

# Urban Diversity Through an Information-Theoretic Lens

---

Phil Chodrow

July 4, 2016

MIT Human Mobility and Networks Laboratory | Operations Research Center

# Outline

---

1. Introduction
2. Diversity and Information Theory
3. Identifying Natural Neighborhoods
4. Wrapping Up

# Introduction

---

# Neighborhoods and Diversity

---

Two questions for today:

# Neighborhoods and Diversity

---

Two questions for today:

1. What is **urban diversity**, and can we measure it in spatially-aware ways?

# Neighborhoods and Diversity

Two questions for today:

1. What is **urban diversity**, and can we measure it in spatially-aware ways?
2. What are **neighborhoods**, and how can we distinguish neighborhoods from each other?

# Neighborhoods and Diversity

Two questions for today:

1. What is **urban diversity**, and can we measure it in spatially-aware ways?
2. What are **neighborhoods**, and how can we distinguish neighborhoods from each other?

## Central Claim

These two questions are deeply related, and we should approach both using *information theory*.

## Three Dimensions of Urban Diversity

---

**Global diversity** is the extent to which different groups are equally represented in the city as a whole.

# Three Dimensions of Urban Diversity

---

**Global diversity** is the extent to which different groups are equally represented in the city as a whole.

**Spatial evenness** “or clustering, refers to the extent to which groups are similarly distributed in residential space.” [13]

# Three Dimensions of Urban Diversity

**Global diversity** is the extent to which different groups are equally represented in the city as a whole.

**Spatial evenness** “*or clustering, refers to the extent to which groups are similarly distributed in residential space.*” [13]

**Spatial exposure** “*refers to the extent that members of one group encounter members of another group (or their own group, in the case of spatial isolation) in their local spatial environments.*” [13]

# Three Dimensions of Urban Diversity

**Global diversity** is the extent to which different groups are equally represented in the city as a whole.

**Spatial evenness** “*or clustering, refers to the extent to which groups are similarly distributed in residential space.*” [13]

**Spatial exposure** “*refers to the extent that members of one group encounter members of another group (or their own group, in the case of spatial isolation) in their local spatial environments.*” [13]

No global diversity  $\implies$  perfect evenness  $\implies$  maximal exposure.

# Neighborhoods, Exposure, and the Checkerboard

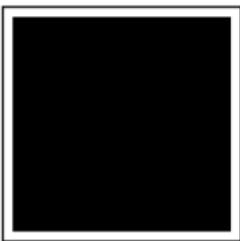
---

**Claim:** Spatial exposure is about the number, scale, and pattern of neighborhoods.

# Neighborhoods, Exposure, and the Checkerboard

Claim: Spatial exposure is about the number, scale, and pattern of neighborhoods.

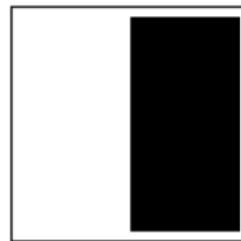
(a)



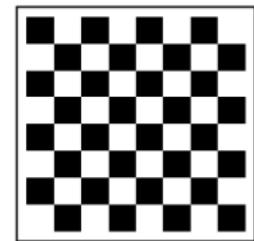
(b)



(c)



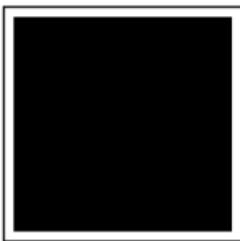
(d)



# Neighborhoods, Exposure, and the Checkerboard

Claim: Spatial exposure is about the number, scale, and pattern of neighborhoods.

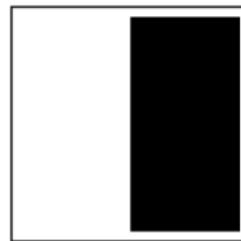
(a)



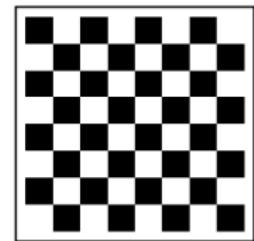
(b)



(c)

