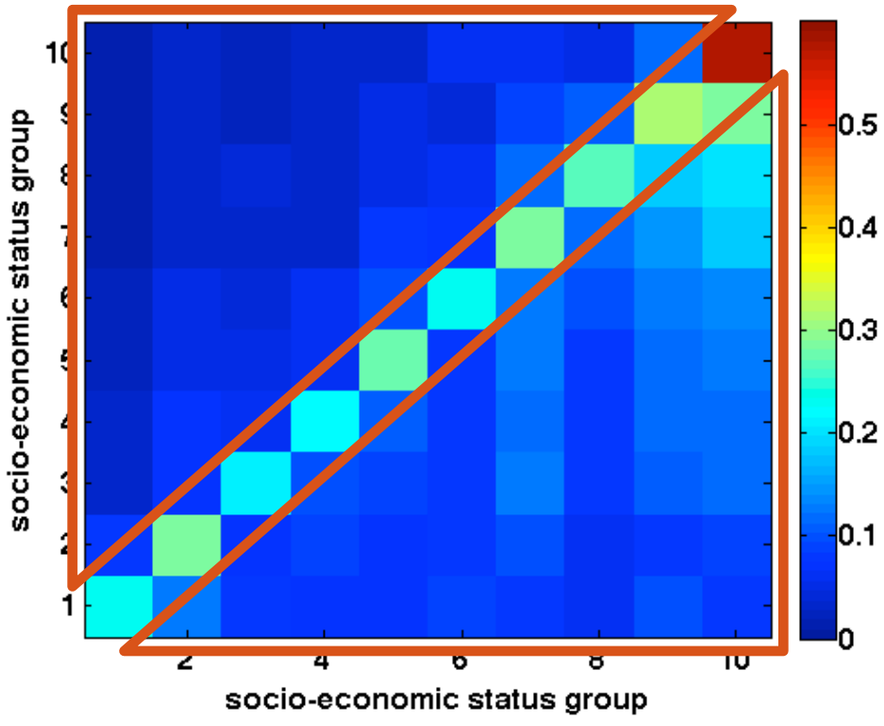
- segregation is largely due to short-distance interaction
  - purchase: less costly in terms of time and money
  - online: interaction of local social groups [Bastos17]

Bastos et al., "The spatial dimension of online echo chambers," arXiv, 2017.

