A New Optimal Matching Approach to Uncovering Neighborhood Sequencing Structure

Wei Kang, Sergio Rey
Center for Geospatial Sciences, University of California Riverside

Presented at NARSC, Nov 16 2019, Pittsburgh PA
This Research is supported by the National Science Foundation under grant 1733705 - Neighborhoods in Space-Time Contexts

Neighborhood Change Research

- Variable diagram
  - Specific variables: racial composition, median household income
  - Composite indexes: urban deprivation index, opportunity index, socioeconomic status
  - Neighborhood change: value increments
    - Continuous
- Contextual mode of analysis
  - Neighborhoods as ensembles of variables
  - Clustering - discrete neighborhood classifications/types
  - Geodemographic typology
  - Neighborhood change: change of types
    - Discrete
  - Statistically suitable for small area estimates potentially suffering from large margins of errors (e.g. American Community Survey (ACS) census tracts) (Spielman and Singleton, 2015)

Multidimensional neighborhood change

- A white-flight/filtering process
- Densification of neighborhoods: single-family house to multi-family dwellings
- White single-family neighborhood obtain higher socioeconomic status
- Gentrification
- Establishment of a multiethnic/global neighborhood
- Stability: highest poverty, majority black, wealthiest and whitest neighborhoods

(Delmelle, 2017)

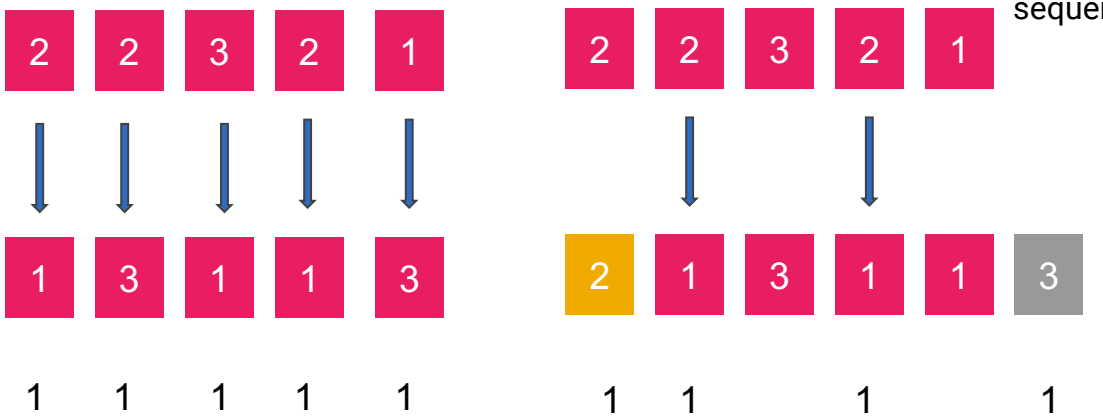
Workflow of Multidimensional Neighborhood Change Analysis

Geodemographic Analysis → Sequence Analysis → Neighborhood sequence typology

Use clustering algorithm to find similar neighborhoods. Each cluster is summarized as a neighborhood type.

To evaluate similarity or distance between every pair of neighborhood sequence. Optimal Matching (OM)

Input the distance matrix of neighborhood sequences into proper clustering algorithm to obtain a neighborhood sequence typology.



Motivation

- Sequence methods origin from DNA matching
- Exploration in its application to the life course research
- Whether the available sequence methods satisfy the needs for neighborhood change research?
    - Theoretical foundation: invasion/succession, neighborhood decline, new pathways of neighborhood change - global neighborhoods?
    - Temporal dimension: contexts?
    - Neighborhoods as social constructs
    - Neighborhoods as spatial constructs
- Current literature: Focus more on the contemporaneous differences

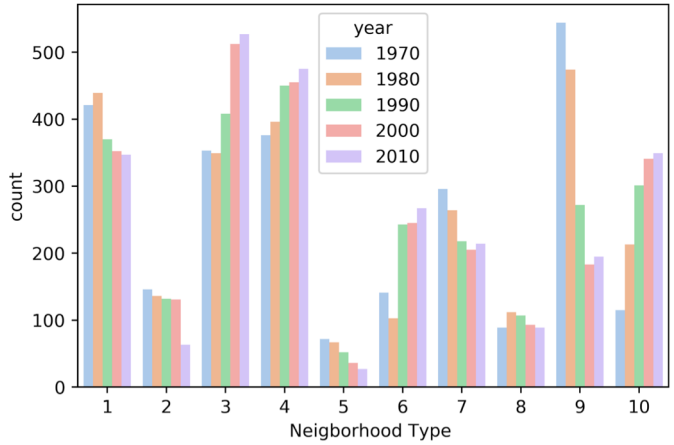
Neighborhood change in LA 1970-2010

- Data: 1970, 1980, 1990, 2000, 2010 census/ACS
- 14 tract-level variables
- Geolytics Neighborhood DataBase 2010 (NCDB 2010)
- POP>500
- 2,553 tracts

TABLE 1 Neighborhoods as ensembles of 14 variables.

Category	Variable	Description
Demographic	CHILD	% persons who are children under 18 years old
	OLD	% persons who are 65+ years old
	SHRWHT	% white population
	SHRBLK	% black/African American population
Socioeconomic	UNEMPRT	% persons 16+ years old who are in the civilian labor force and unemployed
	PRFE	% persons 16+ years old employed in manufacturing, transportation, and public administration
	POVRAT	% total persons below the poverty level last year
	EDUC	% persons 25+ years old with at least a 4-year degree
Housing	BL30OLDPRO	% total housing units built MORE than 30 years ago
	TTMULTI	% total multiunit structures
	YRMV10PRO	% occupied housing units where household heads moved in less than 10 years ago
	MNVALHS	Mean value of specified owner-occupied housing units
	OWNO	% total owner-occupied housing units
	VACHUPRO	% total vacant year-round housing units

Neighborhood Typology (10)