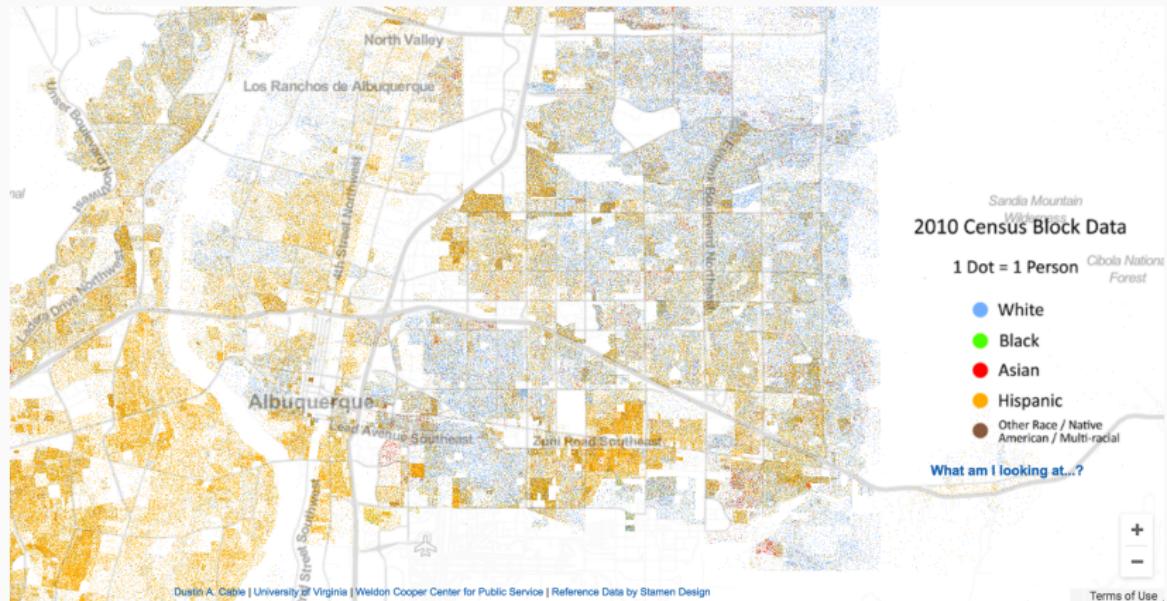


(d)



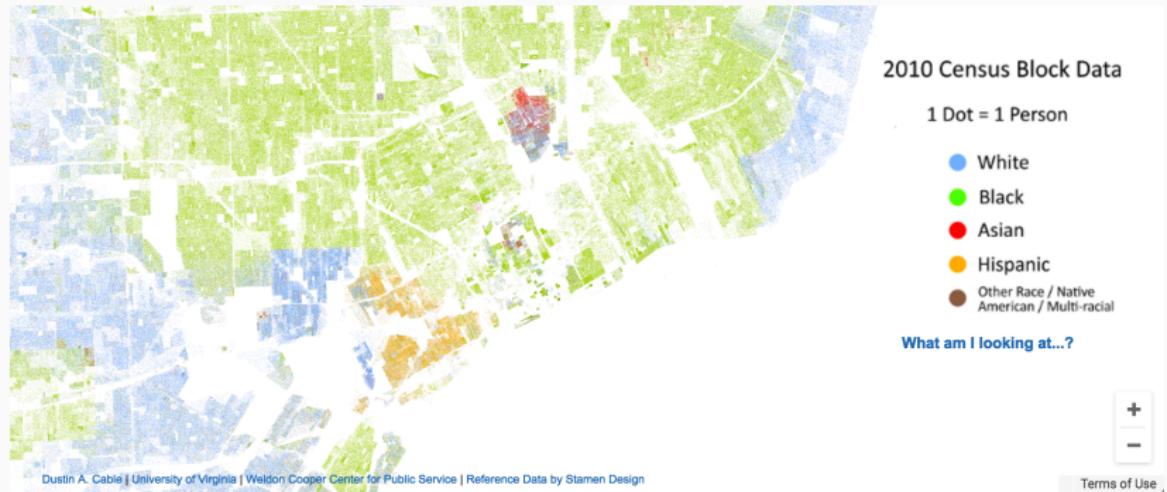
City	Exposure	Neighborhoods
(c)	Low	Few, large
(d)	High	Many, small