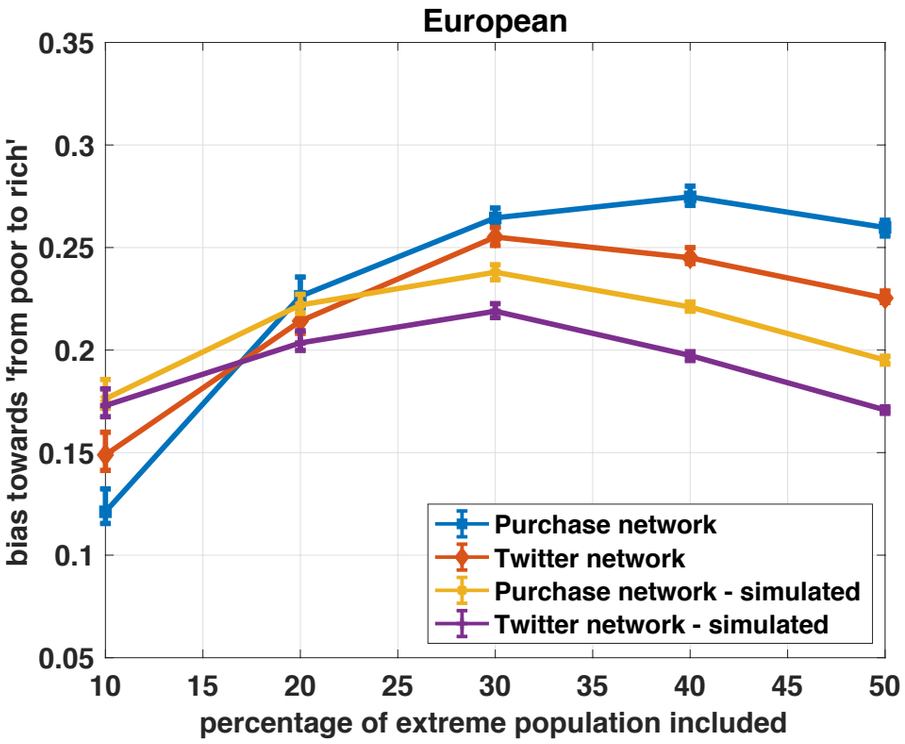
Asymmetry in interaction



Asymmetry in interaction

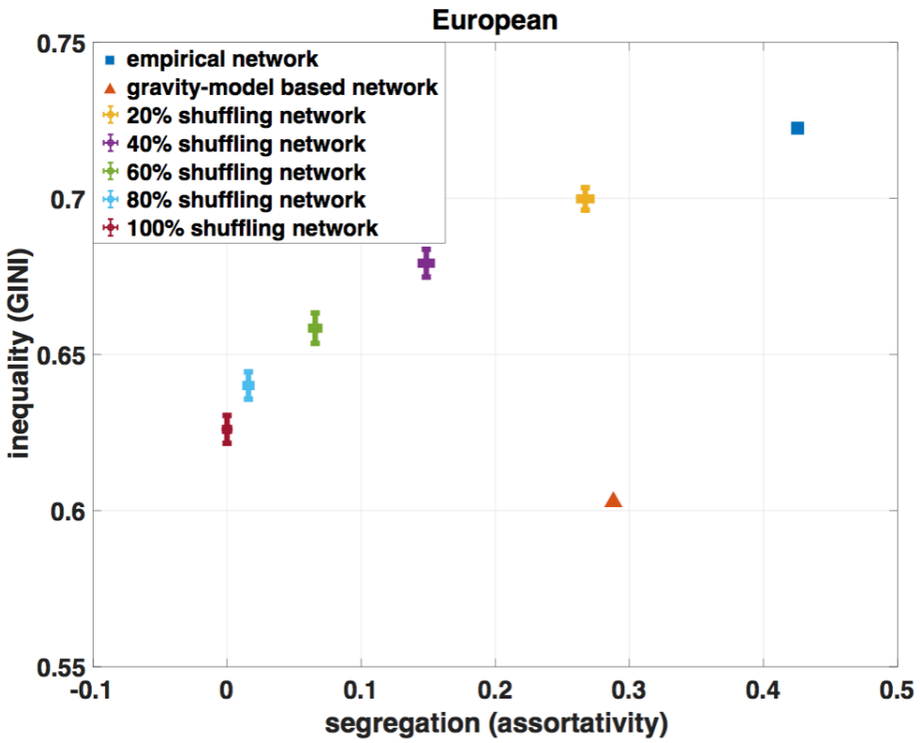


Asymmetry in interaction



- poorer areas interact more with wealthier ones than opposite
  - between extreme SES groups: less bias due to lack of interaction
  - with more SES groups: increasing asymmetry in interaction
- this asymmetry cannot be simply attributed to geography

Segregation and inequality



- there is evidence that segregation is linked to inequality [Cortright16]
- measure neighbourhood inequality via GINI coefficient of sales revenue of neighbourhoods [Louail17]
- segregated interaction is associated with uneven distribution of money

Cortright, "How racial segregation leads to income inequality," 2016.
Louail et al., "Crowdsourcing the Robin Hood effect in cities," Applied Network Science, 2017.

Discussion

- Summary & Implication
  - segregation in offline & online interaction is stronger than expected by geography

Discussion

- Summary & Implication
  - segregation in offline & online interaction is stronger than expected by geography
  - mitigation of segregation may focus on encouraging physical and social interaction (not merely on exposure [Bail18])

Discussion

- Summary & Implication
  - segregation in offline & online interaction is stronger than expected by geography
  - mitigation of segregation may focus on encouraging physical and social interaction (not merely on exposure [Bail18])
  - interaction across segments of society may be promoted by middle SES groups given their bridging role

Discussion

- Summary & Implication
  - segregation in offline & online interaction is stronger than expected by geography
  - mitigation of segregation may focus on encouraging physical and social interaction (not merely on exposure [Bail18])
  - interaction across segments of society may be promoted by middle SES groups given their bridging role
  - asymmetry in interaction may suggest existence of social hierarchy (wealthier areas attract unequal amount of money and information)

Discussion

- Summary & Implication
  - segregation in offline & online interaction is stronger than expected by geography
  - mitigation of segregation may focus on encouraging physical and social interaction (not merely on exposure [Bail18])
  - interaction across segments of society may be promoted by middle SES groups given their bridging role
  - asymmetry in interaction may suggest existence of social hierarchy (wealthier areas attract unequal amount of money and information)
  - better strategy in store allocation may lead to more even distribution of capital and less segregated interaction pattern

Discussion

- Summary & Implication
  - segregation in offline & online interaction is stronger than expected by geography
  - mitigation of segregation may focus on encouraging physical and social interaction (not merely on exposure [Bail18])
  - interaction across segments of society may be promoted by middle SES groups given their bridging role
  - asymmetry in interaction may suggest existence of social hierarchy (wealthier areas attract unequal amount of money and information)
  - better strategy in store allocation may lead to more even distribution of capital and less segregated interaction pattern

- Limitation
  - representativity of data sets
  - unobserved confounding variables
  - correlation but no causation

Micro-level segregation

- People of different backgrounds do not frequent the same places

Micro-level segregation

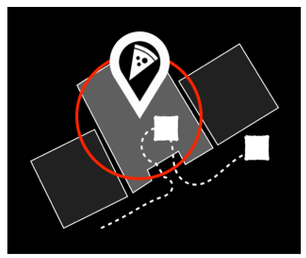
- People of different backgrounds do not frequent the same places
- The Atlas of Inequality

Framework

- Data sets
  - anonymised high-resolution mobile device location pings in 11 US census core-based statistical areas (CBSAs) in six months
  - verified Foursquare venues with more than 5 check-ins in all CBSAs