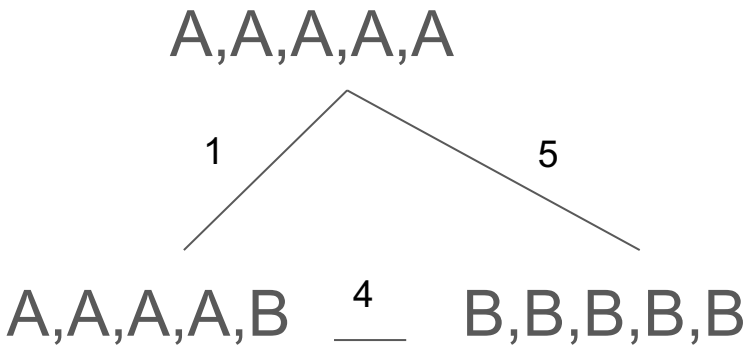


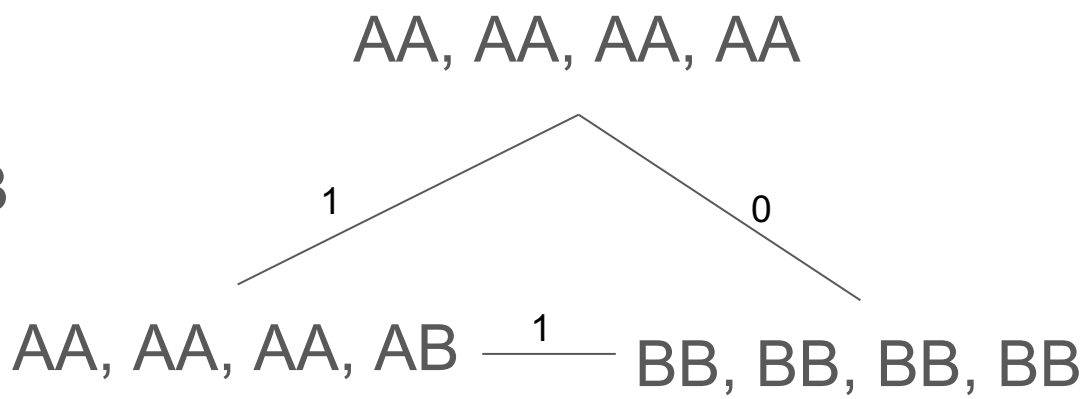
Index	Composition	Classification
1	Mixed race, renters, recent in-movers, median house value	Middle-class Renter
2	Black, high poverty, high unemployment, older homes	Struggling African American
3	Mixed race, low poverty, low unemployment, owners	Mixed Race Owner
4	Less educated, blue collar	Blue Collar
5	Educated, high vacancy, newer homes	New Starts
6	White, educated, wealthy, owners	White Elite
7	Old residents, educated, renters	Aging Community
8	Black, medium poverty/unemployment, older homes, some owners	Ascending African American
9	Owners, newer homes	Newer Suburb
10	Mixed race, high poverty, high unemployment, renters	Struggling Renter

Process of sequence

- Common ancestors
  - DNA sequences: difference due to mutation
  - OM: works by reverse the mutation
- Unfolding process (Biemann, 2011)
  - Social sciences: life/career course
  - Contextual factors
  - Recoding sequences as sequences of transitions