# Albuquerque: Low diversity, fairly even, high exposure



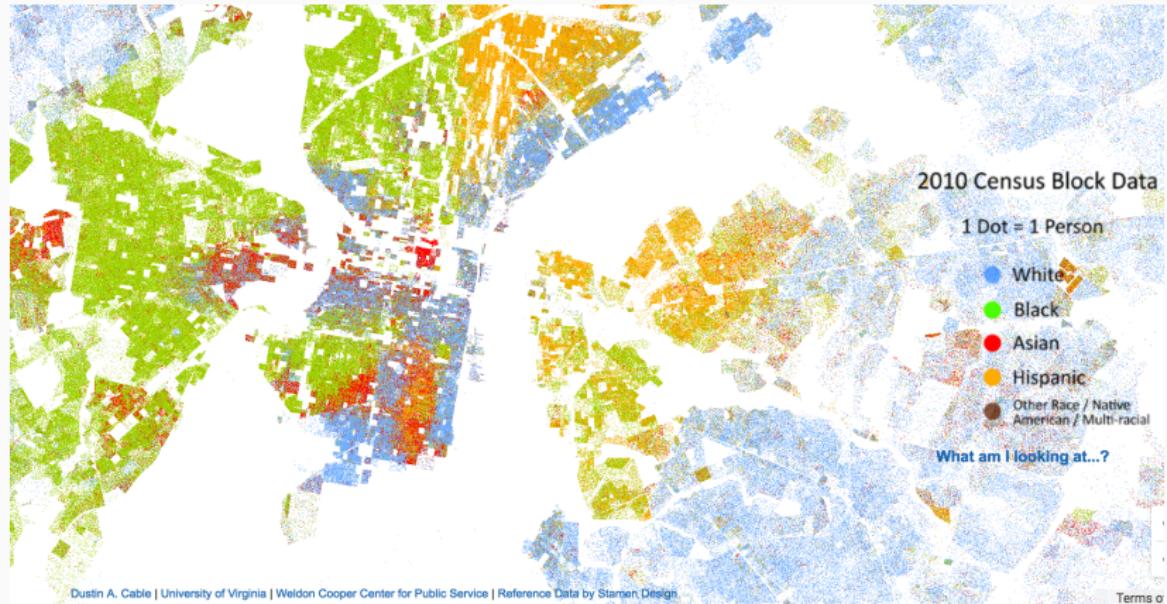
*Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)*

# Detroit: High diversity, highly uneven, low exposure



*Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)*

# Philadelphia: High diversity, highly uneven, high exposure



*Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)*

## Diversity and Information Theory

---

## Why Information Theory?

---

Information theory quantifies concepts like uncertainty and predictability. Why use it for studying diversity?

# Why Information Theory?

Information theory quantifies concepts like uncertainty and predictability. Why use it for studying diversity?

## Core Insight

- Complex systems allow **some** prediction...
- ...But good prediction is **hard**.

# Why Information Theory?

Information theory quantifies concepts like uncertainty and predictability. Why use it for studying diversity?

## Core Insight

- Complex systems allow **some** prediction...
- ...But good prediction is **hard**.

## Examples of prediction:

- If you pick a random person from a city, is it hard to guess their race? “Yes”  $\Rightarrow$  more diverse city.
- Does it help if you tell me where they live? “Yes”  $\Rightarrow$  more uneven city.

# Entropy as Global Diversity

Let  $p(X, Y)$  be the joint distribution of location  $X$  and race  $Y$ .

The marginal distribution of race alone is  $p(Y) = \sum_{x \in \mathcal{X}} p(x, Y)$ .

## Entropy

$$H(Y) \triangleq -\mathbb{E}_Y[\log p(Y)] = -\sum_{y \in \mathcal{Y}} p(Y) \log p(Y)$$

# Entropy as Global Diversity

Let  $p(X, Y)$  be the joint distribution of location  $X$  and race  $Y$ .

The marginal distribution of race alone is  $p(Y) = \sum_{x \in \mathcal{X}} p(x, Y)$ .

## Entropy

$$H(Y) \triangleq -\mathbb{E}_Y[\log p(Y)] = -\sum_{y \in \mathcal{Y}} p(Y) \log p(Y)$$

$H(Y) = 0 \implies$  city is monoracial

$H(Y) = \log |\mathcal{Y}| \implies$  all races are represented equally

# Mutual Information as Evenness

The mutual information quantifies the degree of dependence between two random variables.