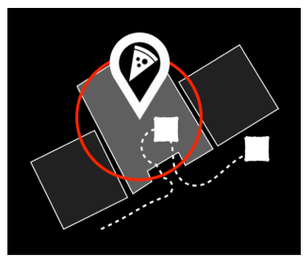
Framework

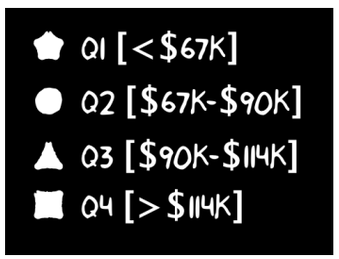
- Data sets
  - anonymised high-resolution mobile device location pings in 11 US census core-based statistical areas (CBSAs) in six months
  - verified Foursquare venues with more than 5 check-ins in all CBSAs
- Methodology
  - extract stays and attribute stays to places



Framework

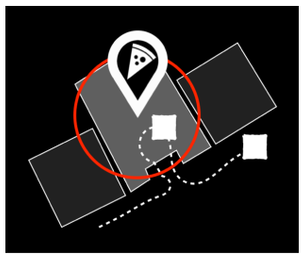
- Data sets
  - anonymised high-resolution mobile device location pings in 11 US census core-based statistical areas (CBSAs) in six months
  - verified Foursquare venues with more than 5 check-ins in all CBSAs
- Methodology
  - extract stays and attribute stays to places
  - identify user home location and income status (in four quantiles)

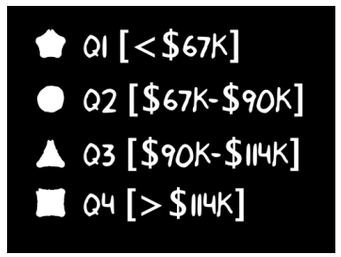


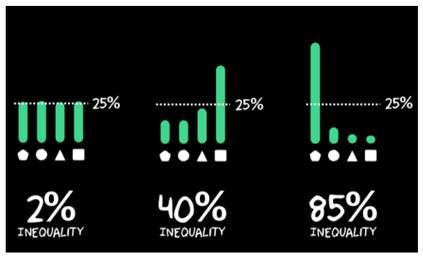


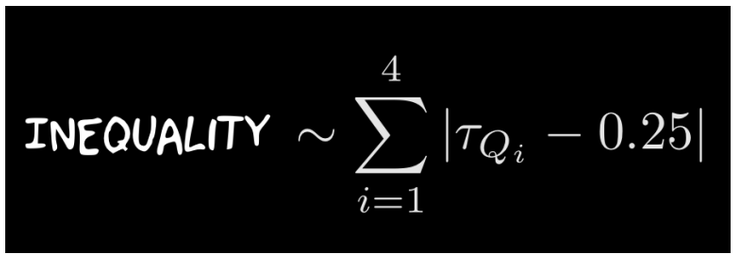
Framework

- Data sets
  - anonymised high-resolution mobile device location pings in 11 US census core-based statistical areas (CBSAs) in six months
  - verified Foursquare venues with more than 5 check-ins in all CBSAs
- Methodology
  - extract stays and attribute stays to places
  - identify user home location and income status (in four quantiles)
  - compute inequality measure at places

