FINDING THE FAULT LINES:

DETECTING URBAN SOCIAL BOUNDARIES USING SOCIAL DATA SCIENCE

NSF-SES
1733705

University of BRISTOL
UNIVERSITY OF CALIFORNIA UCRIVERSIDE

LEVI JOHN WOLF
ELIJAH KNAAP
SERGIO REY

levi.john.wolf@bristol.ac.uk
@levijohnwolf
ljwolf.org

BOUNDARIES

GOODNESS OF FIT: THE SILHOUETTE

WHOSE “GOOD” IS IT ANYWAY?

EXAMPLE: BROOKLYN

BOUNDARIES

AN EMINENTLY-GEOGRAPHICAL CONSTRUCT

GOODNESS OF FIT: THE SILHOUETTE

WHOSE "GOOD" IS IT ANYWAY?

EXAMPLE: BROOKLYN

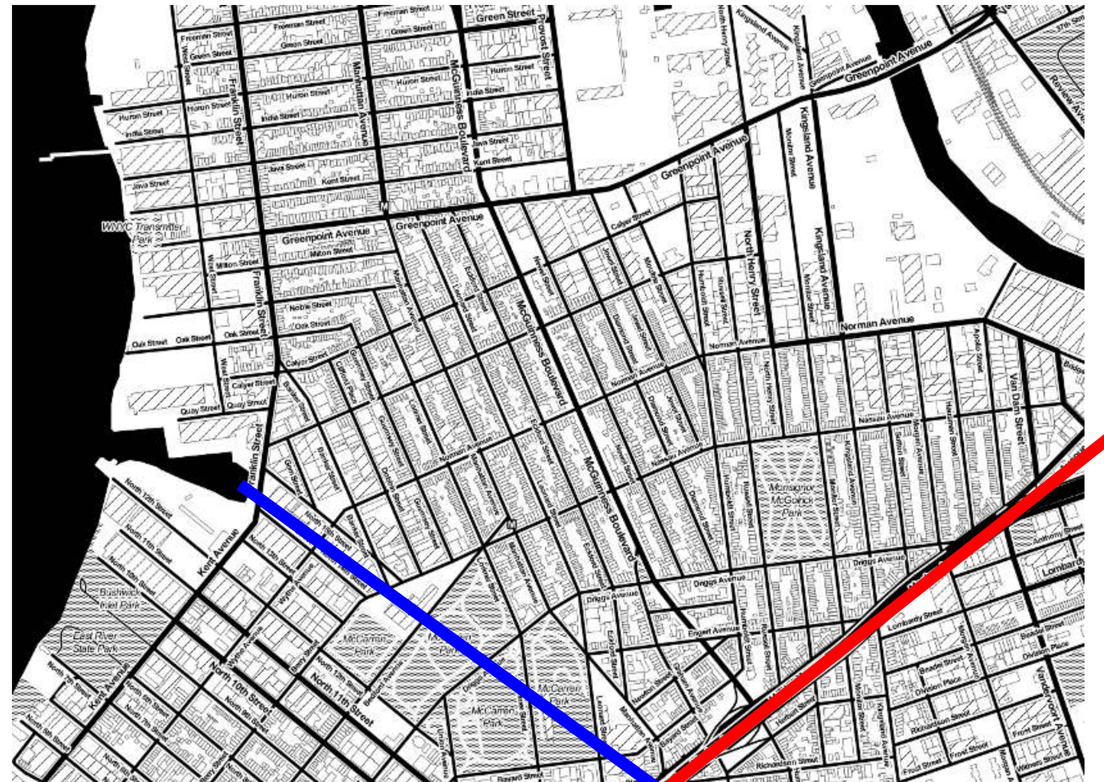
CONCEPTUALIZING BOUNDARIES

“Williamsburg becomes Greenpoint at the Bushwick Inlet”

CONCEPTUALIZING BOUNDARIES

“Williamsburg becomes Greenpoint at the Bushwick Inlet”

“Greenpoint is bordered on the southeast by the BQE”

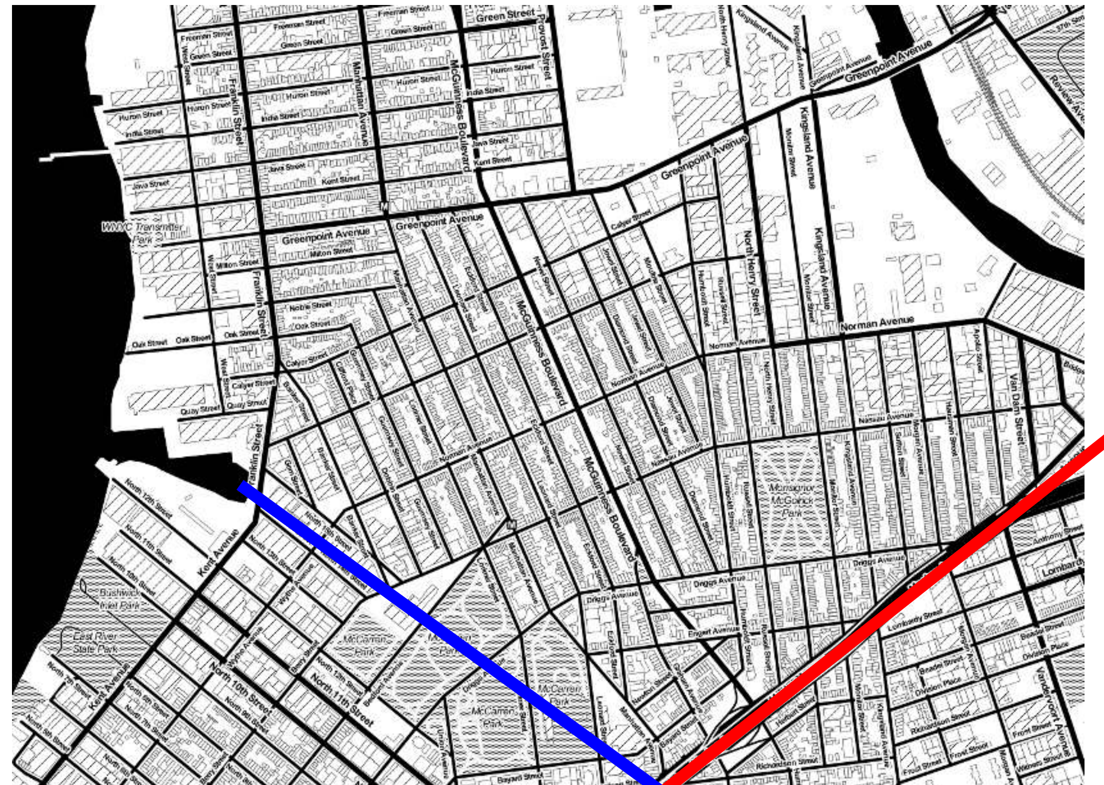


CONCEPTUALIZING BOUNDARIES

BOUNDARIES AS NATURALISTIC DIVISIONS OF URBAN LIFE

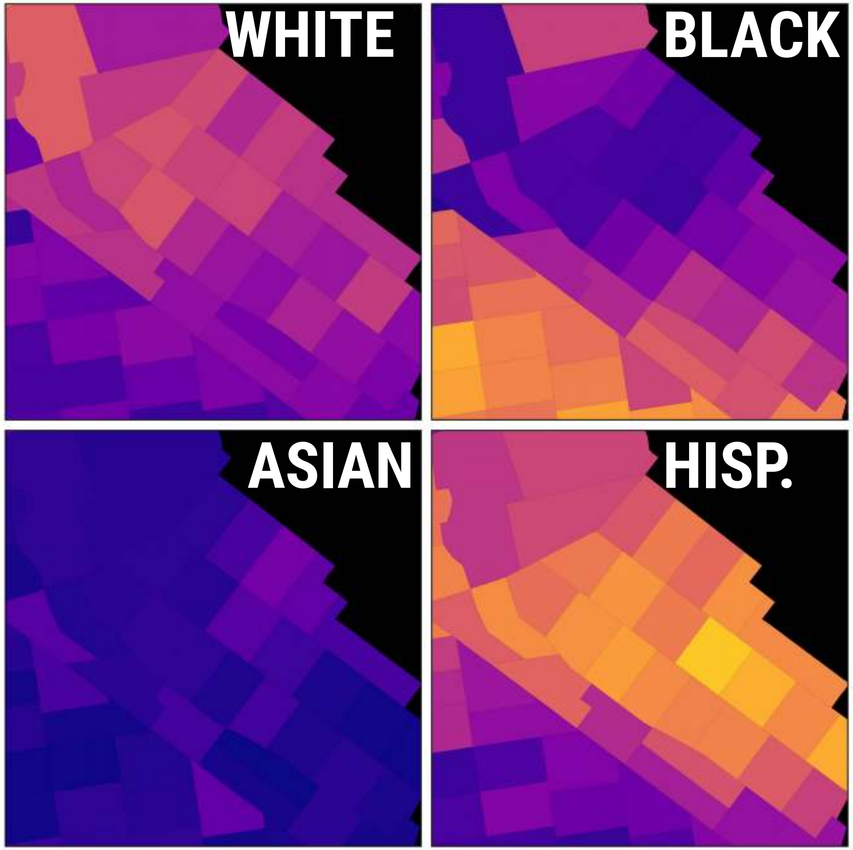
“Williamsburg becomes Greenpoint at the Bushwick Inlet”

“Greenpoint is bordered on the southeast by the BQE”



