Neighborhoods and Segregation: An Information-Theoretic Lens

Phil Chodrow

July 3, 2016

MIT Human Mobility and Networks Laboratory & Operations Research Center

Table of contents

1. Introduction
2. Diversity and Information Theory
3. Information-Theoretic Identification of Neighborhoods
4. Wrapping Up

Introduction

Two questions for today:

Two questions for today:

1. What is urban diversity, and how should we measure it?

Two questions for today:

1. What is **urban diversity**, and how should we measure it?
2. What is a **neighborhood**, and how can we distinguish neighborhoods in principled ways?

Neighborhoods and Diversity

Two questions for today:

1. What is **urban diversity**, and how should we measure it?
2. What is a **neighborhood**, and how can we distinguish neighborhoods in principled ways?

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."* [3]

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."* [3]

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."* [3]

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.”* [3]

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.”* [3]

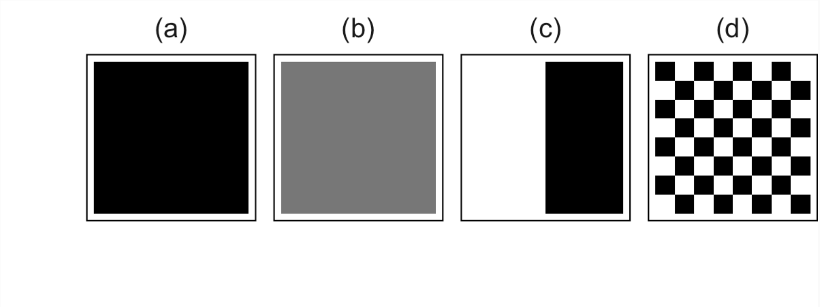
No global diversity $\implies$ perfect evenness $\implies$ maximal exposure.

Neighborhoods, Exposure, and the Checkerboard

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

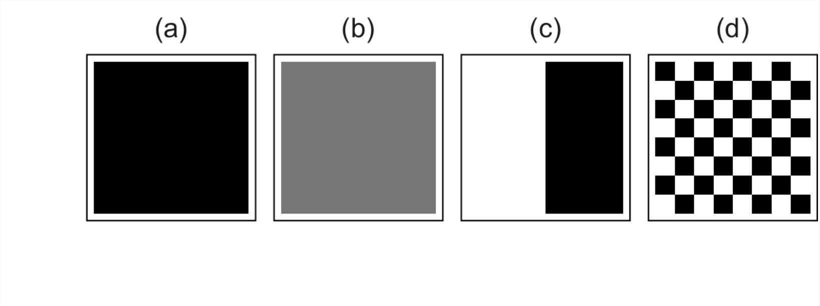
Neighborhoods, Exposure, and the Checkerboard

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



Neighborhoods, Exposure, and the Checkerboard

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



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

Albuquerque: Low diversity, fairly even, high exposure

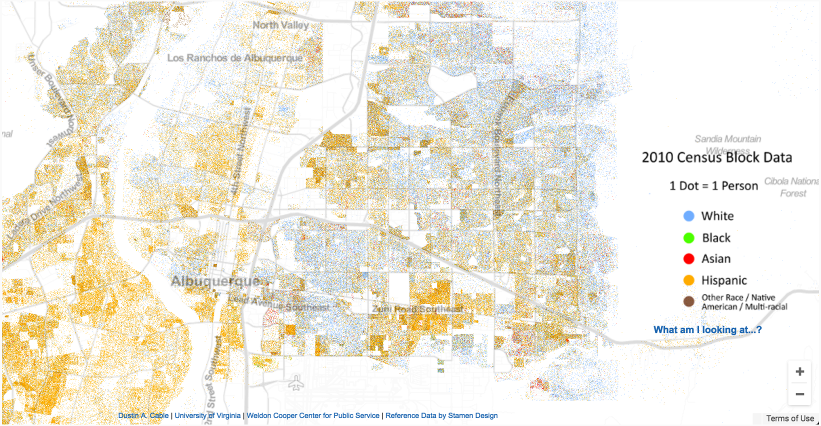


Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)