## Mutual Information

$$\begin{aligned} I(X, Y) &\triangleq \mathbb{E}_{X,Y} \left[ \log \frac{p(X, Y)}{p(X)p(Y)} \right] \\ &= \int_{\mathbb{R}^n} d^n x \sum_{y \in \mathcal{Y}} p(x, Y) \log \frac{p(x, Y)}{p(x)p(Y)}. \end{aligned}$$

# Mutual Information as Evenness

The mutual information quantifies the degree of dependence between two random variables.

## Mutual Information

$$\begin{aligned} I(X, Y) &\triangleq \mathbb{E}_{X,Y} \left[ \log \frac{p(X, Y)}{p(X)p(Y)} \right] \\ &= \int_{\mathbb{R}^n} d^n x \sum_{y \in \mathcal{Y}} p(x, Y) \log \frac{p(x, Y)}{p(x)p(Y)}. \end{aligned}$$

$I(X, Y) = 0 \implies$  location and race are independent

$I(X, Y) = \log |\mathcal{Y}| \implies$  location determines race

## Spatializing Mutual Information

Let  $I_r(x_0)$  be local the mutual information between  $X$  and  $Y$ , restricted to a small area of radius  $r$  centered at  $x_0$ . Think of this as the **spatial variation of racial trends in a small area**.

# Spatializing Mutual Information

Let  $I_r(x_0)$  be local the mutual information between  $X$  and  $Y$ , restricted to a small area of radius  $r$  centered at  $x_0$ . Think of this as the **spatial variation of racial trends in a small area**.

## Theorem

Under appropriate regularity conditions, for small  $r$ ,

$$\frac{I_r(x_0)}{r^2} \cong \frac{1}{4} \text{trace } J_Y(x_0), \quad (\text{proof on request})$$

where  $J_Y(x_0)$  is the Fisher information in  $Y$  about  $x_0$

# Spatializing Mutual Information

Let  $I_r(x_0)$  be local the mutual information between  $X$  and  $Y$ , restricted to a small area of radius  $r$  centered at  $x_0$ . Think of this as the **spatial variation of racial trends in a small area**.

## Theorem

Under appropriate regularity conditions, for small  $r$ ,

$$\frac{I_r(x_0)}{r^2} \cong \frac{1}{4} \text{trace } J_Y(x_0), \quad (\text{proof on request})$$

where  $J_Y(x_0)$  is the Fisher information in  $Y$  about  $x_0$

**Significance:** the local information  $I_r(x_0)$  is related to an intrinsic statistical measure  $J_Y(x)$  that is independent of the spatial resolution  $r$ .

## Mean Local Information Measures Spatial Exposure

---

We're interested in the average value of local information:

# Mean Local Information Measures Spatial Exposure

We're interested in the average value of local information:

## Mean Local Information

$$J(X, Y) \triangleq \mathbb{E}_X[\text{trace } J_Y(X)] .$$

# Mean Local Information Measures Spatial Exposure

We're interested in the average value of local information:

## Mean Local Information

$$J(X, Y) \triangleq \mathbb{E}_X[\text{trace } J_Y(X)] .$$

$J(X, Y) = 0 \implies$  No neighborhoods, low exposure

$J(X, Y)$  large  $\implies$  Lots of neighborhoods, high exposure

# Mean Local Information Measures Spatial Exposure

We're interested in the average value of local information:

## Mean Local Information

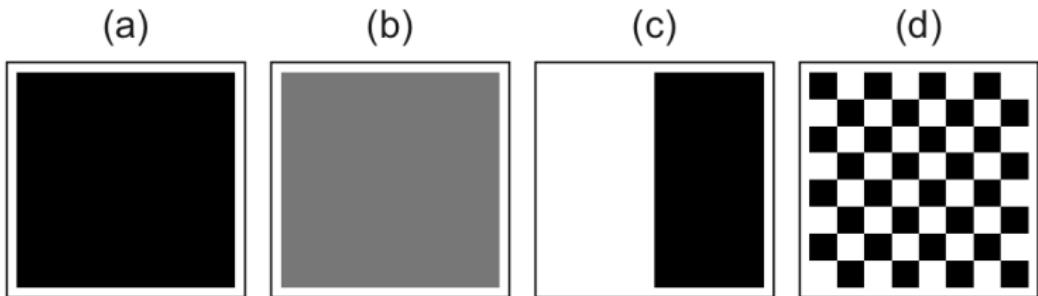
$$J(X, Y) \triangleq \mathbb{E}_X[\text{trace } J_Y(X)] .$$

$J(X, Y) = 0 \implies$  No neighborhoods, low exposure

$J(X, Y)$  large  $\implies$  Lots of neighborhoods, high exposure

This is one way of “spatializing” the mutual information  $I(X, Y)$ ; for another see [14, 15].