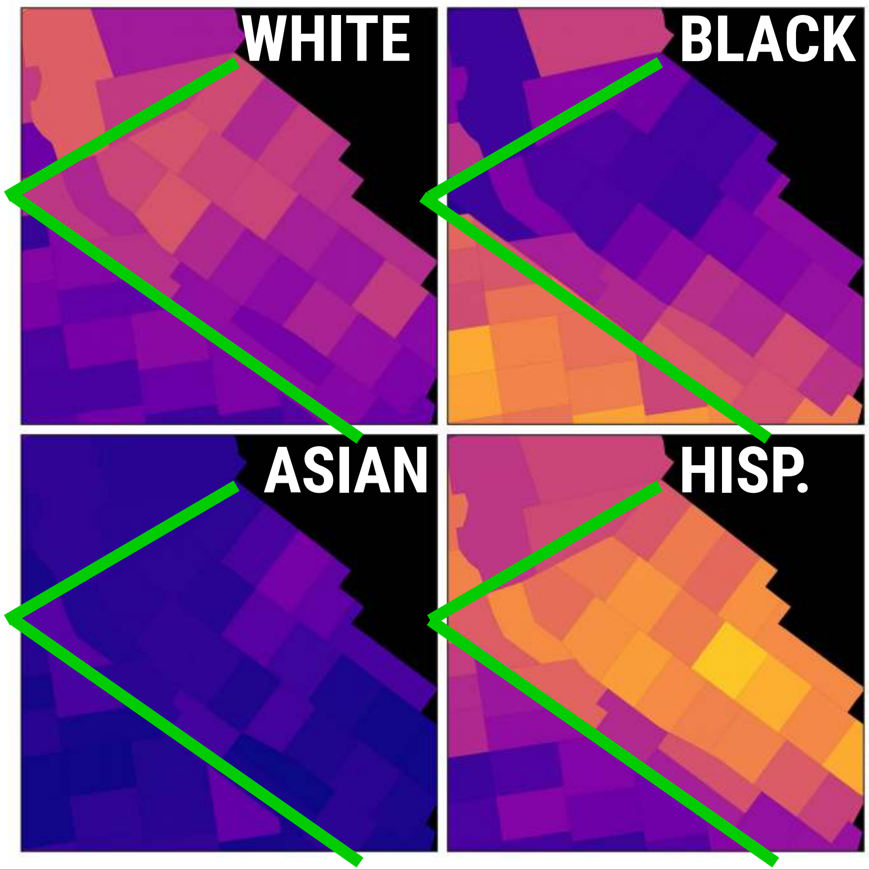
CONCEPTUALIZING BOUNDARIES

"Though an ethnic neighborhood, Bushwick's population is, for a NYC neighborhood, relatively heterogeneous"



CONCEPTUALIZING BOUNDARIES

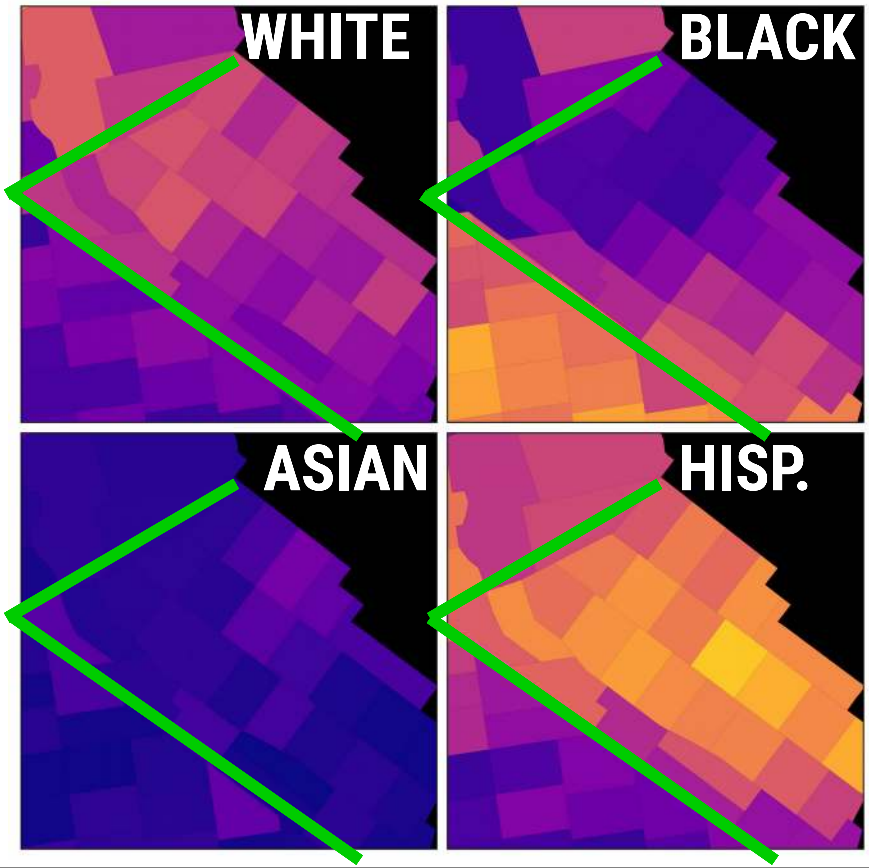
“Though an ethnic neighborhood, Bushwick’s population is, for a NYC neighborhood, relatively heterogeneous”



CONCEPTUALIZING BOUNDARIES

BOUNDARIES AS SOCIALLY CONSTRUCTED DIVISIONS OF URBAN LIFE

“Though an ethnic neighborhood, Bushwick’s population is, for a NYC neighborhood, relatively heterogeneous”



CONCEPTUALIZING BOUNDARIES

BOUNDARIES AS SOCIALLY CONSTRUCTED DIVISIONS OF URBAN LIFE

"Though an ethnic neighborhood, Bushwick's population is, for a NYC neighborhood, relatively heterogeneous"

SCHELLING (1971) Selective segregation

SUTTLES (1972) Defended communities

GRIGSBY (1987) Real income is everything

GRANNIS (1998) Transit network barriers

GALSTER (2001) House Attribute "bundles"

HEDMAN et al. (2011) Choice geographies

HIPP & BOESSEN (2013) Access areas

LEGEWIE & SCHAEFFER (2016) Friction

KWAN (2018) Contingent social contexts

DEAN (2019) Social frontiers

CONCEPTUALIZING BOUNDARIES

BOUNDARIES AS SOCIALLY CONSTRUCTED DIVISIONS OF URBAN LIFE

"Though an ethnic neighborhood, Bushwick's population is, for a NYC neighborhood, relatively heterogeneous"

SCHELLING (1971) Selective segregation

SUTTLES (1972) Defended communities

GRIGSBY (1987) Real income is everything

GRANNIS (1998) Transit network barriers

GALSTER (2001) House Attribute "bundles"

HEDMAN et al. (2011) Choice geographies

HIPP & BOESSEN (2013) Access areas

LEGEWIE & SCHAEFFER (2016) Friction

KWAN (2018) Contingent social contexts

DEAN (2019) Social frontiers

Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict[1]

Joscha Legewie
Yale University

Merlin Schaeffer
University of Cologne

"We propose the *contested boundaries hypothesis* … conflict arises at poorly-defined boundaries that separate ethnic and racial groups … because [boundaries] threaten homogeneous community life and foster ambiguities about group rank."

Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict[1]

Joscha Legewie
Yale University

Merlin Schaeffer
University of Cologne

"We propose the *contested boundaries hypothesis* … conflict arises at poorly-defined boundaries that separate ethnic and racial groups … because **[boundaries] threaten homogeneous community life** and foster ambiguities about group rank."

Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict[1]

Joscha Legewie
Yale University

Merlin Schaeffer
University of Cologne

"We propose the *contested boundaries hypothesis* ... conflict arises at poorly-defined boundaries that separate ethnic and racial groups ... because [boundaries] threaten homogeneous community life and foster ambiguities about group rank."

Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict[1]

Joscha Legewie
Yale University

Merlin Schaeffer
University of Cologne

Communities are neighborhoods, territories that delimit a social group.

When territory is unclear, communities come into conflict.

SPACE

PLACE

SPACE PLACE

Understanding the New Human Dynamics in Smart Spaces and Places: Toward a Splatial Framework

Shih-Lung Shaw* and Daniel Sui†

*Department of Geography, University of Tennessee
†Department of Geosciences, University of Arkansas

SPACE

PLACE

SPACE

The geographic system over which objects of study are related.

- ***Earth Surface***
- ***Road Systems***
- ***Social Networks***
- ***Economic Relations***

PLACE

Geographic entities that are constructed by distinctiveness.

- ***Regions***
- ***Neighborhoods***
- ***Home/Staying locales***
- ***Functional classifications***

SPACE

The geographic system over which objects of study are related.

- *Earth Surface*
- *Road Systems*
- *Social Networks*
- *Economic Relations*

PLACE

Geographic entities that are constructed by distinctiveness.

Geographic information science II:
less space, more places in smart cities

Stéphane Roche

Digital neighborhoods

Luc Anselin[a*] and Sarah Williams[b]

Towards the statistical analysis and visualization of places

René Westerholt et al.

SPACE

The geographic system over which objects of study are related.

- ***Earth Surface***
- ***Road Systems***
- ***Social Networks***
- ***Economic Relations***

PLACE

Geographic entities that are constructed by distinctiveness.

How or why do they emerge?
What are their properties?
What are their purpose?
Do they have effects on things we care about?

SPACE

The geographic system over which objects of study are related.

How do things interact?
Over what spatial systems?
In what manner?
What impact do entities have on others nearby?

PLACE

Geographic entities that are constructed by distinctiveness.

How or why do they emerge?
What are their properties?
What are their purpose?
Do they have effects on things we care about?

SPACE

The geographic system over which objects of study are related.

Boundary:

a division or discontinuity in the field of interactions.

PLACE

Geographic entities that are constructed by distinctiveness.

Boundary:

where one places becomes distinct from another.

SPACE

The geographic system over which objects of study are related.

Bounda

a di

in th

PLACE

Geographic entities that are constructed by distinctiveness.

Article

B Urban Analytics and City Science

EPB: Urban Analytics and City Science
0(0) 1–19

DOI: 10.1177/2399808319875752
journals.sagepub.com/home/epb
SAGE

Geosilhouettes: Geographical measures of cluster fit

Levi J Wolf
School of Geographical Sciences, University of Bristol, UK

Elijah Knaap and Sergio Rey
Center for Geospatial Sciences, University of California Riverside, USA

BOUNDARIES

AN EMINENTLY-GEOGRAPHICAL CONSTRUCT

GOODNESS OF FIT: THE SILHOUETTE

SIMILARITY IN A COUNTERFACTUAL

WHOSE “GOOD” IS IT ANYWAY?