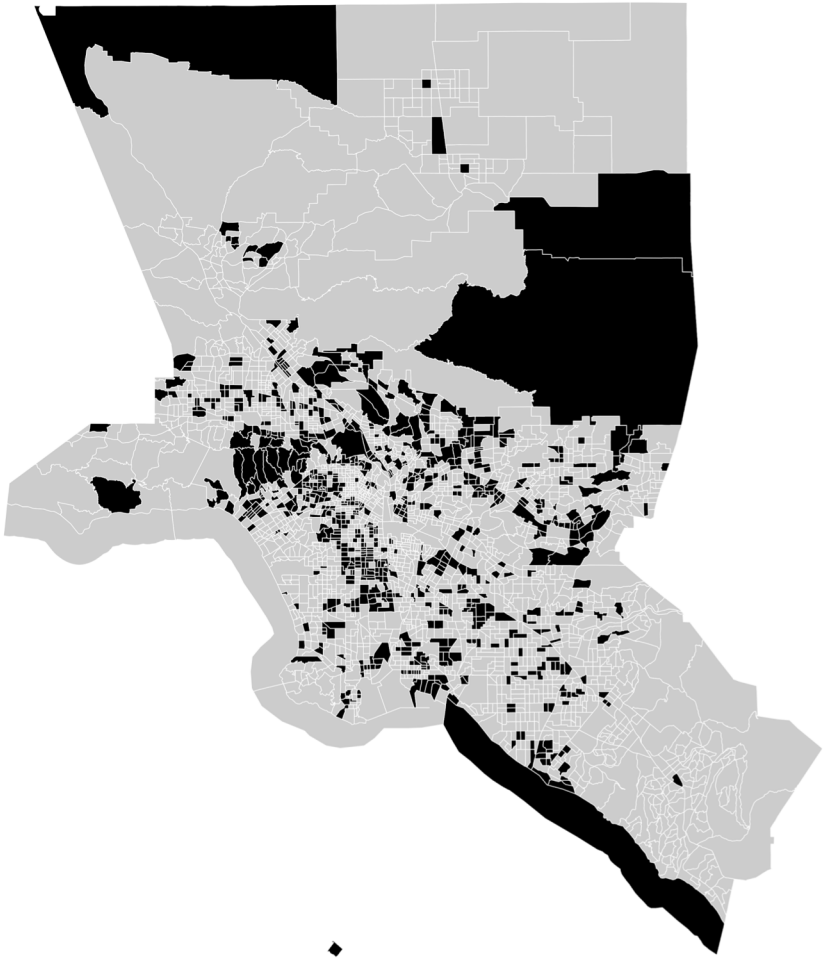
Recoding sequences





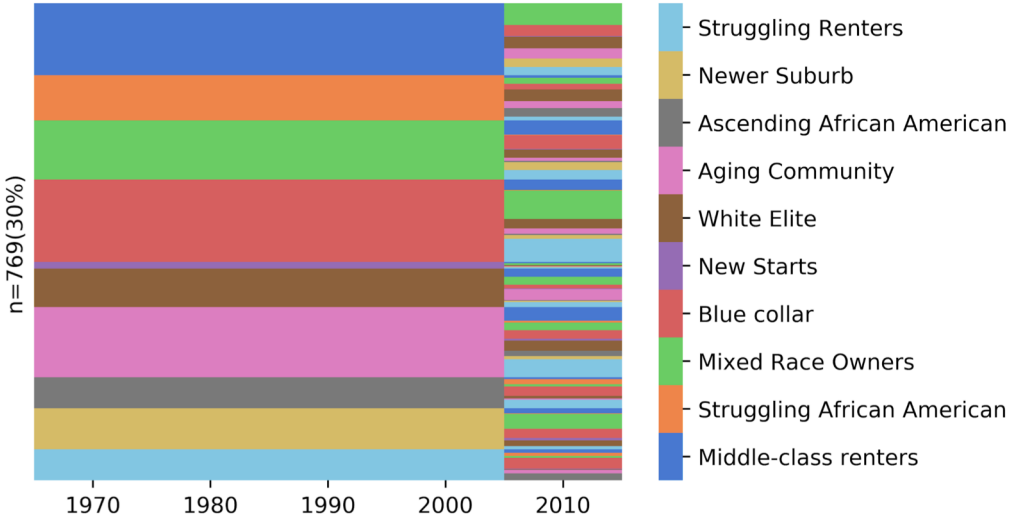
Second order property of the sequence (16)

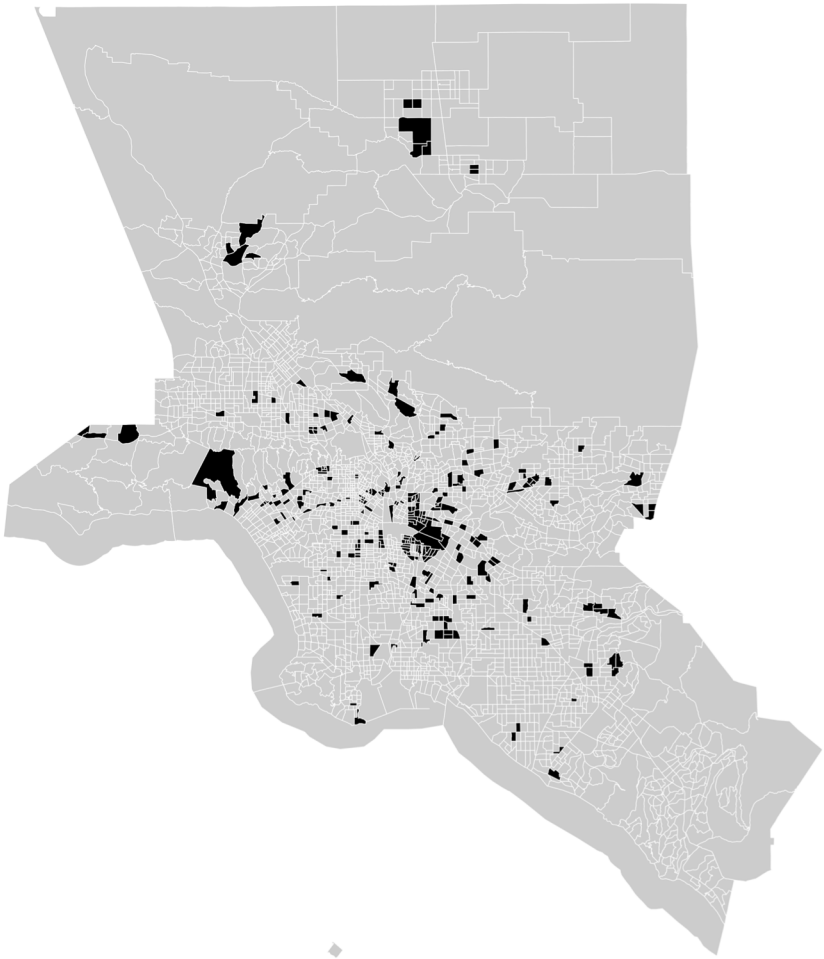
- Stability
  - Stay: AA, BB
  - Change: AB, BA
- Neighborhood sequence clusters/typology
  - How many stays/changes?
  - When/where are the stays/changes?