# Back to the checkerboard



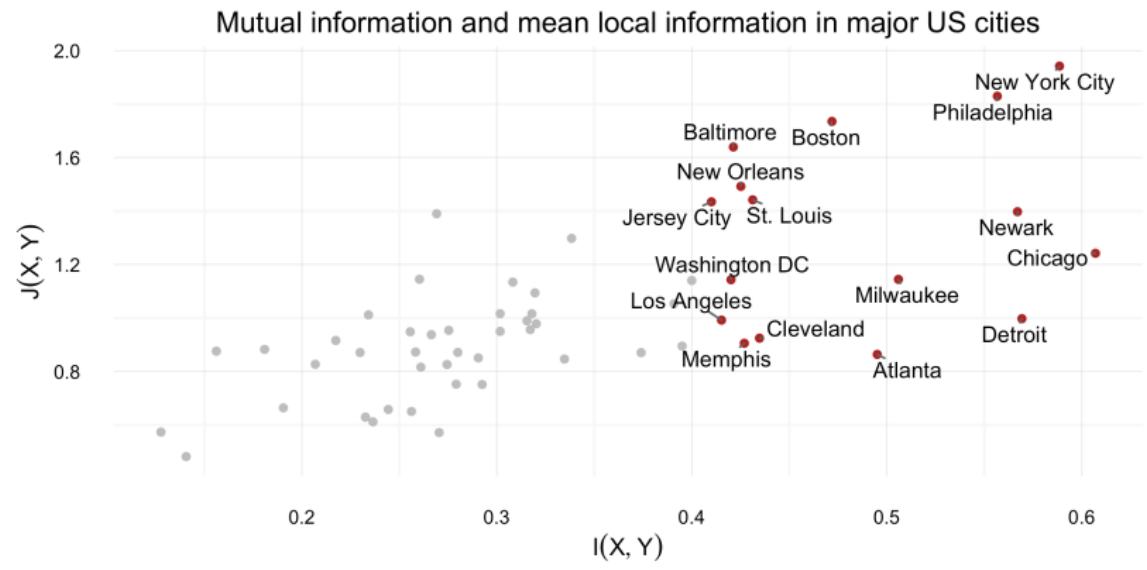
City	$H(Y)$	$I(X, Y)$	$J(X, Y)$
(a)	0.0	0.0	0.0
(b)	0.7	0.0	0.0
(c)	0.7	0.7	0.6
(d)	0.7	0.7	2.7

# A Three-Dimensional Characterization of Urban Diversity

City	$H(Y)$	$I(X, Y)$	$J(X, Y)$
Albuquerque	1.08	0.19	0.66
Detroit	1.12	0.57	1.00
Philadelphia	1.31	0.56	1.83

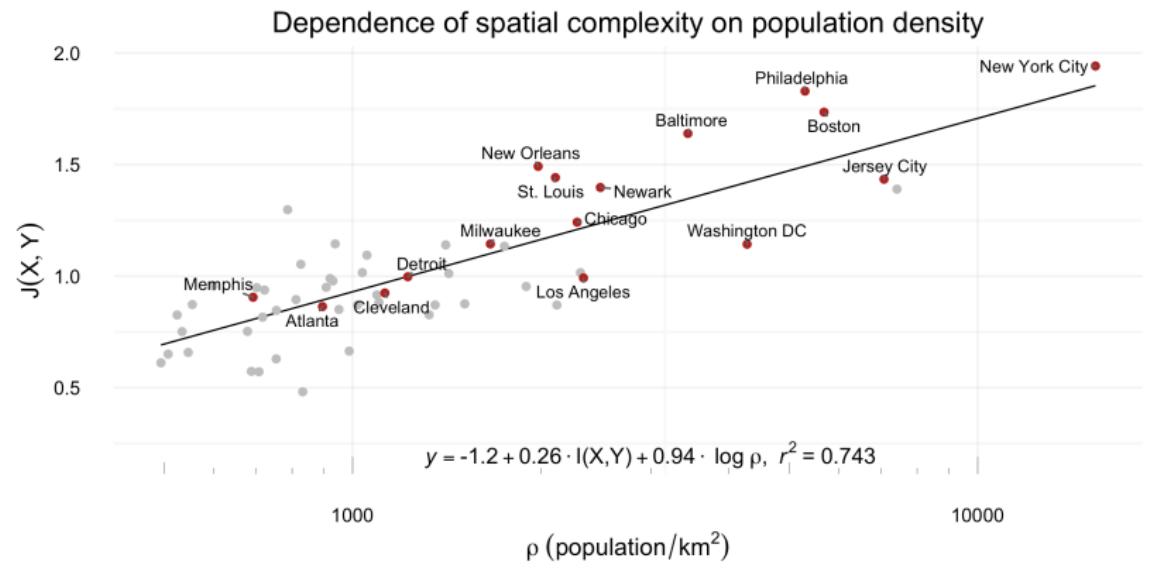
*Data accessed from the American Community Survey of the U.S. Census [1].*