EXAMPLE: BROOKLYN

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

ROUSSEEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

AVERAGE DISSIMILARITY FROM i

TO EVERYONE ELSE IN ITS CURRENT CLUSTER c

ROUSSEEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

AVERAGE DISSIMILARITY FROM i TO EVERYONE ELSE IN SOME OTHER CLUSTER k

ROUSSEEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\left\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\right\}}$$

AVERAGE DISSIMILARITY FROM i TO EVERYONE ELSE IN THE SECOND-BEST CHOICE CLUSTER

ROUSSEEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

DIFFERENCE IN SIMILARITY BETWEEN CURRENT CLUSTER AND NEXT BEST FIT CLUSTER

ROUSSEEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

NORMALIZING FACTOR SO THAT $s(i)$ IS LIKE A CORRELATION COEFFICIENT

ROUSSEEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

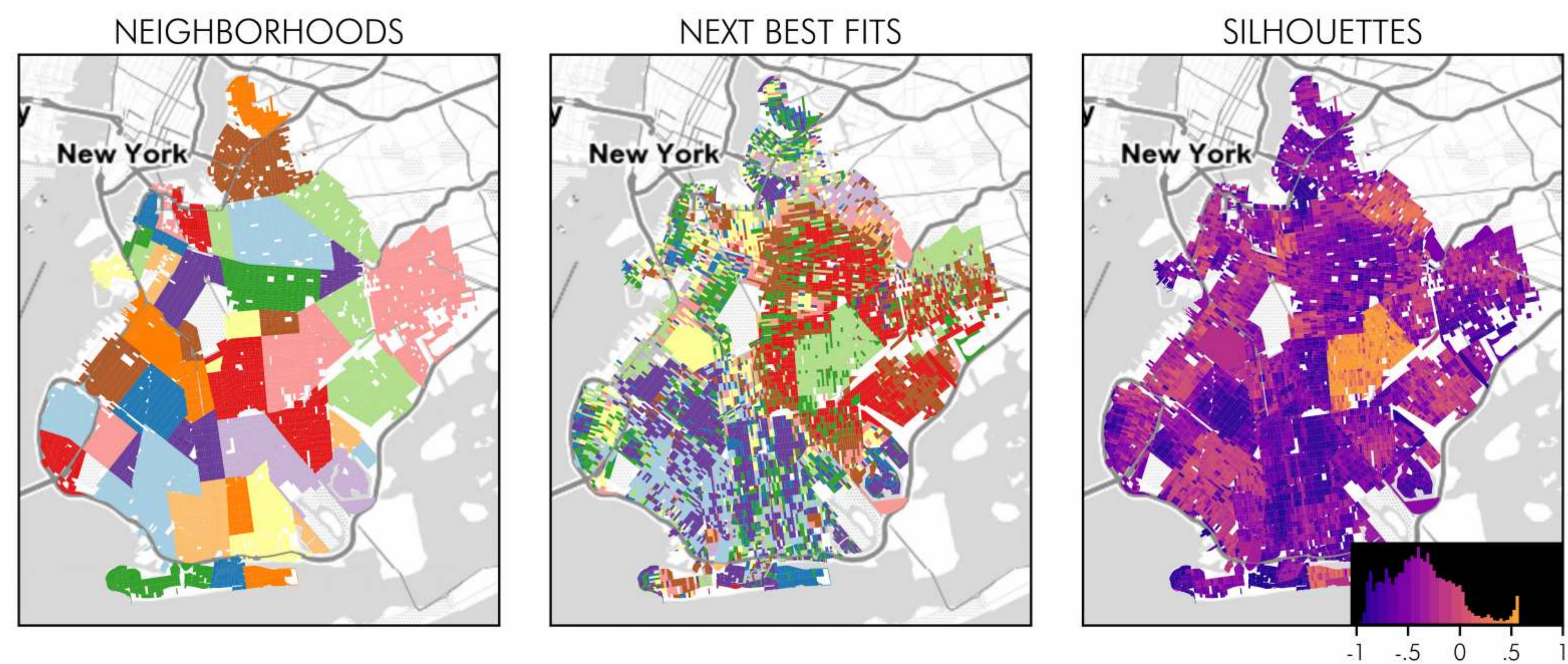
Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

HOW MUCH MORE SIMILAR IS i TO ITS SECOND CHOICE CLUSTER THAN TO ITS CURRENT CLUSTER?

ROUSSEEUW (1987)

SILHOUETTE SCORES



ROUSSEUW (1987)

SILHOUETTE SCORES

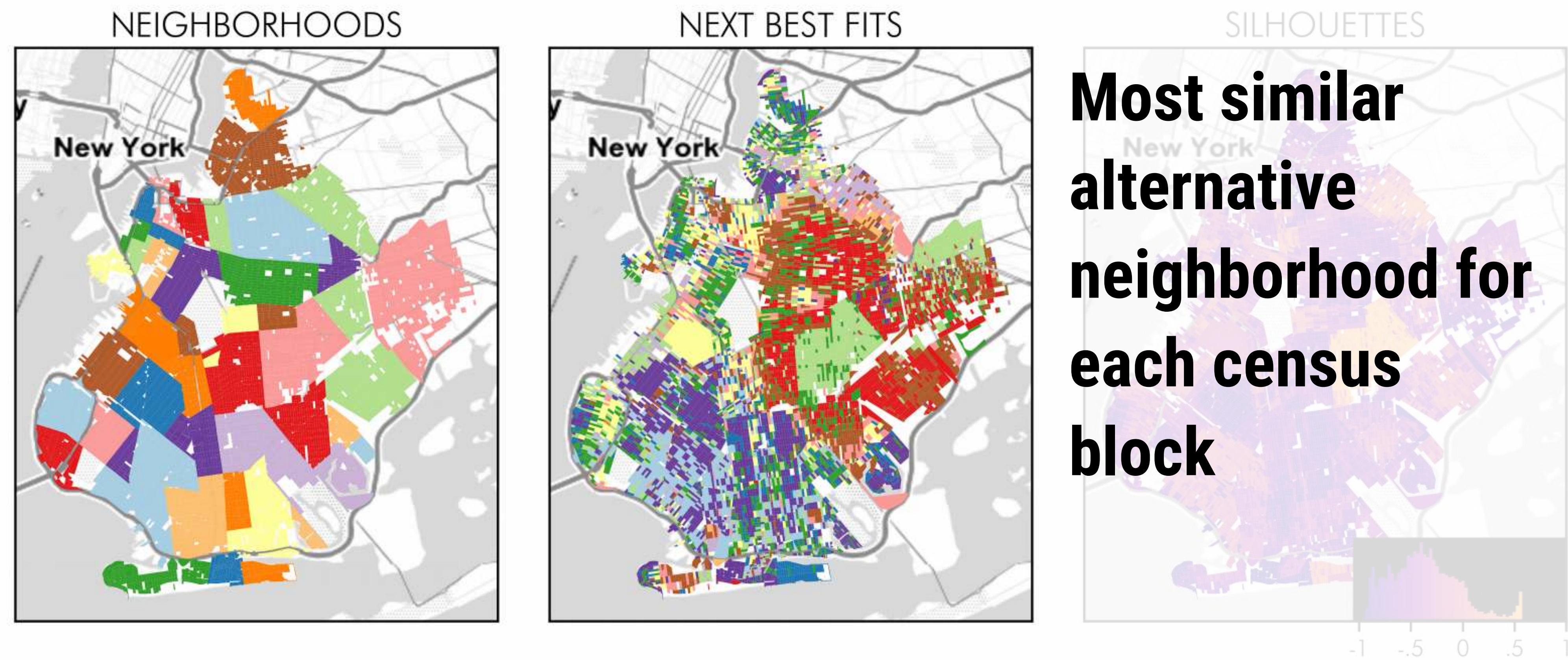
Zillow neighborhoods built from

online housing markets

official boundaries (NYCTA)

ROUSSEEUW (1987)