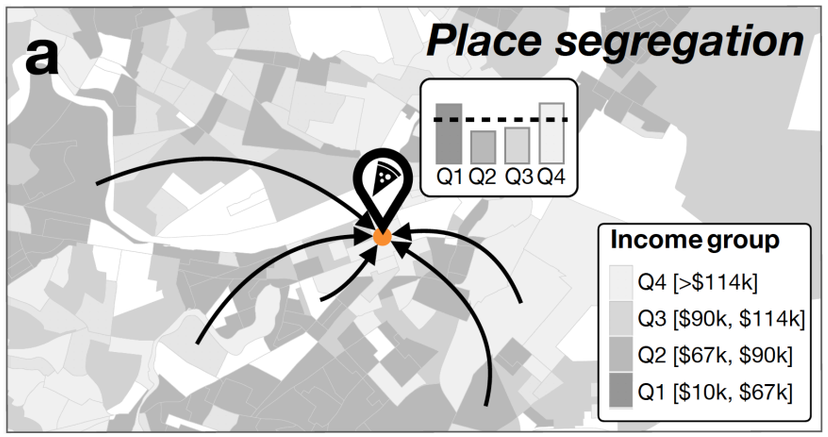


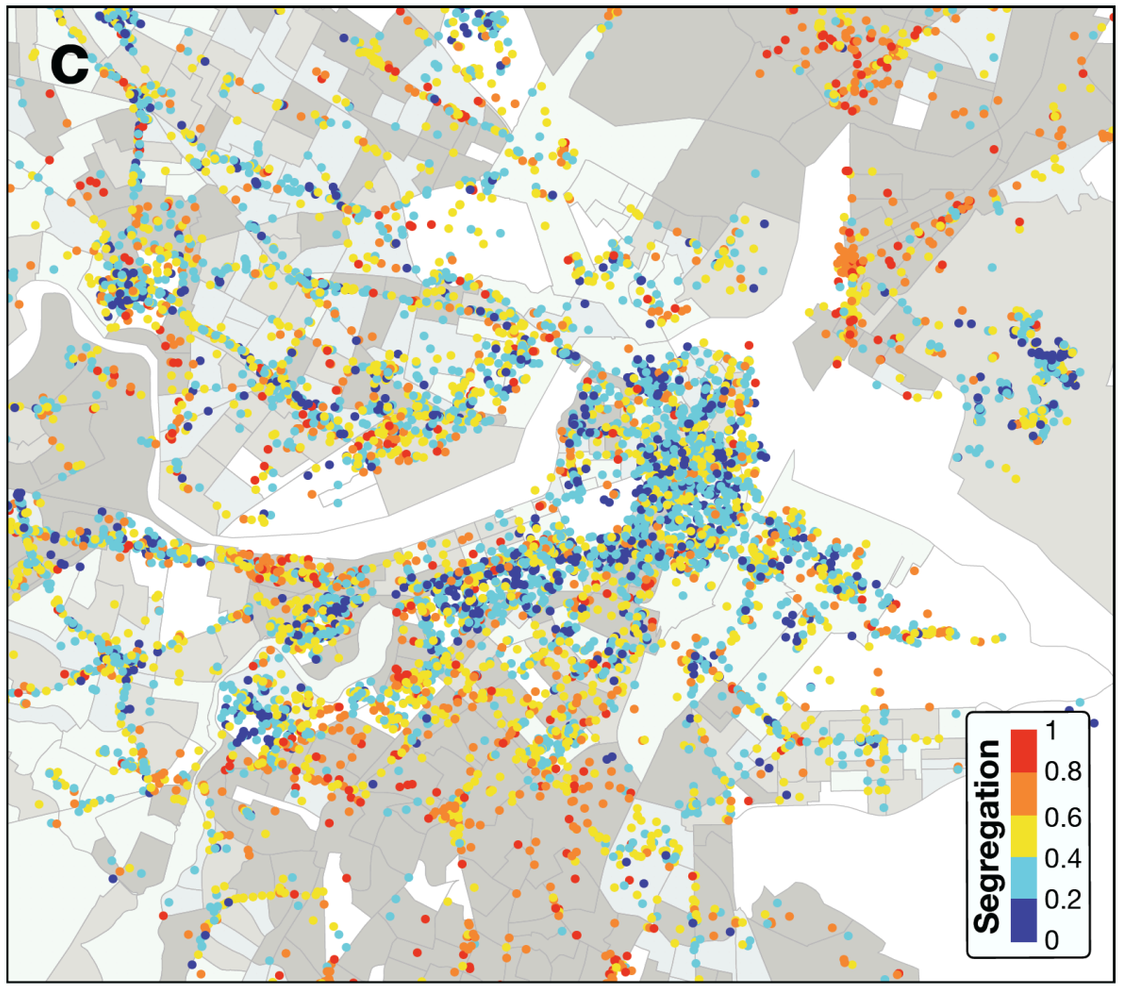






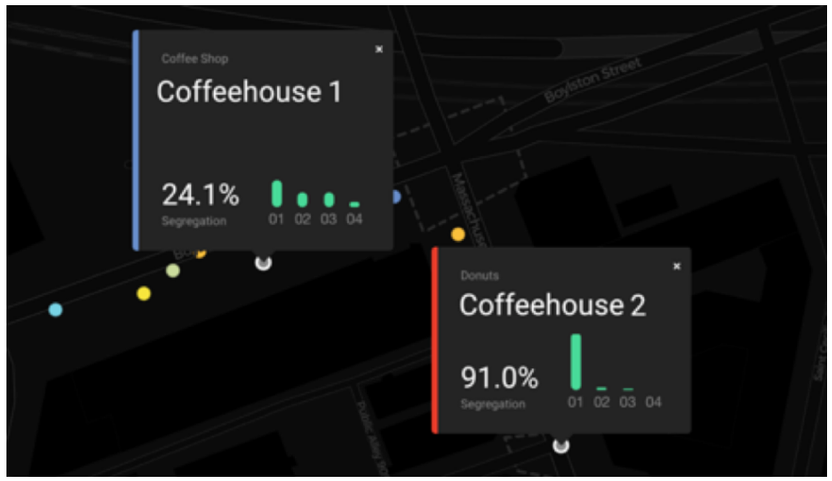