# Visualizing Diversity Profiles in Major Cities



- Philadelphia has greater exposure / more neighborhood structure than Detroit.

# Urban Density and the Compression of Social Space



- Denser cities have higher exposure and more neighborhood structure

## Identifying Natural Neighborhoods

---

## How can we use this?

---

$H(Y)$ ,  $I(X, Y)$ , and  $J(X, Y)$  measure diversity, evenness, and spatial exposure.

Also, spatial exposure is equivalent to neighborhood structure.

## How can we use this?

---

$H(Y)$ ,  $I(X, Y)$ , and  $J(X, Y)$  measure diversity, evenness, and spatial exposure.

Also, spatial exposure is equivalent to neighborhood structure.

So, can we use these measures to **find** neighborhoods?

## How can we use this?

---

$H(Y)$ ,  $I(X, Y)$ , and  $J(X, Y)$  measure diversity, evenness, and spatial exposure.

Also, spatial exposure is equivalent to neighborhood structure.

So, can we use these measures to **find** neighborhoods?

Yes!

- We can use  $I(X, Y)$  to find neighborhoods
- We can use  $H(Y)$ ,  $I(X, Y)$ , and  $J(X, Y)$  to estimate how neighborhoods many we need to preserve key patterns.
- So, we can view  $H(Y)$ ,  $I(X, Y)$ , and  $J(X, Y)$  as measurements of **complexity** in spatiotemporal structure.

## Key Property of Mutual Information

Suppose we cluster the locations  $X$  into groups with labels  $C$ .

“Additive Organizational Decomposability” (Chain Rule)

$$I(X, Y) = I(C, Y) + I(X, Y|C) .$$

## Key Property of Mutual Information

Suppose we cluster the locations  $X$  into groups with labels  $C$ .

“Additive Organizational Decomposability” (Chain Rule)

$$I(X, Y) = I(C, Y) + I(X, Y|C) .$$

We can use this to identify neighborhoods!

# Key Property of Mutual Information

Suppose we cluster the locations  $X$  into groups with labels  $C$ .

## Recipe for Neighborhood Identification

$$I(X, Y) = \underbrace{I(C, Y)}_{\text{maximize this}} + \overbrace{I(X, Y|C)}^{\text{minimize this}} .$$

# Key Property of Mutual Information

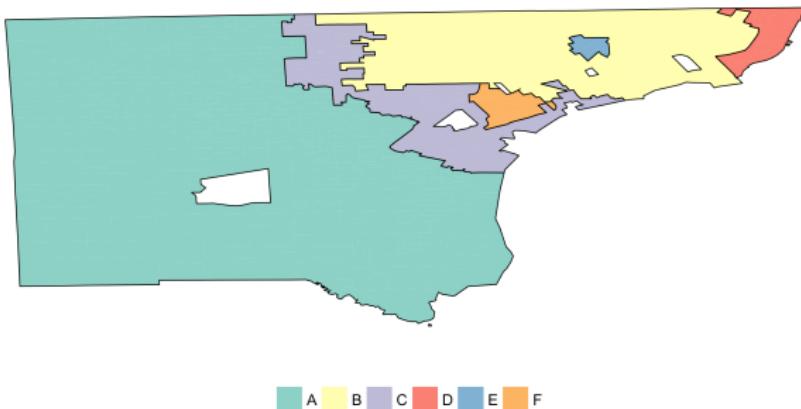
Suppose we cluster the locations  $X$  into groups with labels  $C$ .