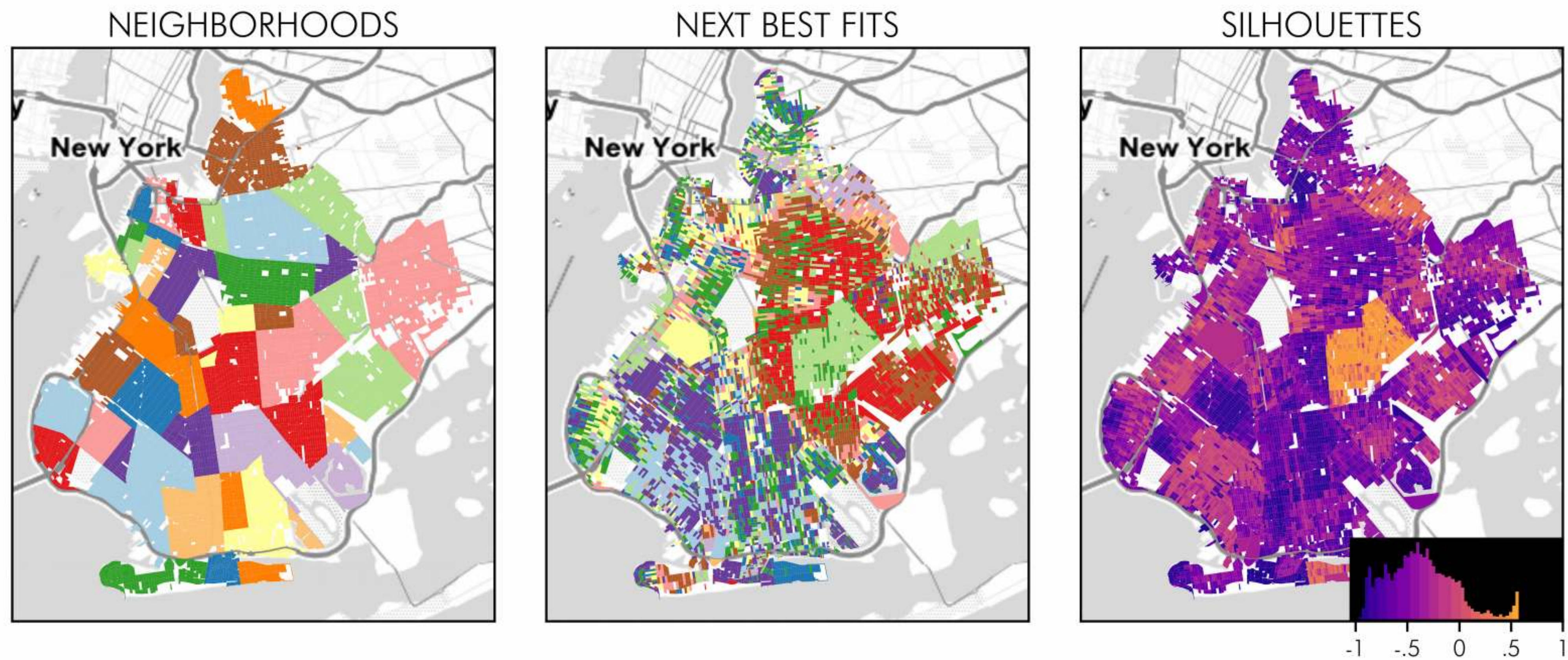
SILHOUETTE SCORES



Most similar alternative neighborhood for each census block

ROUSSEEUW (1987)

SILHOUETTE SCORES



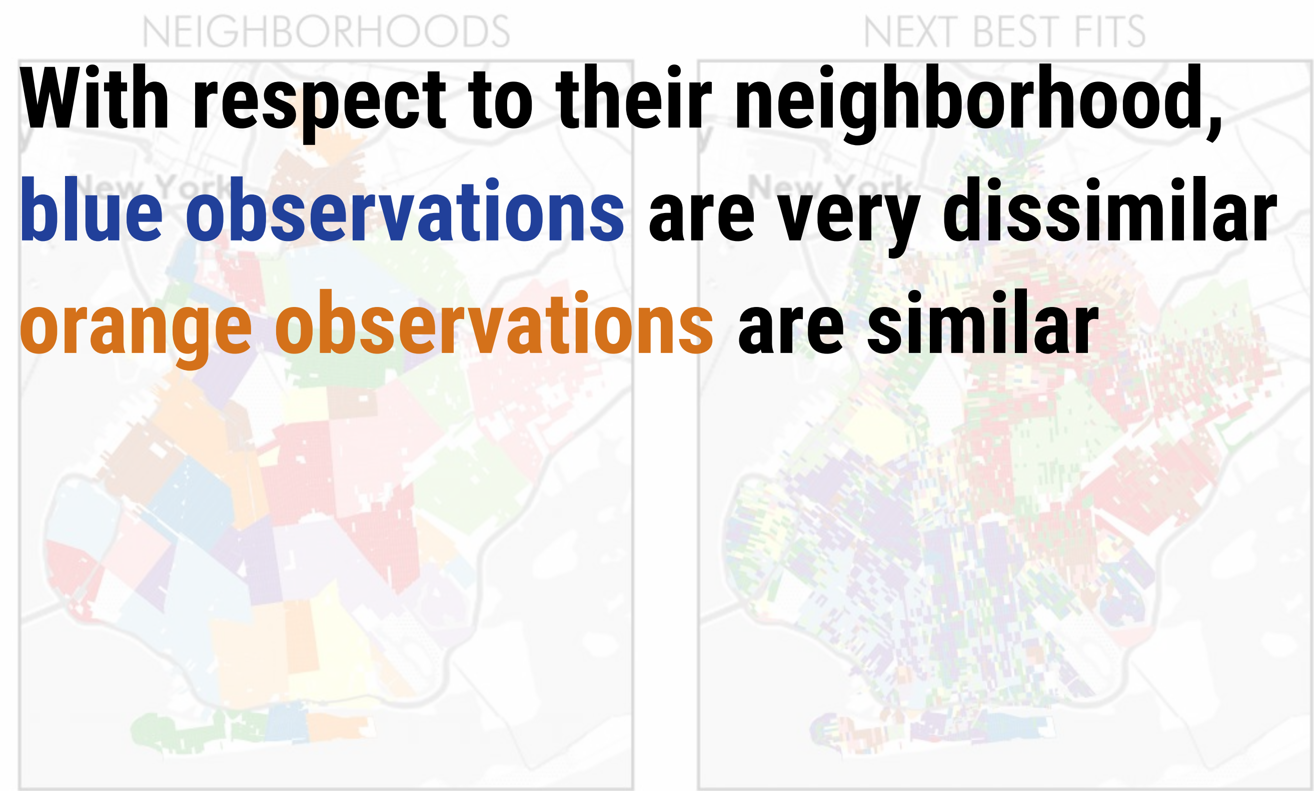
ROUSSEEUW (1987)

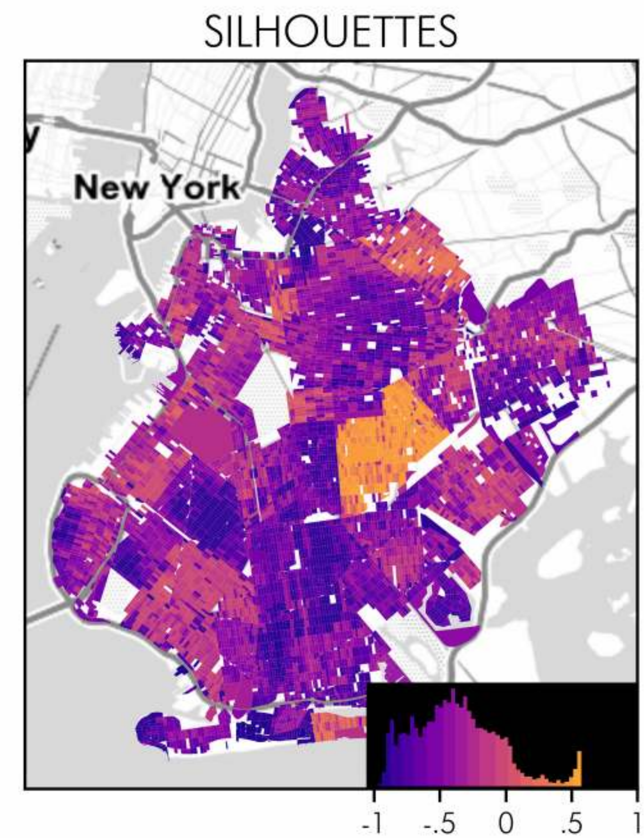
SILHOUETTE SCORES

With respect to their neighborhood,

blue observations are very dissimilar

orange observations are similar





ROUSSEEUW (1987)

BOUNDARIES

AN EMINENTLY-GEOGRAPHICAL CONSTRUCT

GOODNESS OF FIT: THE SILHOUETTE

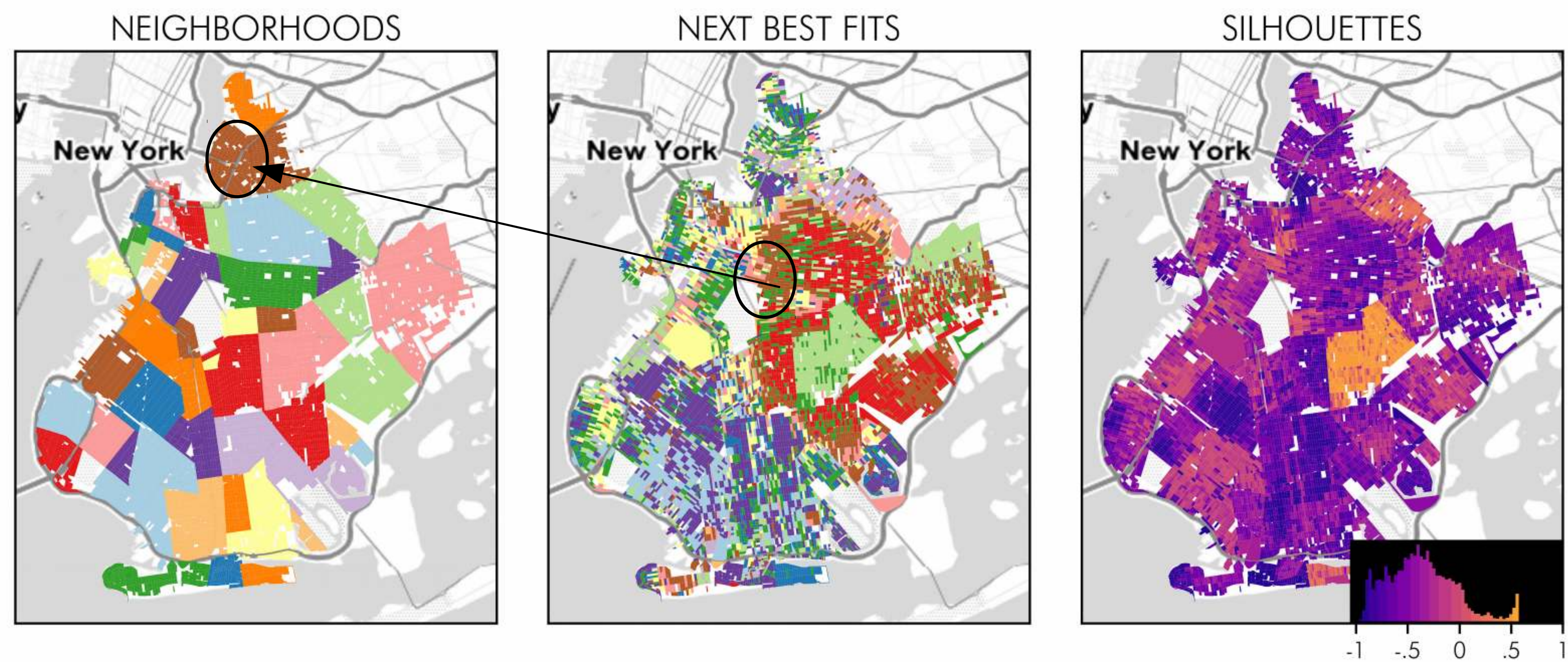
SIMILARITY IN A COUNTERFACTUAL

WHOSE “GOOD” IS IT ANYWAY?