Place segregation

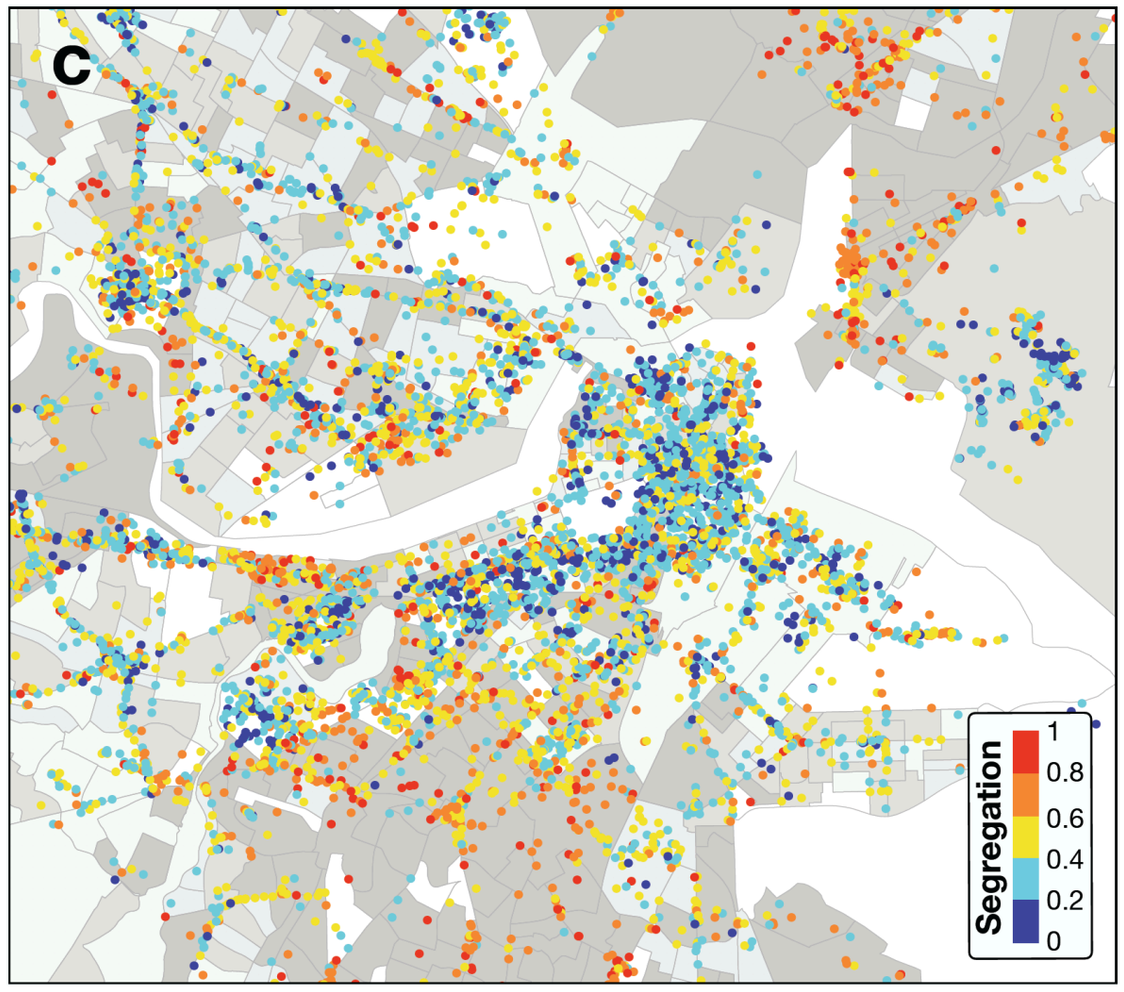




segregation occurs at micro level (within 25 meters)