Stable until last period

- Most frequent neighborhood trajectory type!
- Spatially compact

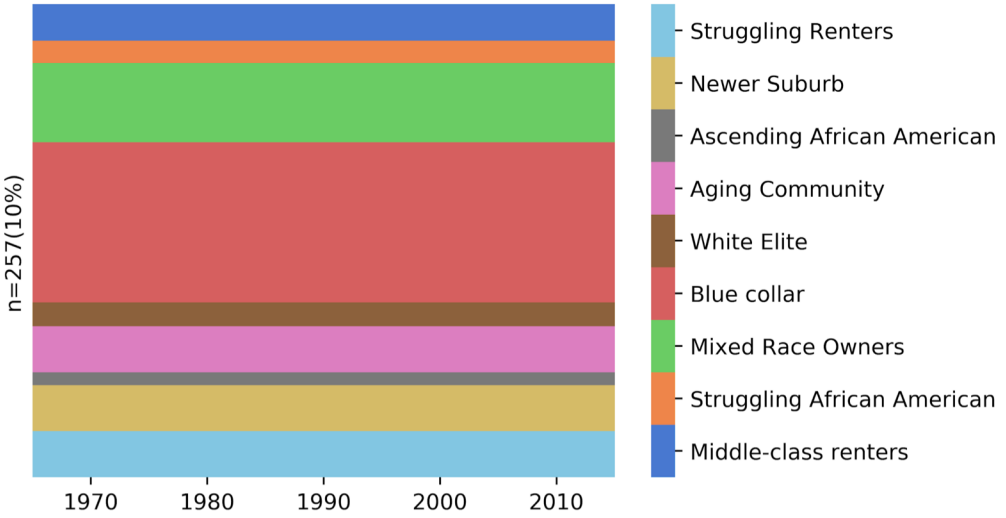


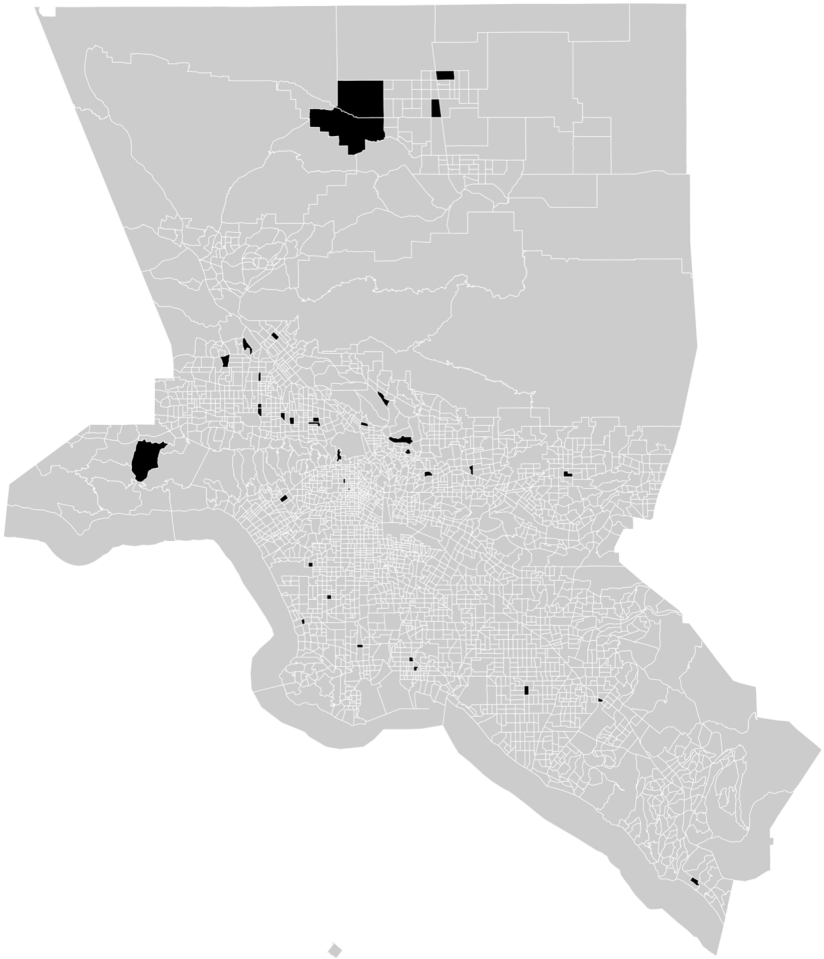


Stability

Not as frequent as the last trajectory type - not that stable after all?

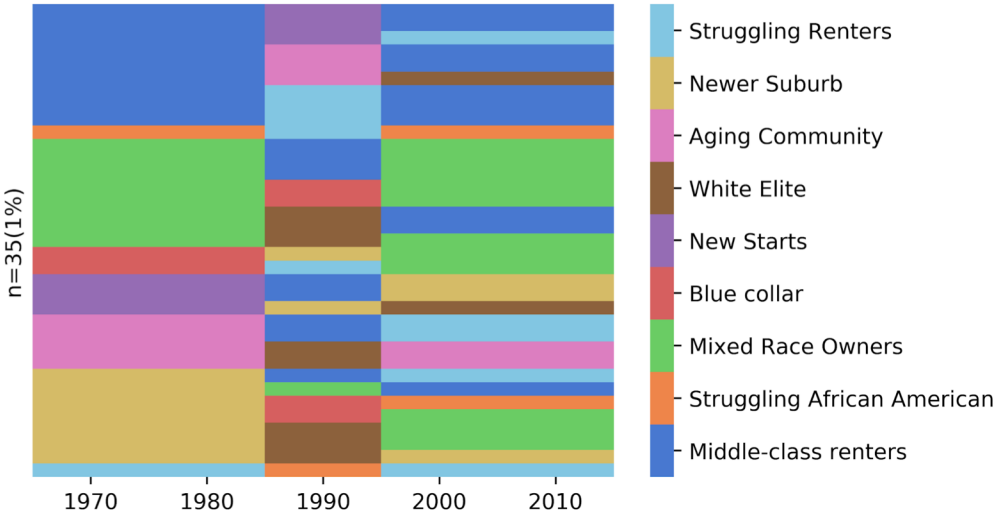
- “New Starts” missing
- Spatially compact