## Recipe for Neighborhood Identification

$$I(X, Y) = \underbrace{I(C, Y)}_{\text{maximize this}} + \overbrace{I(X, Y|C)}^{\text{minimize this}} .$$

This optimization is hard to do exactly, but we can use a greedy approach based on hierarchical clustering.

# Practical Example: Racial Neighborhoods in Detroit

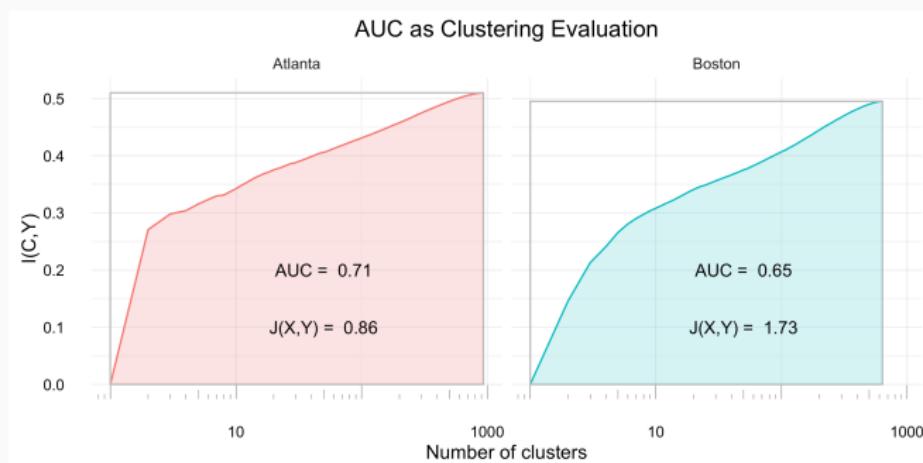


	A	B	C	D	E	F
White	36	2	7	2	1	1
Black	5	29	4			
Hispanic	2		1			2
Asian	2				1	
Other	2	1	1			

*Percentage of population by race and neighborhood.*

# Information Measures and “Clusterability”

Evaluate clusterings using “Area Under the Curve”:



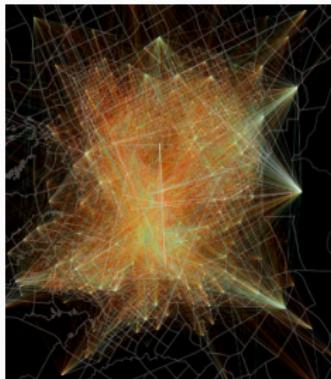
Together,  $H(Y)$ ,  $I(X, Y)$  and  $J(X, Y)$  explain 75% of the variability in the clusterability (AUC) of cities – they measure neighborhood structure.

## Wrapping Up

---

## Generality and future work

- **Other demographics:** education level, income, occupation type, etc.
- **Other dimensions:** time
  - Evolution of urban demographics (timescale: decades)
  - Dynamics of diversity in daily mobility (timescale: minutes)



Daily commuting flows in  
Riyad, Saudi Arabia.

*Image credit: Shan Jiang  
and MIT HuMNetLab*

## Download our tools!

R Package for Information-Theoretic Analysis

<https://github.com/PhilChodrow/compx>

This project: analysis and presentations

[https://github.com/PhilChodrow/spatial\\_complexity](https://github.com/PhilChodrow/spatial_complexity)

## Recap

---

Information theory offers:

- A **unified mathematical approach** to traditional concerns in segregation studies
- A **practical toolbox** for applied analysis of spatio-social trends
- A **theoretical platform** for formulating and asking new questions about human difference in space and time.

THANK YOU! Questions? Feedback?

## Supplementary Materials

---

## References I

2014 American Community Survey 5-Year Estimates, Table B01003, 2016.

Shun-Ichi Amari and Hiroshi Nagaoka.

**Methods of Information Geometry**, 2000.

Luis M A Bettencourt, Joe Hand, and José Lobo.

**Spatial Selection and the Statistics of Neighborhoods**.

2015.

Roger Bivand, Tim Keitt, and Barry Rowlingson.

**rgdal: bindings for the Geospatial Data Abstraction Library**,  
2014.

## References II

Roger Bivand and Nicholas Lewin-Koh.

`maptools`: Tools for reading and handling spatial objects,  
2014.

Roger Bivand and Colin Rundel.

`rgeos`: Interface to Geometry Engine - Open Source (GEOS),  
2014.