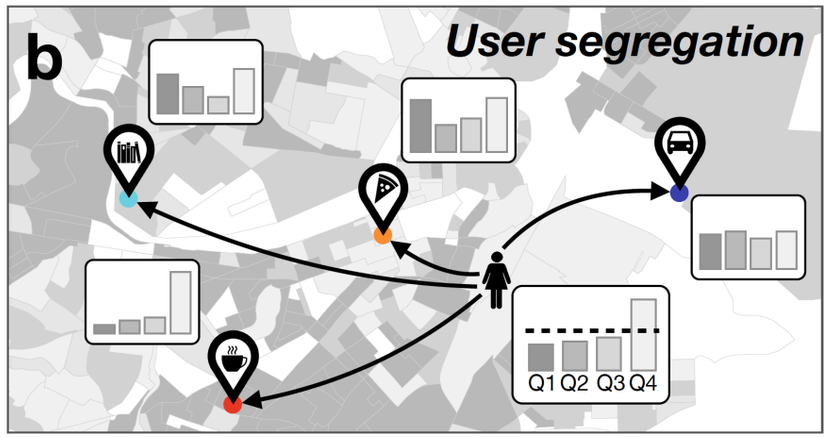
Place segregation

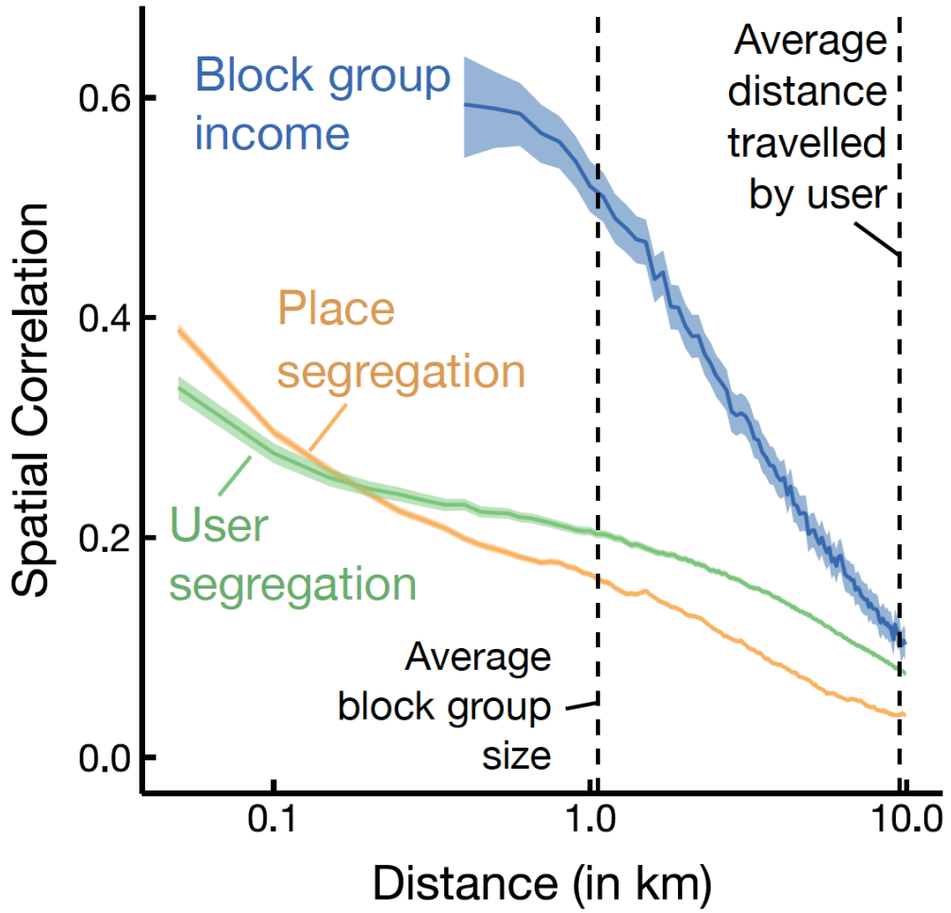




segregation occurs at micro level (within 25 meters)

User segregation

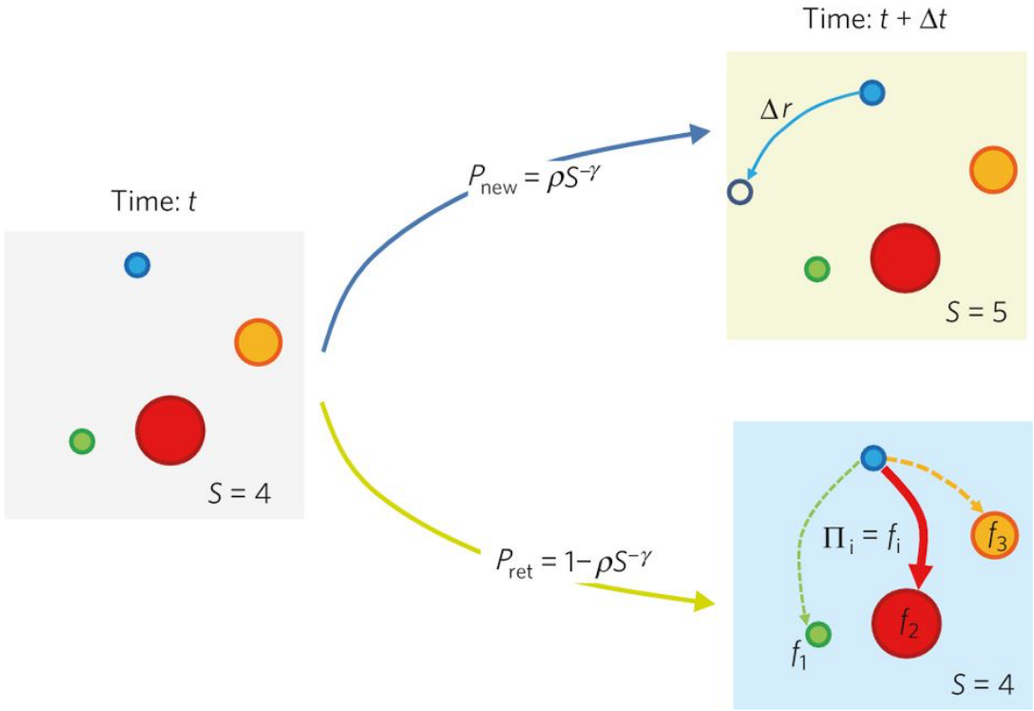




individual segregation is not only determined by where people live

Modelling segregation

- Exploration and Preferential Return (EPR) model [Song10]