Two changes

Spatially dispersed



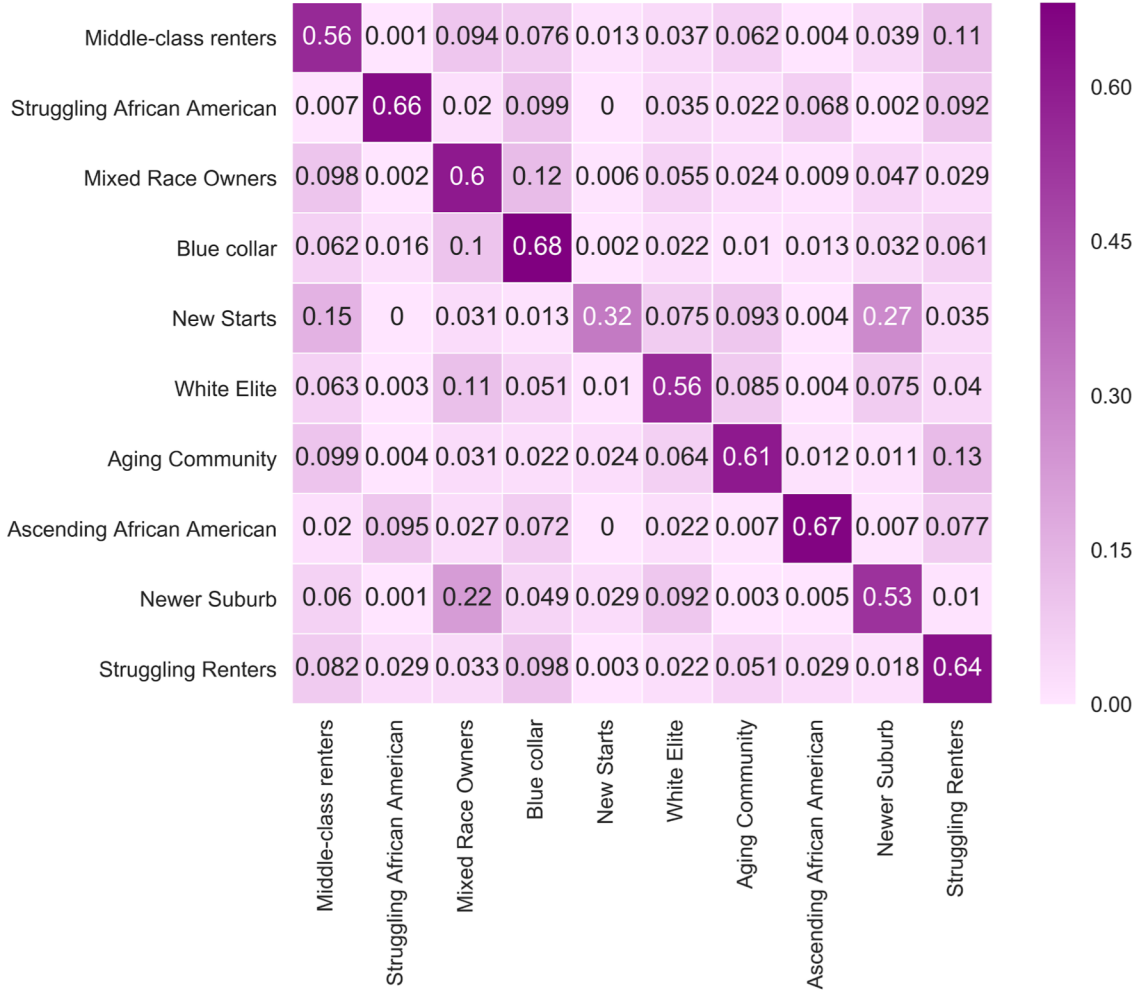
“Stability” defined with empirical transition rates (21)

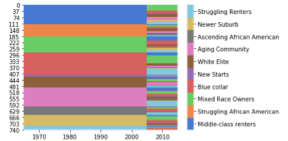
Substitution cost

$$s(a \to b, c \to d) = \left| p(a \to b) - p(c \to d) \right|$$

↑ *transition rate*

↑ *stability (easier to occur in the system)*





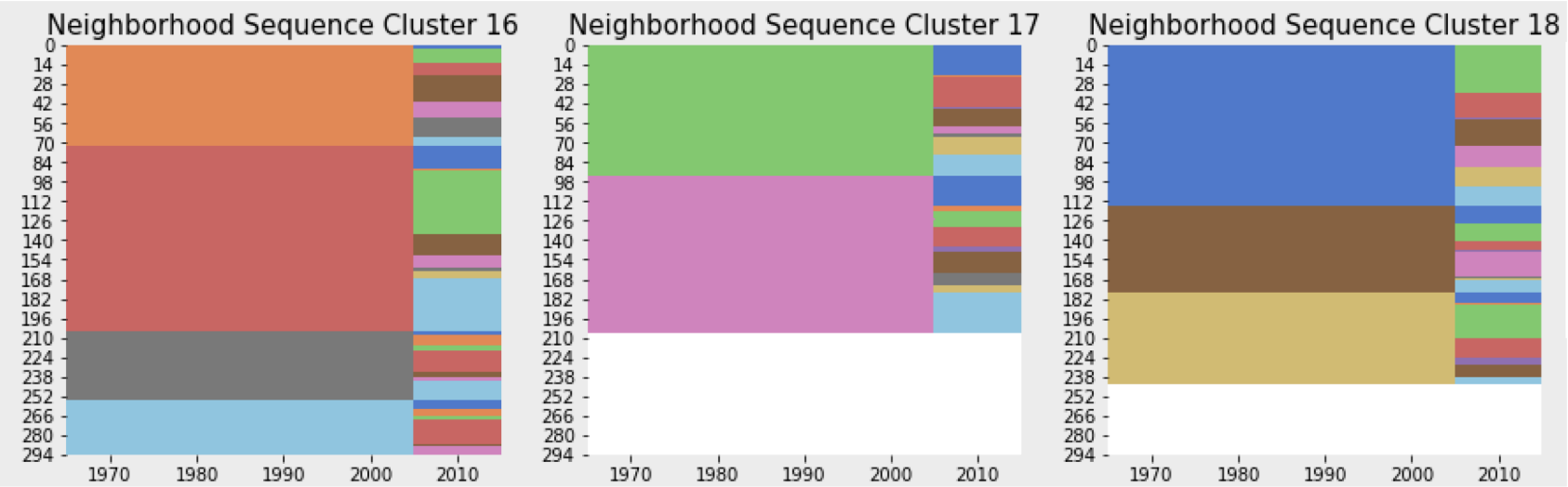
Stable until last period - more details

3 instead of 1 sequence clusters

Higher

Middle

Lower



First + Second Order (39)

- First order:
  - How different are two neighborhood types
  - Neighborhood category as a social construct:
    - Euclidean distance between two respective cluster centers
- Difference between states + Difference between transitions