GEOSILHOUETTES: MAKING SPACE FOR BOUNDARIES

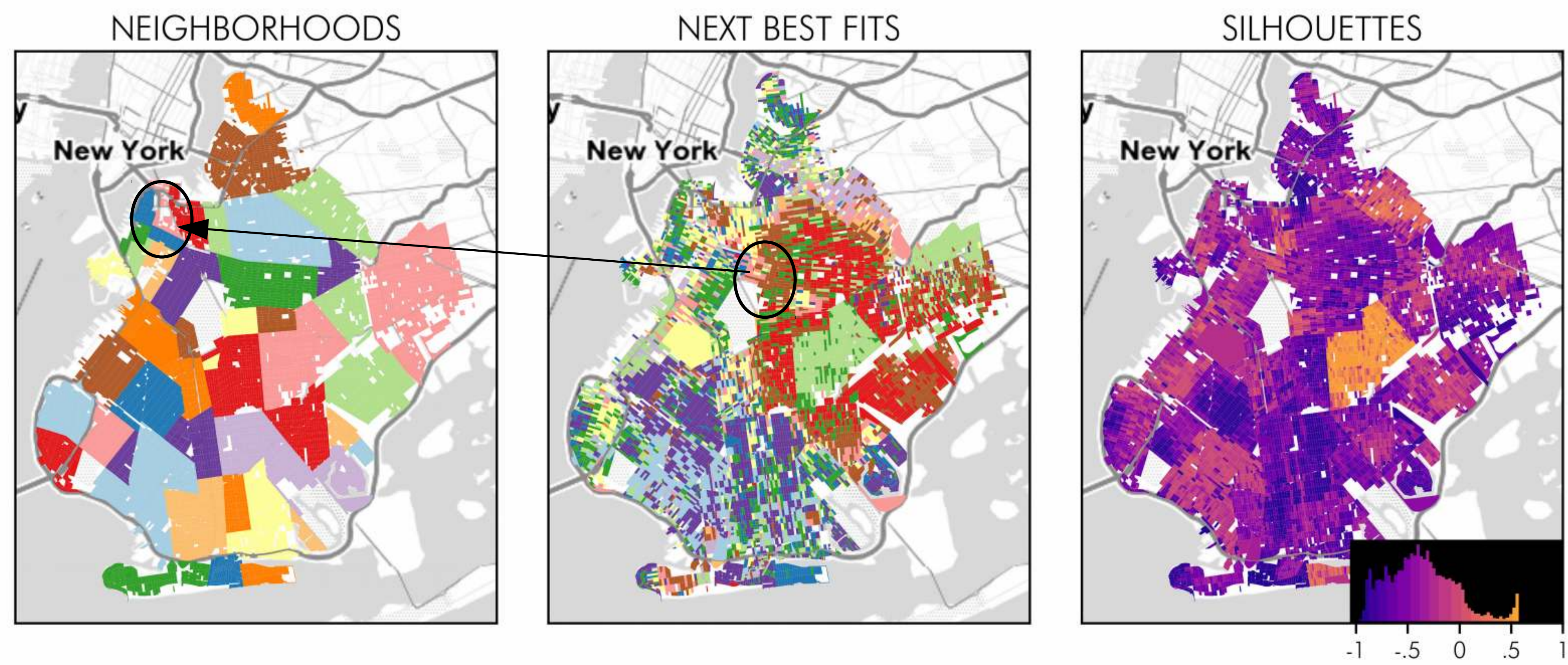
EXAMPLE: BROOKLYN

SILHOUETTE SCORES



ROUSSEEUW (1987)

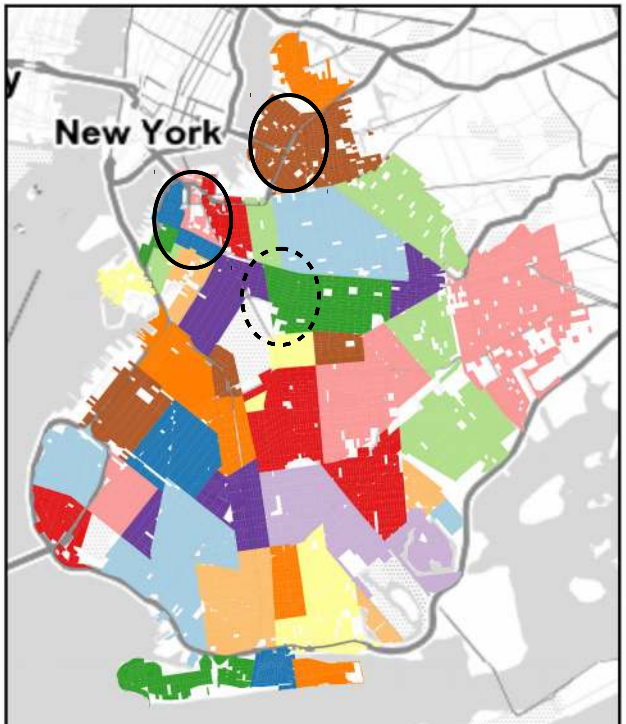
SILHOUETTE SCORES



ROUSSEEUW (1987)

SILHOUETTE SCORES

NEIGHBORHOODS



NEXT BEST FITS

SILHOUETTES

Williamsburg is (relatively) far away from south Crown Heights.

Census blocks can't really "move," so the second choice cluster isn't "real."

ROUSSEUW (1987)

PATH
SILHOUETTE

Williamsburg is (relatively) far away from south Crown Heights.

BOUNDARY
SILHOUETTE

Census blocks can’t really “move,” so the second choice cluster isn’t “real.”

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

HOW MUCH MORE SIMILAR IS i TO ITS SECOND CHOICE CLUSTER THAN TO ITS CURRENT CLUSTER?

ROUSSEUW (1987)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Then, make a few friends:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

REDEFINE d SO THAT IT INCLUDES GEOGRAPHY!

WOLF et al. (2019)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Say that observation i is embedded in a graph G

G has an adjacency matrix, W, where

$w_{ij} = 1$ if i is connected to j, zero otherwise.

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Say that observation i is embedded in a graph G

G has an adjacency matrix, W, where

$w_{ij} = 1$ if i is connected to j, zero otherwise.

$$\mathbf{C}_1 = \mathbf{D} \circ \mathbf{W}$$

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Say that observation i is embedded in a graph G

G has an adjacency matrix, W, where

$w_{ij} = 1$ if i is connected to j, zero otherwise.

C_1 **DISSIMILARITY BETWEEN ADJACENT OBSERVATIONS**

WOLF et al. (2019)

SILHOUETTE SCORES

Say that observation i is assigned to cluster c

Say that observation i is embedded in a graph G

G has an adjacency matrix, W, where

$w_{ij} = 1$ if i is connected to j, zero otherwise.

$\mathbf{C}$ **DISSIMILARITY "COST" TO CONNECT ANY TWO OBSERVATIONS**

PATH SILHOUETTE

Say that i in G is assigned to cluster c

The PATH SILHOUETTE is:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

WHERE DISTANCES USE C

PATH SILHOUETTE

Say that i in G is assigned to cluster c

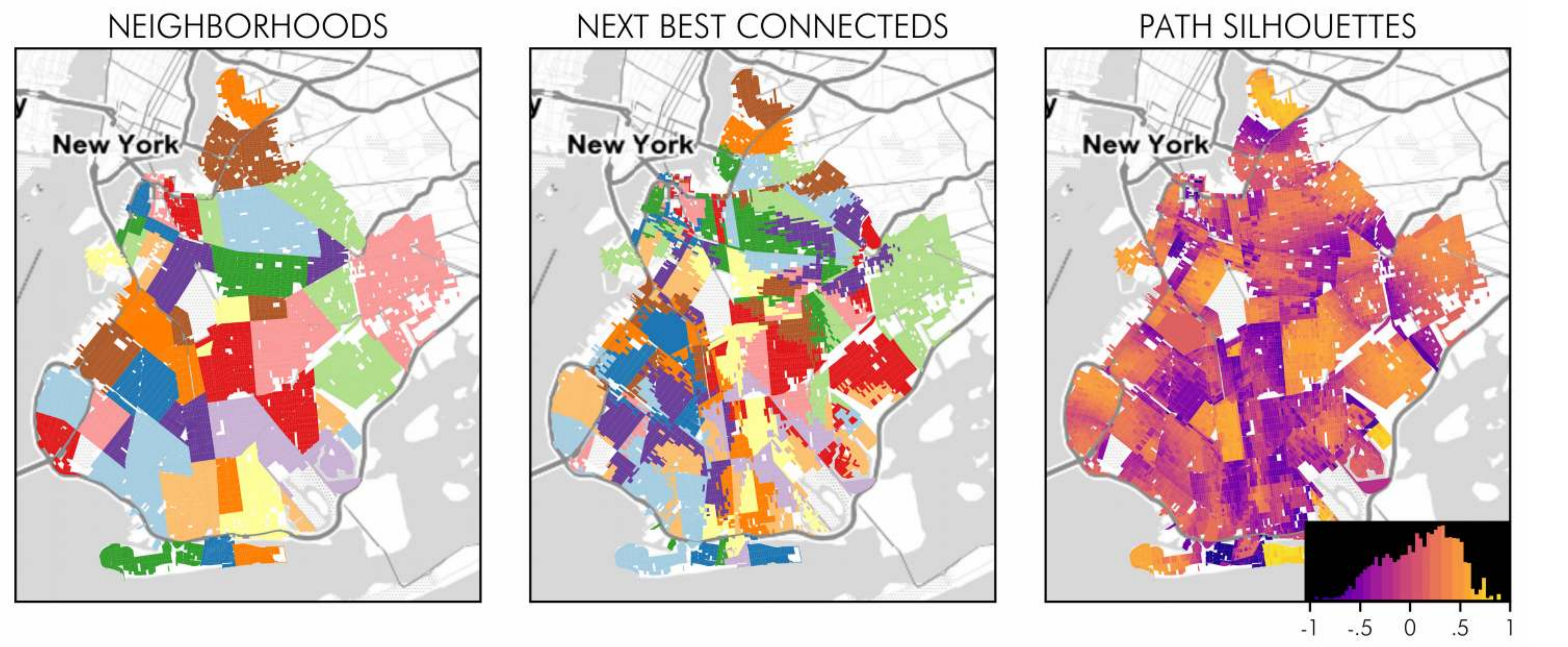
The PATH SILHOUETTE is:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

HOW MUCH MORE SIMILAR IS i TO k THAN TO c

WHEN PROXIMITY MATTERS FOR SIMILARITY?