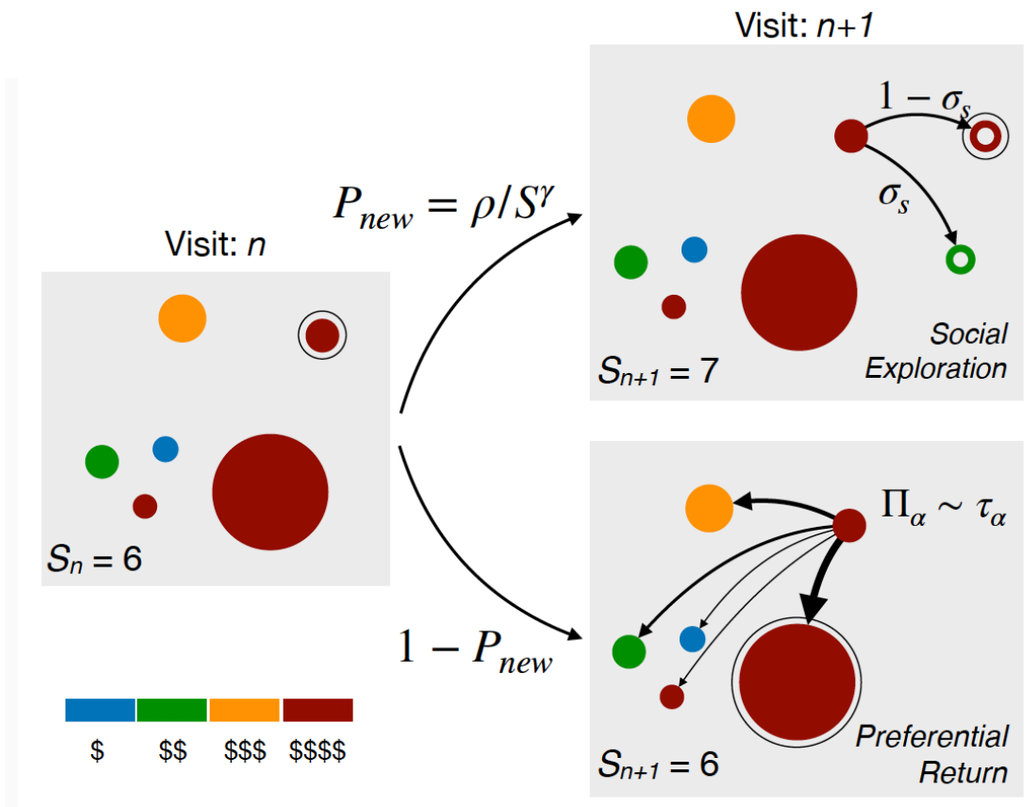


- exploration vs exploitation
- if exploration, visit a new location according to distance (place exploration)
- if exploitation, return to a previous location with preferential attachment

Song et al., "Modelling the scaling properties of human mobility," Nature Physics, 2010.

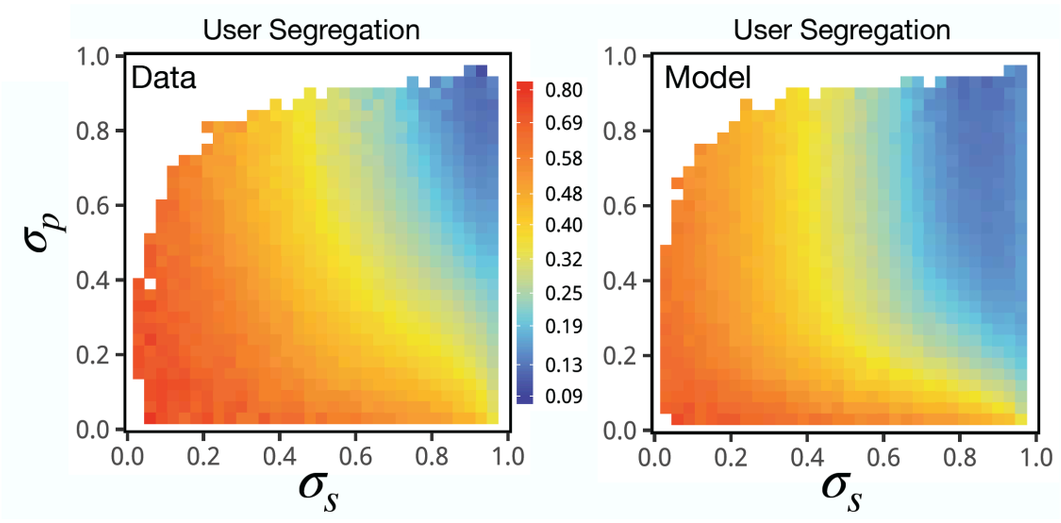
Modelling segregation

- Social Exploration and Preferential Return (social-EPR) model



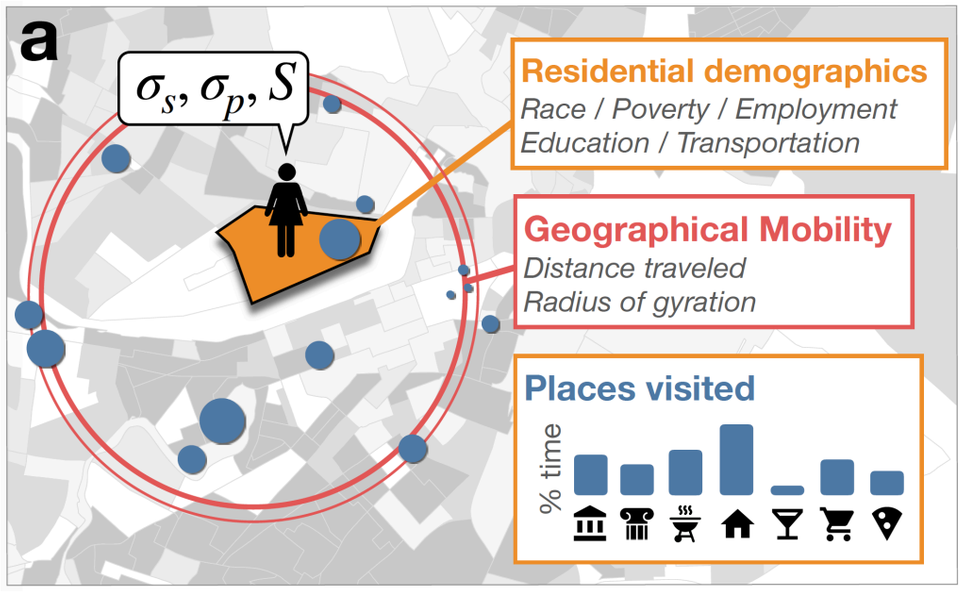
- exploration vs exploitation
- if exploration, with probability visit a new location of different income status (place and social exploration)
- if exploitation, return to a previous location with preferential attachment