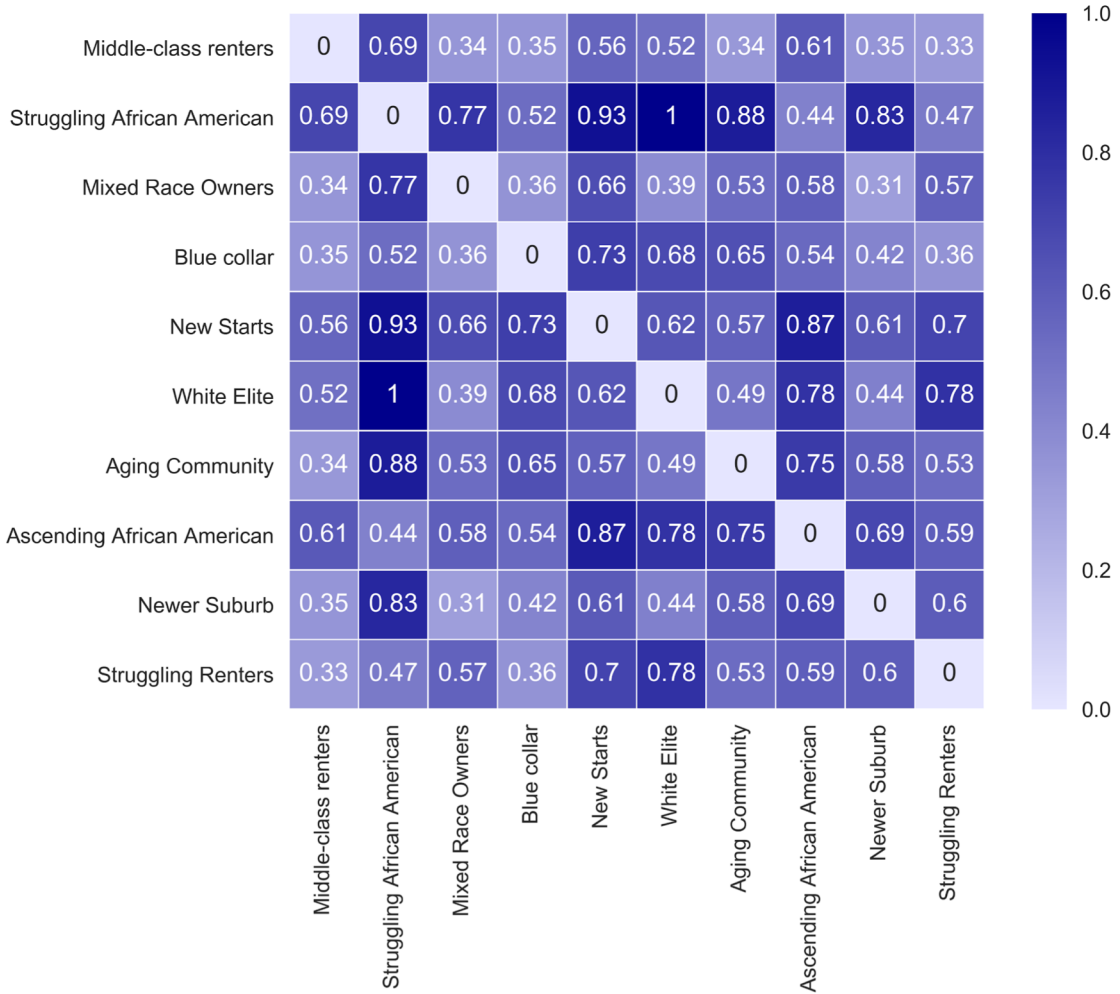
$$s(a \to b, c \to d) = ws(a, c) + (1 - w)\left|p(a \to b) - p(c \to d)\right|$$

or

$$s(a \to b, c \to d) = ws(b, d) + (1 - w)\left|p(a \to b) - p(c \to d)\right|$$

First-order difference

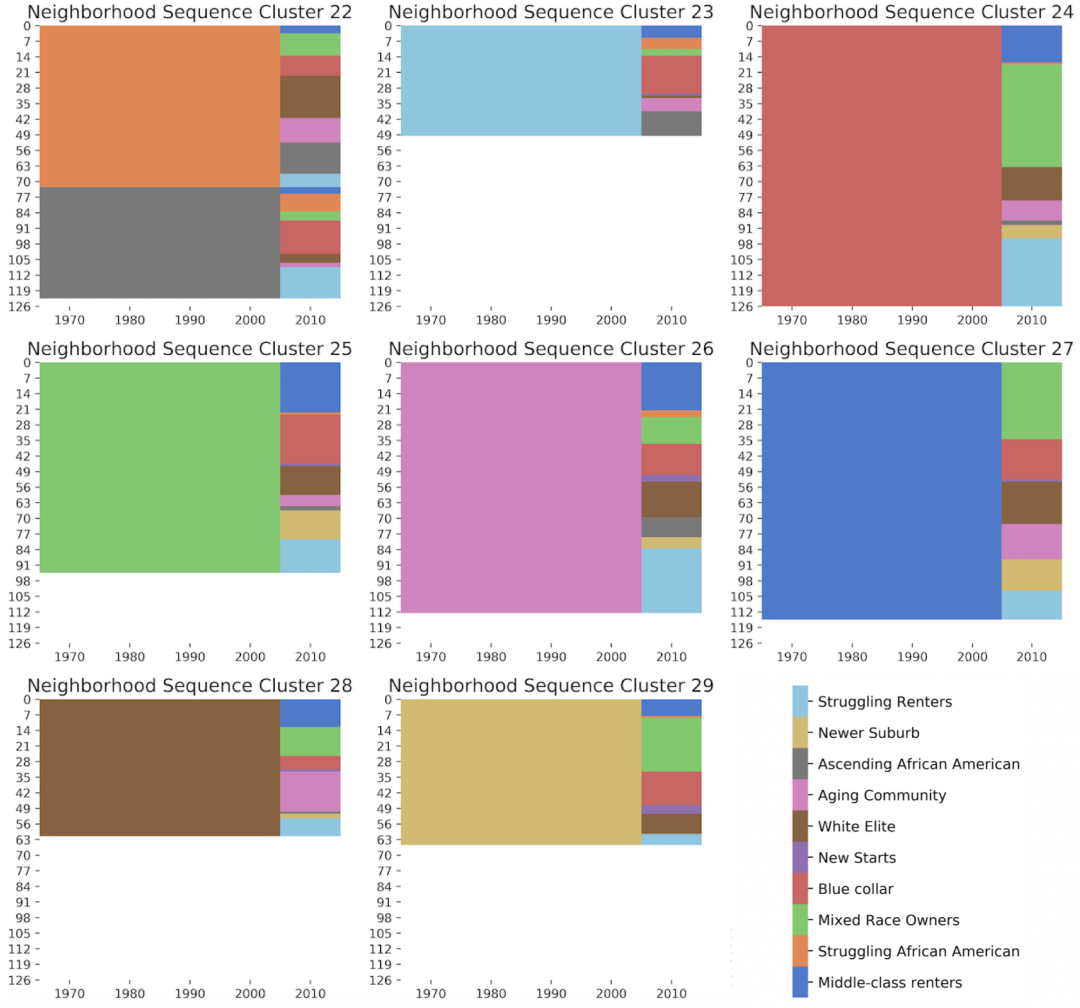
Euclidean distances between neighborhood types