Detroit: High diversity, highly uneven, low exposure

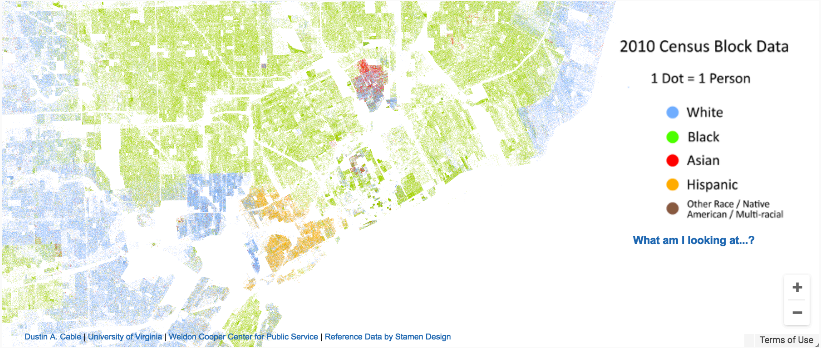


Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)

Philadelphia: High diversity, highly uneven, high exposure

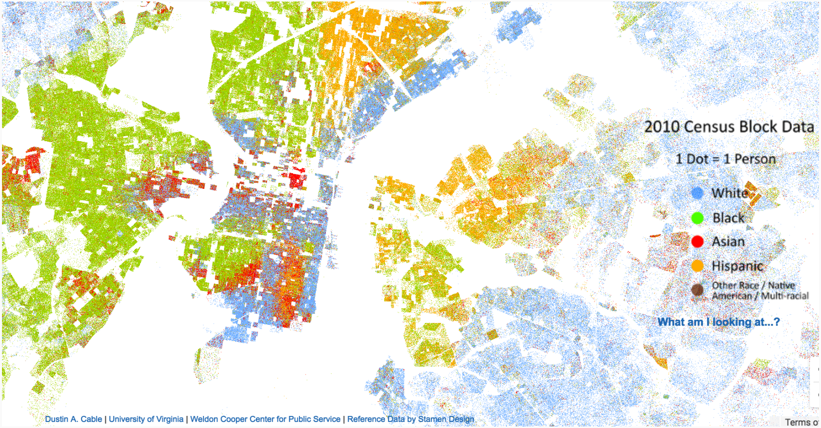


Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)

Diversity and Information Theory

Why Information Theory?

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

Why Information Theory?

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

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 race?

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" $\implies$ more diverse city.
- Does it help if you tell me where they live? "Yes" $\implies$ 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

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

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

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

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

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

$$\begin{aligned} H(Y) = 0 &\implies \text{city is monoracial} \\ H(Y) = \log|\mathcal{Y}| &\implies \text{all races are represented equally} \end{aligned}$$

Entropy measures global diversity: more entropy, more diversity.

Mutual Information as Evenness

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

Mutual Information

$$I(X, Y) \triangleq \sum_{x,y \in \mathcal{X} \times \mathcal{Y}} p(X, Y) \log \frac{p(X, Y)}{p(X)p(Y)} \, .$$

Mutual Information as Evenness

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

Mutual Information

$$I(X, Y) \triangleq \sum_{x,y \in \mathcal{X} \times \mathcal{Y}} p(X, Y) \log \frac{p(X, Y)}{p(X)p(Y)} \, .$$

$$\begin{aligned} I(X, Y) = 0 &\implies \text{location and race are independent} \\ I(X, Y) = \log |Y| &\implies \text{location determines race} \end{aligned}$$

Mutual Information as Evenness

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

Mutual Information

$$I(X, Y) \triangleq \sum_{x,y \in \mathcal{X} \times \mathcal{Y}} p(X, Y) \log \frac{p(X, Y)}{p(X)p(Y)} \ .$$

$$I(X, Y) = 0 \implies \text{location and race are independent}$$
$$I(X, Y) = \log |Y| \implies \text{location determines race}$$

Mutual information measures evenness: more information, the more uneven the spatial distribution of demographics.