Thomas M Cover and Joy A Thomas.

*Elements of Information Theory.*

John Wiley and Sons, New York, 1991.

## References III

Imre Csiszar and Paul C Shields.

**Information Theory and Statistics: A Tutorial.**

*Foundations and Trends™ in Communications and Information Theory*, 1(4):417–528, 2004.

Barret A Lee, Sean F Reardon, Glenn Firebaugh, Chard R Farrell, Stephen A Matthews, and David O'Sullivan.

**Beyond the Census Tract: Patterns and Determinants of Racial Segregation at Multiple Geographic Scales.**

*American Sociological Review*, 73(5):766–791, 2008.

Douglas S Massey and Nancy A Denton.

**The Dimensions of Residential Segregation.**

*Social Forces*, 67(2):281–315, 1988.

## References IV

R Core Team.

**R: A Language and Environment for Statistical Computing,**  
2016.

Sean F Reardon and G Firebaugh.

**Measures of multigroup segregation.**

*Sociological Methodology*, 32:33–67, 2002.

Sean F. Reardon and David O'Sullivan.

**Measures of Spatial Segregation.**

*Sociological Methodology*, 34(1):121–162, 2004.

Elizabeth Roberto.

**Measuring Inequality and Segregation.**

2015.

## References V

---

Elizabeth Roberto.

**The Spatial Context of Residential Segregation.**

*arXiv.org*, pages 1–27, 2015.

## Relation of local and Fisher informations

Suppose that  $p(y|x) > 0$  and that  $p(y|x)$  is differentiable as a function of  $x$  for all  $x, y \in \mathbb{R}^n \times \mathcal{Y}$ . Define

$B_r \triangleq B_r(x_0) \triangleq \{x \in \mathbb{R}^n \mid |x - x_0| \leq r\}$ . Additionally, define the *local mutual information* in  $B_r$  as the mutual information between  $X$  and  $Y$  where  $X$  is restricted to  $B_r$ :

$$I_r(x_0) \triangleq \mathbb{E}_X[D[p(\cdot|X)\|p(\cdot|X \in B_r)]|X \in B_r] \quad (1)$$

$$= \int_{B_r} p(x|X \in B_r) D[p(\cdot|x)\|p(\cdot|X \in B_r)] d^n x. \quad (2)$$

where  $D[p\|q] \triangleq \sum_y p(y) \log \frac{p(y)}{q(y)}$  is the Kullback-Leibler divergence of  $q$  from  $p$ .

# Relation of local and Fisher informations

## Theorem

Under the stated conditions,

$$\lim_{r \rightarrow 0} \frac{I_r(x_0)}{r^2} = \frac{n}{2(n+2)} \text{trace } J_Y(x_0) . \quad (3)$$

The Fisher information matrix  $J_Y$  is given by

$$J_Y(x) \triangleq \mathbb{E}_Y [\nabla_x S_Y(x) \nabla_x S_Y(x)^T] \quad (4)$$

$$S_Y(x) \triangleq \log p(y|x) . \quad (5)$$

## Information Measures Check the Boxes

Criterion	$H(Y)$	$I(X, Y)$	$J(X, Y)$
Organizational Equivalence	✓	✓	✓
Size/Density Invariance	✓	✓	✓
Additive Group Decomposability	✓	✓	✓
Additive Spatial Decomposability	NA	✓	*
Scale Interpretability	NA	✓	✓
Boundary Independence	NA	✗	✗
Exchanges	NA	✓	✓

\* We contend that [13]'s version is not a workable criterion for explicitly spatial measures.

✓ In the theoretical definition; in practice we are reliant on the data as it is provided.

# Some Comparative Clusters of Equal Information

## Example clusters

Atlanta:  $I(C, Y) = 0.27$

White	32	3
Other	3	1
Hispanic	7	2
Black	9	38
Asian	5	

Boston:  $I(C, Y) = 0.27$

30	6	3	1	3
2		1	2	1
4	2	3	5	9
4		3	12	1
6				1

Chicago:  $I(C, Y) = 0.27$

White	37	2	1	2
Other	2	1		
Hispanic	10	2	8	5
Black	5	14	4	
Asian	6			

Detroit:  $I(C, Y) = 0.27$

44	3	2
3	1	
5		
9	29	
2	1	

A

B

C

D

E

A

B

C

D

E