Reveal More details

8 trajectory clusters

Would have a different result if the first-order difference is focused on the latter neighborhood state.



Findings

- The most frequent pathway: change between 2000-2010
  - In contrast to the past findings
  - Possible reasons:
    - Finer neighborhood typology (10 VS 5/9)
    - Choice of neighborhood variables
- Proposed approach
  - Picks up temporal changes, stability structure
  - More details revealed by incorporating first-order + second-order differences

Future directions

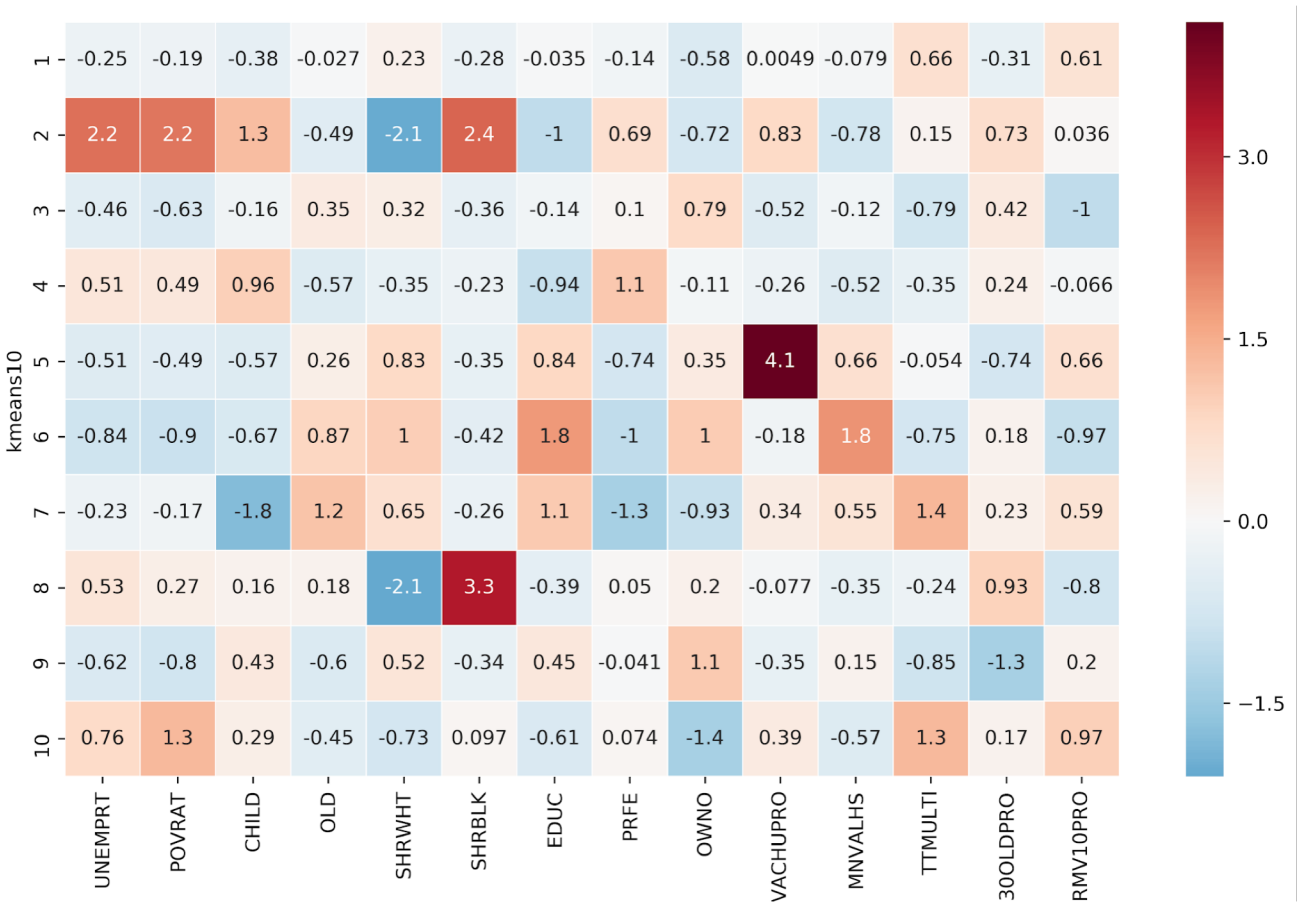
- Controlled simulation experiments
  - duration, sequencing, stability, timing
- Try on different values of weight
- Work with metro areas other than LA
- More thorough interpretation
- Spatially constrained clustering algorithm