Local 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.**

Local 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,

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

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

Local 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,

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

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 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

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

Mean Local Information Measures Spatial Exposure

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

Mean Local Information

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

$$\begin{aligned} J(X, Y) = 0 &\implies \text{No neighborhoods, low exposure} \\ J(X, Y) \text{ large} &\implies \text{Lots of neighborhoods, high exposure} \end{aligned}$$

Mean Local Information Measures Spatial Exposure

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

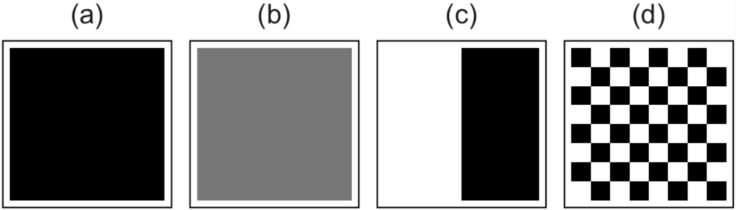
Mean Local Information

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

$$J(X, Y) = 0 \implies \text{No neighborhoods, low exposure}$$
$$J(X, Y) \text{ large} \implies \text{Lots of neighborhoods, high exposure}$$

This is one way of "spatializing" the mutual information $I(X, Y)$; for another see [4, 5].

Back to the checkerboard



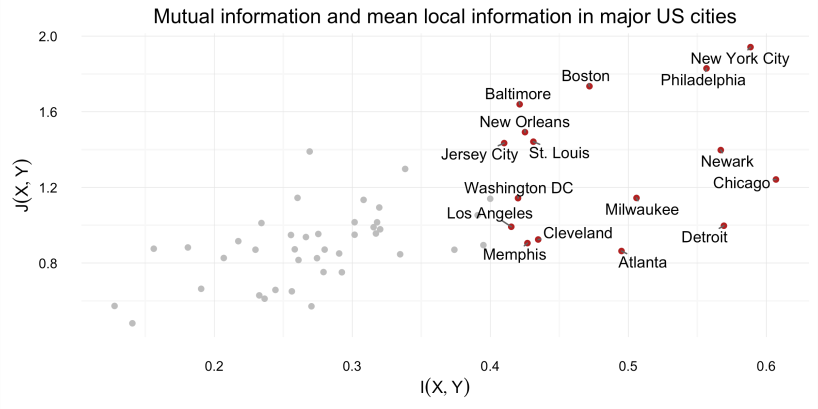
City	$H(Y)$	$I(X,Y)$	$J(X,Y)$
(a)	0.0	0.0	0.0
(b)	0.7	0.0	0.0
(c)	0.7	0.7	0.6
(d)	0.7	0.7	2.7

A Three-Dimensional Characterization of Urban Diversity

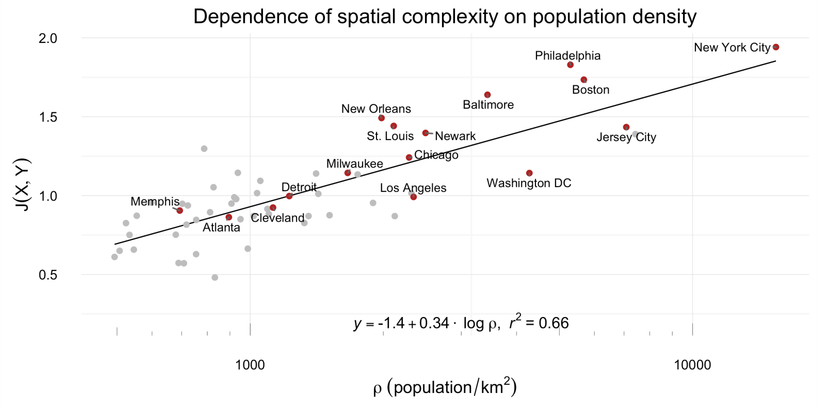
City	$H(Y)$	$I(X,Y)$	$J(X,Y)$
Albuquerque	1.08	0.19	0.66
Detroit	1.12	0.57	1.00
Philadelphia	1.31	0.56	1.83