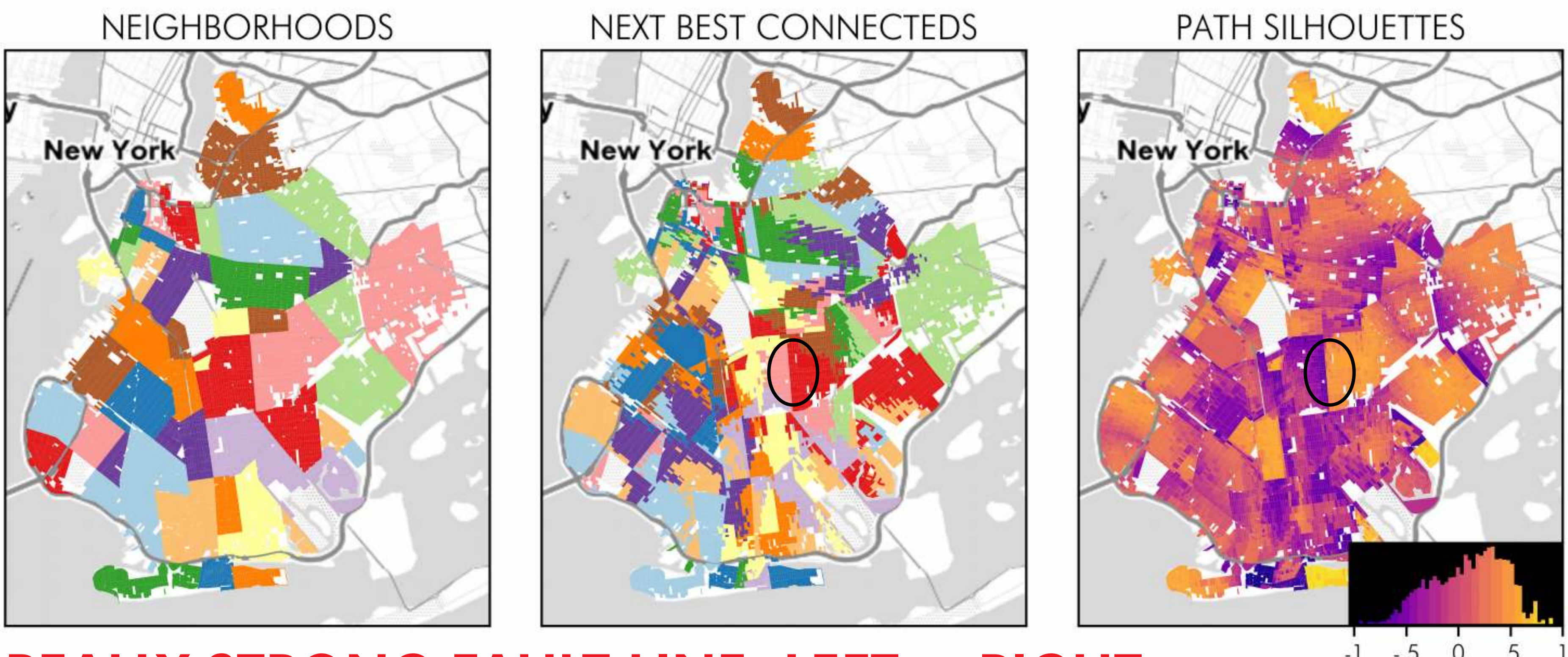
PATH SILHOUETTE



PATH : REMOTENESS x SIMILARITY

WOLF et al. (2019)

PATH SILHOUETTE

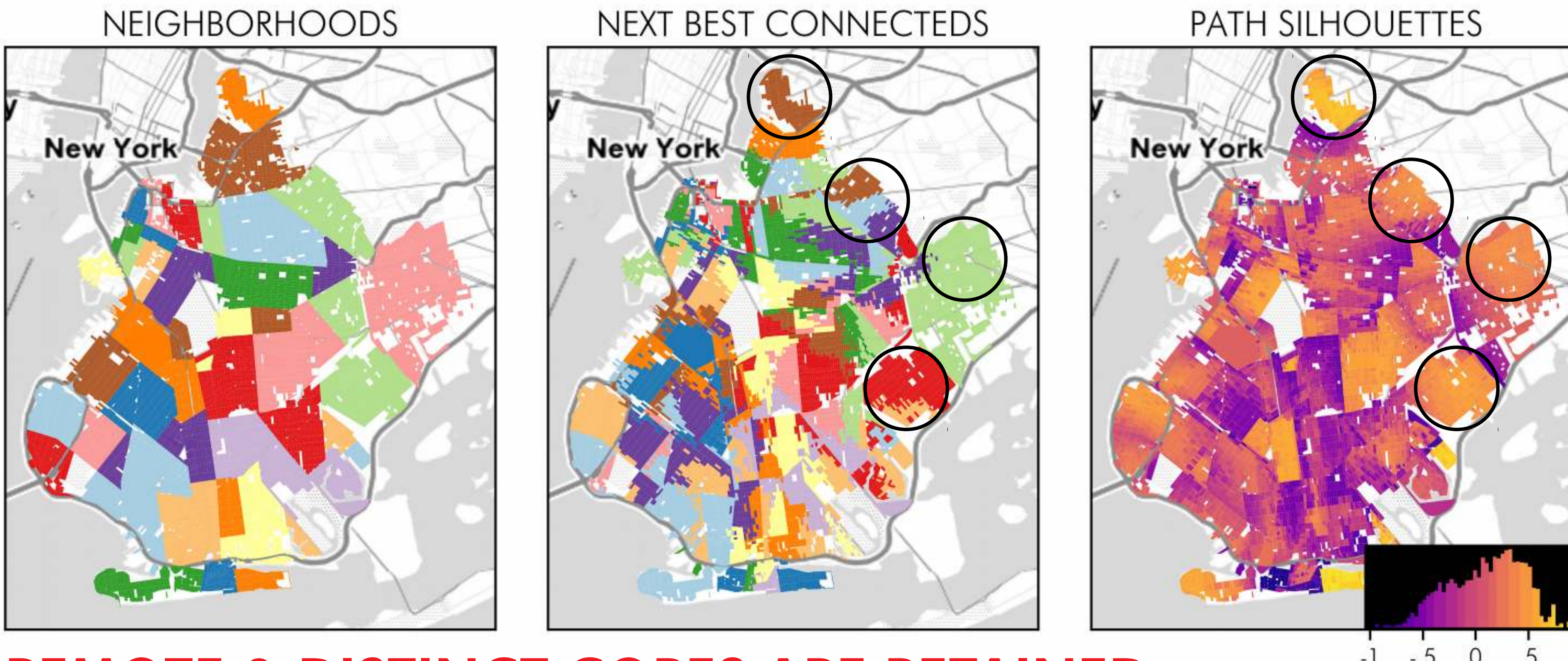


REALLY STRONG FAULT LINE: LEFT → RIGHT

PATH : REMOTENESS x SIMILARITY

WOLF et al. (2019)

PATH SILHOUETTE



REMOTE & DISTINCT CORES ARE RETAINED

PATH : REMOTENESS x SIMILARITY

WOLF et al. (2019)

PATH SILHOUETTE

Williamsburg is (relatively) far away from south Crown Heights.

BOUNDARY SILHOUETTE

Census blocks can’t really “move,” so the second choice cluster isn’t “real.”

BOUNDARY SILHOUETTE

Say that i in G is assigned to cluster c

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

HOW MUCH MORE SIMILAR IS i TO ITS SECOND CHOICE CLUSTER THAN TO ITS CURRENT CLUSTER?

BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE

Say that i in G is assigned to cluster c

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

SECOND CHOICE CLUSTERS ARE ONLY FEASIBLE IF i COULD BE MOVED W/O BREAKING c OR k

BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE

Say that i in G is assigned to cluster c

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

A FEASIBLE SECOND CHOICE CLUSTER IS CALLED THE BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE

Say that i in G is assigned to cluster c, has BLA k

The BOUNDARY SILHOUETTE is:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

WHERE k IS RESTRICTED TO BE

A FEASIBLE (i.e. local) REASSIGNMENT

BOUNDARY SILHOUETTE

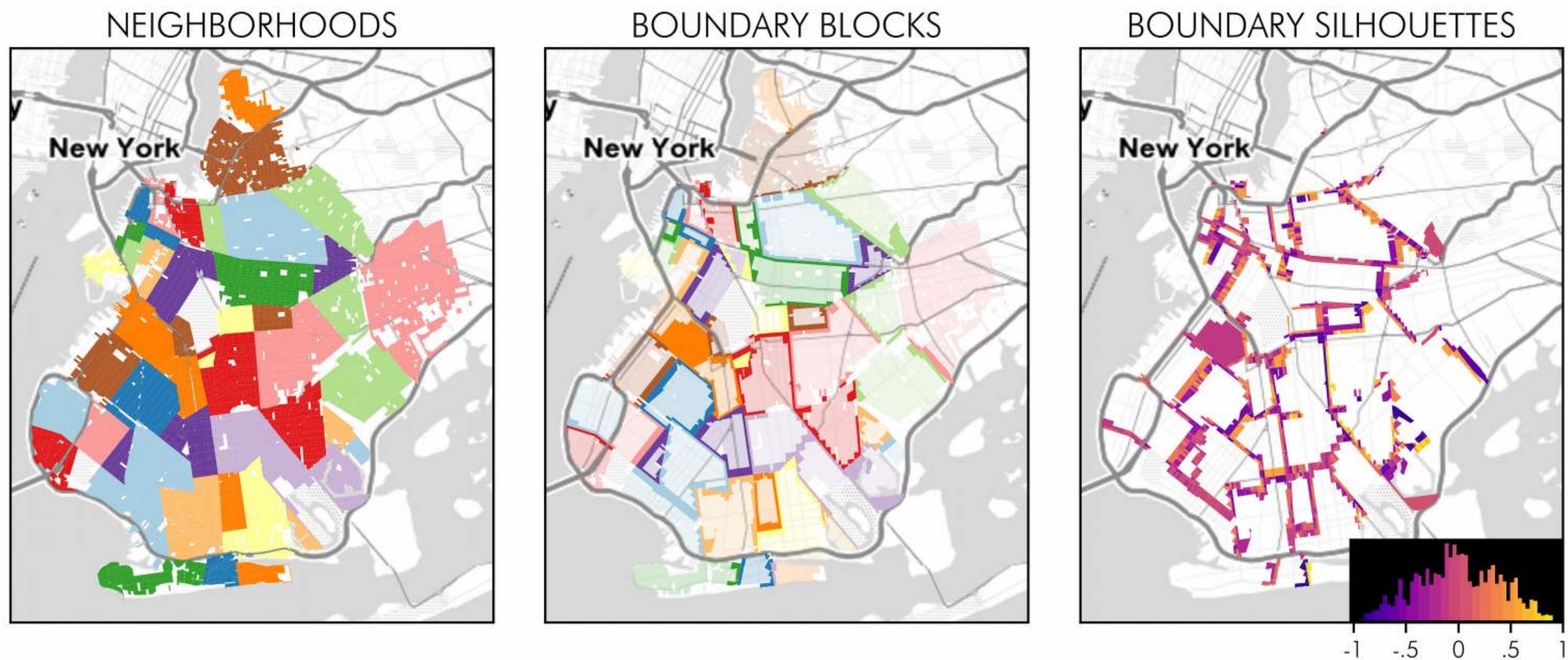
Say that i in G is assigned to cluster c, has **BLA** k

The BOUNDARY SILHOUETTE is:

$$s(i) = \frac{\min\left\{\bar{d}_k(i)\right\} - \bar{d}_c(i)}{\max\{\min\left\{\bar{d}_k(i)\right\}, \bar{d}_c(i)\}}$$

HOW MUCH MORE SIMILAR IS i TO c THAN TO THE BEST LOCAL ALTERNATIVE?

BOUNDARY SILHOUETTE

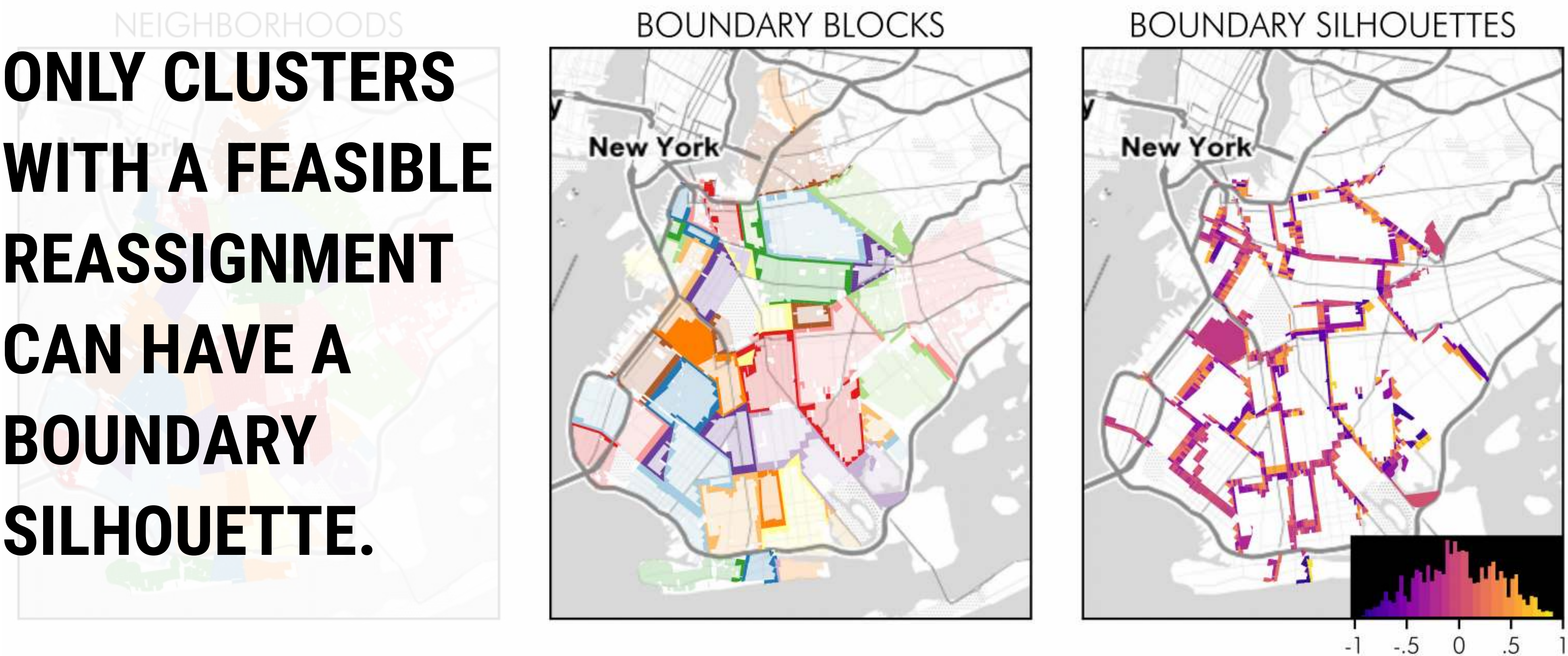


BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE

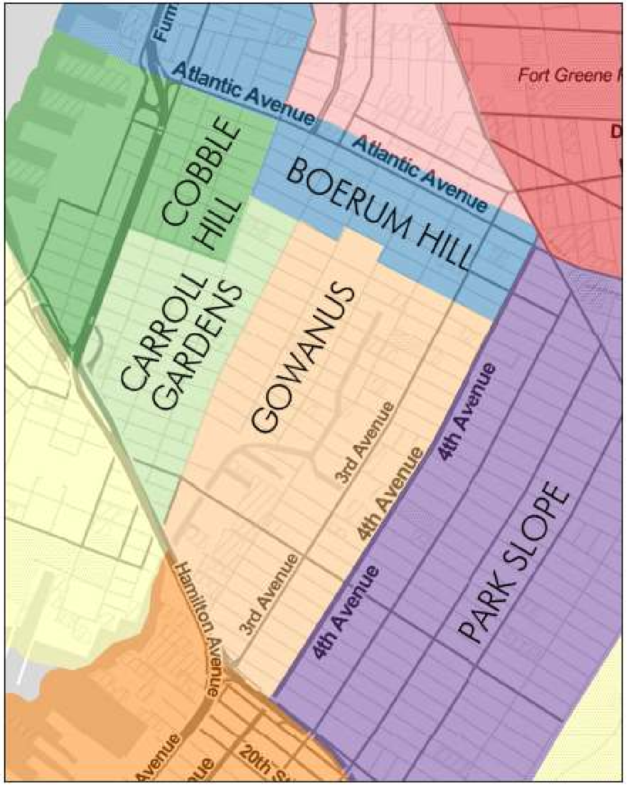
ONLY CLUSTERS WITH A FEASIBLE REASSIGNMENT CAN HAVE A BOUNDARY SILHOUETTE.



BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

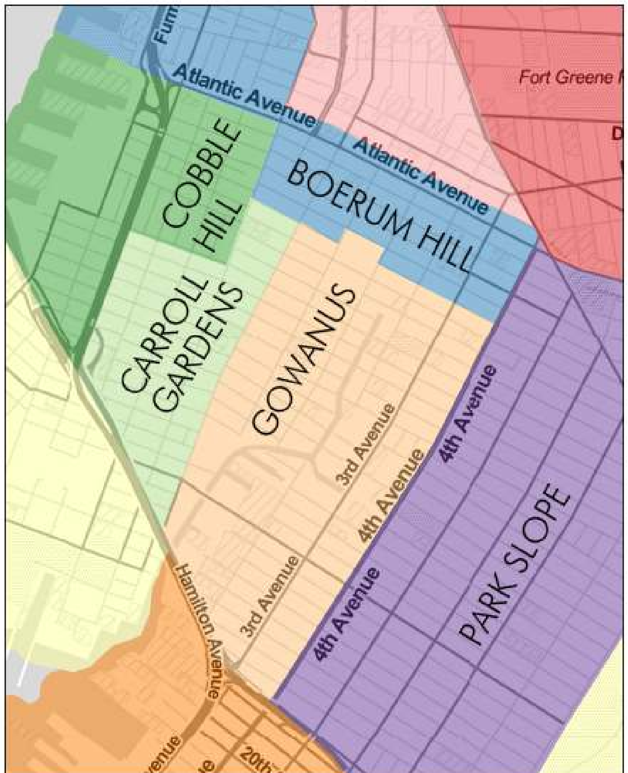
BOUNDARY SILHOUETTE



BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE



BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE

neighbor focal	Boerum Hill	Cobble Hill	Carroll Gardens	Gowanus	Park Slope
Boerum Hill	0.000	-0.32	-0.358	0.274	0.122
Cobble Hill	0.627	0	-0.156	0.639	-
Carroll Gardens	0.339	0.152	0	0.710	-
Gowanus	-0.071	-0.359	-0.647	0.000	-0.168
Park Slope	0.050	-	-	0.390	0

On the Gowanus side, blocks are much more similar to those in Carroll Gardens.