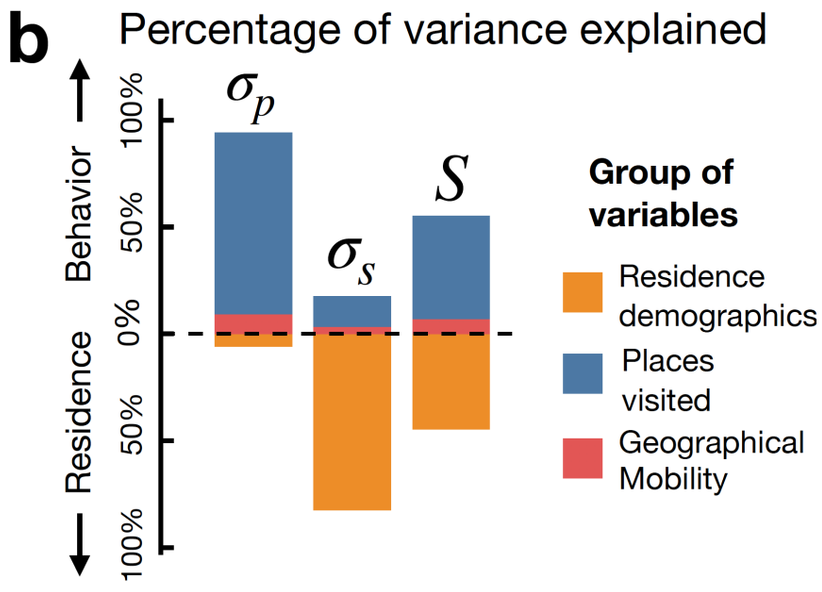
Modelling segregation



- for each user obtain parameters for place and social exploration
  - σ_p : # unique place / # place
  - σ_s : # minority place / # place
- simulate user visits using social-EPR model and compute segregation
- simulated segregation strongly correlated with empirical data (r=0.8)

Explaining place and social exploration





individual segregation explained by behaviour (55% variance) as well as residential (45% variance) factors

Discussion

- Summary & Implication
  - measurement of segregation as an emergent behavioural process in cities

Discussion

- Summary & Implication
  - measurement of segregation as an emergent behavioural process in cities
  - study of micro-level segregation in cities and segregation analysis across different types of venues

Discussion

- Summary & Implication
  - measurement of segregation as an emergent behavioural process in cities
  - study of micro-level segregation in cities and segregation analysis across different types of venues
  - individual segregation explained by patterns of exploration in both physical and social space

Discussion

- Summary & Implication
  - measurement of segregation as an emergent behavioural process in cities
  - study of micro-level segregation in cities and segregation analysis across different types of venues
  - individual segregation explained by patterns of exploration in both physical and social space
  - individual segregation modelled by simple extension (with social factor) of the EPR mobility model

Discussion

- Summary & Implication
  - measurement of segregation as an emergent behavioural process in cities
  - study of micro-level segregation in cities and segregation analysis across different types of venues
  - individual segregation explained by patterns of exploration in both physical and social space
  - individual segregation modelled by simple extension (with social factor) of the EPR mobility model
  - from residential redevelopment to city development that impacts who residents have opportunities to interact with

Discussion

- Summary & Implication
  - measurement of segregation as an emergent behavioural process in cities
  - study of micro-level segregation in cities and segregation analysis across different types of venues
  - individual segregation explained by patterns of exploration in both physical and social space
  - individual segregation modelled by simple extension (with social factor) of the EPR mobility model
  - from residential redevelopment to city development that impacts who residents have opportunities to interact with

- Limitation
  - focused on individuals for whom home location could be identified
  - focused on venues available via the Foursquare API
  - focused on segregation by income (not race/ethnicity)

Acknowledgement & Resources

- Collaborators
  - Alfredo Morales, MIT Media Lab & New England Complex Systems Institute
  - Eaman Jahani, MIT Institute for Data, Systems, and Society
  - Esteban Moro, MIT Media Lab & University Carlos III de Madrid
  - Bruno Lepri, Fondazione Bruno Kessler
  - Burçin Bozkaya, Sabancı University
  - Carlos Sarraute, Grandata
  - Yaneer Bar-Yam, New England Complex Systems Institute
  - Alex 'Sandy' Pentland, MIT Media Lab
  - Dan Calacci, MIT Media Lab

- Resources
  - Alfredo et al., "Segregation and polarization in urban areas," Royal Society Open Science, 2019.
  - Esteban Moro, "Overcoming urban isolation," TEDx Cambridge, 2018.
  - The Atlas of Inequality: https://inequality.media.mit.edu