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

Visualizing Diversity Profiles in Major Cities



Urban Density and the Compression of Social Space



Information-Theoretic Identification of Neighborhoods

How can we use this?

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

$J(X, Y)$ also measures "neighborhoodiness."

How can we use this?

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

$J(X,Y)$ also measures "neighborhoodiness."

Does 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.

$J(X, Y)$ also measures “neighborhoodiness.”

Does can we use these measures to **find** neighborhoods?

Yes!

- We can use $I(X, Y)$ to find neighborhoods
- We can use $I(X, Y)$ and $J(X, Y)$ to estimate how 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 spatiosocial 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

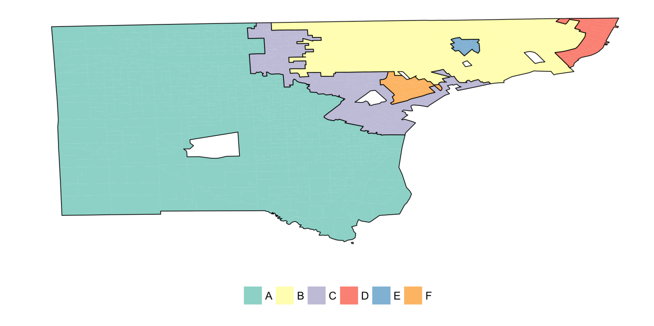
Suppose we cluster the locations X into groups with labels C.

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.

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			

Information Measures and "Clusterability"

We would expect that cities with low spatial exposure $J(X, Y)$ are "easy" to cluster. Evaluate 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.

Wrapping Up

Learnings

Information theory provides a natural language for the conceptualization and measurement of sociospatial variability.

Open Directions

Let’s learn about...

Thank you!

Questions?

Feedback?

References I

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

Shun-Ichi Amari and Hiroshi Nagaoka.
Methods of Information Geometry, 2000.

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.

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:

$$
\begin{aligned}
I_r(x_0) &\triangleq \mathbb{E}_X[D[p(\cdot|X)\|p(\cdot|X \in B_r)]|X \in B_r] &\quad (1)\\
&= \int_{B_r} p(x|X \in B_r) D[p(\cdot|x)\|p(\cdot|X \in B_r)] d^n x \,. &\quad (2)
\end{aligned}
$$

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,

$$\lim_{r \to 0} \frac{I_r(x_0)}{r^2} = \frac{n}{2(n+2)} \operatorname{trace} J_Y(x_0) \ . \tag{3}$$

The Fisher information matrix J_Y is given by

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

$$S_y(x) \triangleq \log p(y|x) \ . \tag{5}$$

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	✓	✓

* In the full paper, we argue that [3]'s version is not a workable criterion for 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

	A	B	C	D	E
White	32	3			
Other	3	1			
Hispanic	7	2			
Black	9	38			
Asian	5				

Boston: I(C,Y) = 0.27

	A	B	C	D	E
White	30	6	3	1	3
Other	2		1	2	1
Hispanic	4	2	3	5	9
Black	4		3	12	1
Asian	6				1

Chicago: I(C,Y) = 0.27

	A	B	C	D	E
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

	A	B	C	D	E
White	44	3	2		
Other	3	1			
Hispanic	5				
Black	9	29			
Asian	2	1			