WOLF et al. (2019)

BOUNDARY SILHOUETTE

neighbor focal	Boerum Hill	Cobble Hill	Carroll Gardens	Gowanus	Park Slope
Boerum Hill	0.000	-0.32	-0.358	0.274	0.122
Cobble Hill	0.627	0	-0.156	0.639	-
Carroll Gardens	0.339	0.152	0	0.710	-
Gowanus	-0.071	-0.359	-0.647	0.000	-0.168
Park Slope	0.050	-	-	0.390	0

On the Gowanus side, blocks are much more similar to those in Carroll Gardens.
On the Carroll Gardens side, blocks are much more similar to Carroll Gardens.

WOLF et al. (2019)

BOUNDARY SILHOUETTE

neighbor focal	Boerum Hill	Cobble Hill	Carroll Gardens	Gowanus	Park Slope
Boerum Hill	0.000	-0.32	-0.358	0.274	0.122
Cobble Hill	0.627	0	-0.156	0.639	-
Carroll Gardens	0.339	0.152	0	0.710	-
Gowanus	-0.071	-0.359	-0.647	0.000	-0.168
Park Slope	0.050	-	-	0.390	0

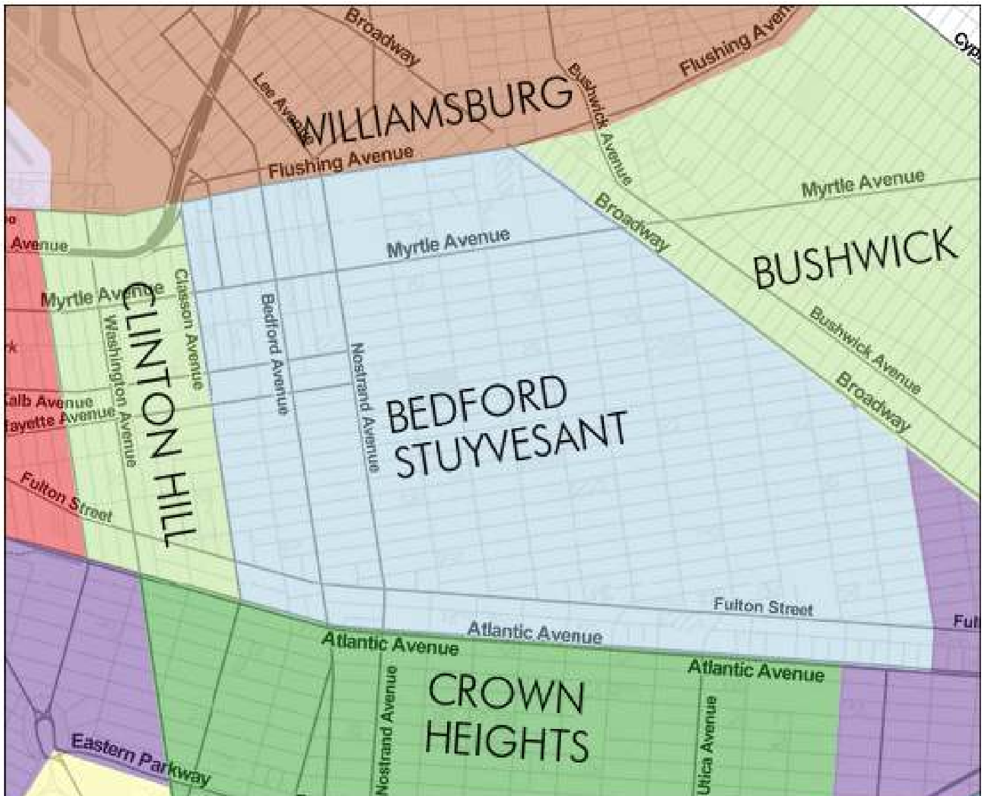
On the Gowanus side, blocks are much more similar to those in Carroll Gardens.

On the Carroll Gardens side, blocks are much more similar to Carroll Gardens.

The boundary is asymmetric! Gowanus border leans towards Carroll Gardens.

WOLF et al. (2019)

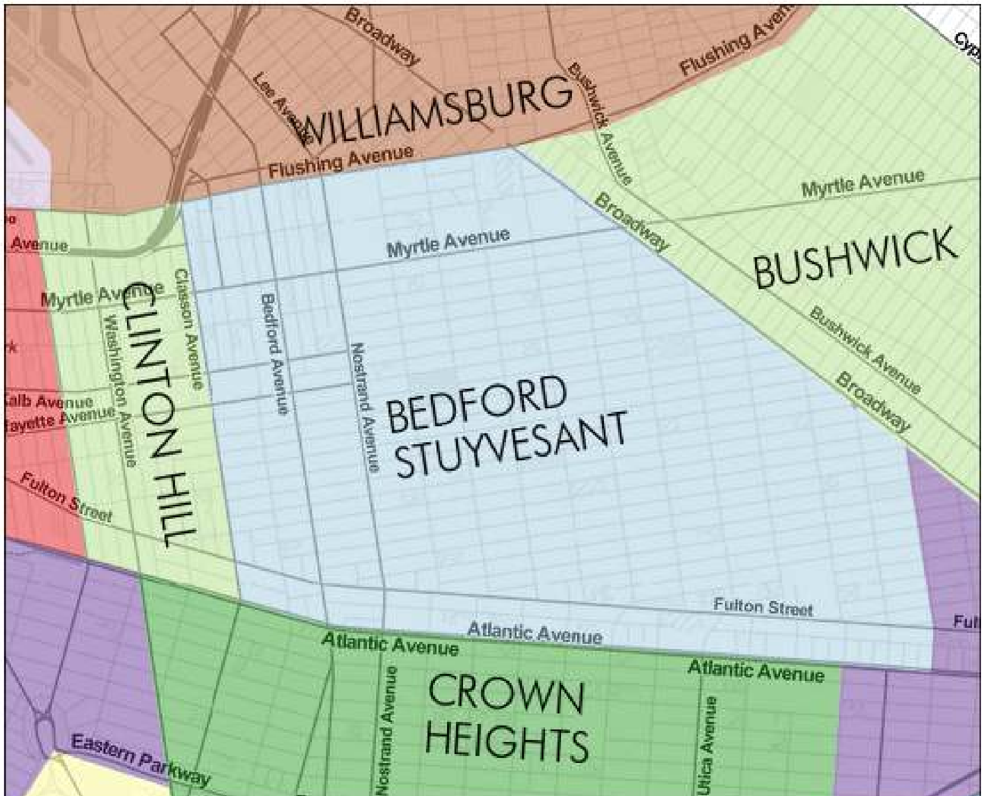
BOUNDARY SILHOUETTE



BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE



BOUNDARY: BEST LOCAL ALTERNATIVE

WOLF et al. (2019)

BOUNDARY SILHOUETTE

neighbor focal	Williamsburg	Bushwick	Bedford Stuyvesant	Clinton Hill	Crown Heights
Williamsburg	0	-0.096	0.693	0.516	-
Bushwick	0.288	0	0.482	-	-
Bedford Stuyvesant	-0.478	0.198	0.000	0.006	-0.059
Clinton Hill	-0.355	-	0.358	0	0.296
Crown Heights	-	-	0.077	-0.427	0

On the BedStuy side, blocks remain slightly more similar to blocks in BedStuy.

BOUNDARY SILHOUETTE

neighbor focal	Williamsburg	Bushwick	Bedford Stuyvesant	Clinton Hill	Crown Heights
Williamsburg	0	-0.096	0.693	0.516	-
Bushwick	0.288	0	0.482	-	-
Bedford Stuyvesant	-0.478	0.198	0.000	0.006	-0.059
Clinton Hill	-0.355	-	0.358	0	0.296
Crown Heights	-	-	0.077	-0.427	0

On the BedStuy side, blocks remain slightly more similar to blocks in BedStuy.
On the Bushwick side, blocks are more similar to blocks in Bushwick.

WOLF et al. (2019)

BOUNDARY SILHOUETTE

neighbor focal	Williamsburg	Bushwick	Bedford Stuyvesant	Clinton Hill	Crown Heights
Williamsburg	0	-0.096	0.693	0.516	-
Bushwick	0.288	0	0.482	-	-
Bedford Stuyvesant	-0.478	0.198	0.000	0.006	-0.059
Clinton Hill	-0.355	-	0.358	0	0.296
Crown Heights	-	-	0.077	-0.427	0

On the BedStuy side, blocks remain slightly more similar to blocks in BedStuy.

On the Bushwick side, blocks are more similar to blocks in Bushwick.

The boundary is symmetric! Not a "contested boundary"!

WOLF et al. (2019)

SILHOUETTES

Numerically robust

Multidimensional

Not “predictive” of another variate

Straightforward interpretation

SILHOUETTES

Numerically robust

Multidimensional

Not “predictive” of another variate

Straightforward interpretation

Not probabilistic (no “certainty” about strength)

Compares place boundaries, not spatial boundaries

BOUNDARIES

AN EMINENTLY-GEOGRAPHICAL CONSTRUCT

GOODNESS OF FIT: THE SILHOUETTE

SIMILARITY IN A COUNTERFACTUAL

GEOSILHOUETTES

PATH SILHOUETTE: REMOTENESS & SIMILARITY

BOUNDARY SILHOUETTE: ADJACENCY & DIRECTION

FINDING THE FAULT LINES:

DETECTING URBAN SOCIAL BOUNDARIES USING SOCIAL DATA SCIENCE

NSF-SES 1733705

University of Bristol
UC Riverside (University of California)
SOC ARXIV osf.io/vd3uk

LEVI JOHN WOLF
ELIJAH KNAAP
SERGIO REY

levi.john.wolf@bristol.ac.uk
@levijohnwolf
ljwolf.org