References

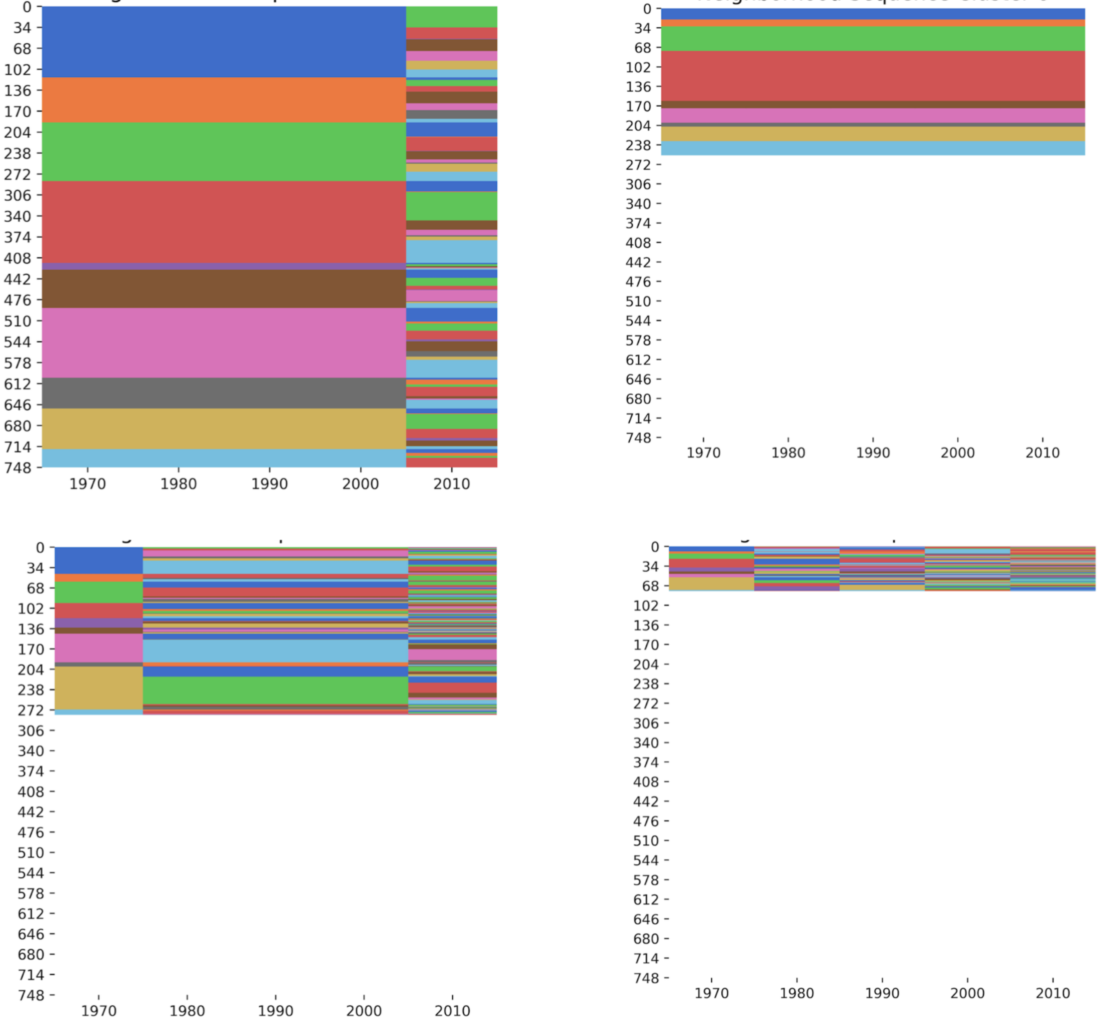
- Elizabeth C Delmelle. Differentiating pathways of neighborhood change in 50 U.S. metropolitan areas. Environment and Planning A, 49(10):2402–2424, 2017.
- Torsten Biemann. A transition-oriented approach to optimal matching. Sociological Methodology, 41(1):195–221, 2011.
- S. E. Spielman and A. Singleton. Studying neighborhoods using uncertain data from the American community survey: A contextual approach. Annals of the Association of American Geographers, 105(5):1003–1025, 2015.
- Matthias Studer and Gilbert Ritschard. What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. Journal of the Royal Statistical Society: Series A (Statistics in Society), 179(2):481–511, 2016.

Thank you!

Questions?


kmeans10
1 -0.25 -0.19 -0.38 -0.027 0.23 -0.28 -0.035 -0.14 -0.58 0.0049 -0.079 0.66 -0.31 0.61
2 2.2 2.2 1.3 -0.49 -2.1 2.4 -1 0.69 -0.72 0.83 -0.78 0.15 0.73 0.036
3 -0.46 -0.63 -0.16 0.35 0.32 -0.36 -0.14 0.1 0.79 -0.52 -0.12 -0.79 0.42 -1
4 0.51 0.49 0.96 -0.57 -0.35 -0.23 -0.94 1.1 -0.11 -0.26 -0.52 -0.35 0.24 -0.066
5 -0.51 -0.49 -0.57 0.26 0.83 -0.35 0.84 -0.74 0.35 4.1 0.66 -0.054 -0.74 0.66
6 -0.84 -0.9 -0.67 0.87 1 -0.42 1.8 -1 1 -0.18 1.8 -0.75 0.18 -0.97
7 -0.23 -0.17 -1.8 1.2 0.65 -0.26 1.1 -1.3 -0.93 0.34 0.55 1.4 0.23 0.59
8 0.53 0.27 0.16 0.18 -2.1 3.3 -0.39 0.05 0.2 -0.077 -0.35 -0.24 0.93 -0.8
9 -0.62 -0.8 0.43 -0.6 0.52 -0.34 0.45 -0.041 1.1 -0.35 0.15 -0.85 -1.3 0.2
10 0.76 1.3 0.29 -0.45 -0.73 0.097 -0.61 0.074 -1.4 0.39 -0.57 1.3 0.17 0.97
UNEMPRT
POVRAT
CHILD
OLD
SHRWHT
SHRBLK
EDUC
PRFE
OWNO
VACHUPRO
MNVALHS
TTMULTI
30OLDPRO
RMV10PRO
3.0
1.5
0.0
−1.5

Stability?



0
34
68
102
136
170
204
238
272
306
340
374
408
442
476
510
544
578
612
646
680
714
748
1970
1980
1990
2000
2010

Approaches to Studying Neighborhood Change

- A dynamic perspective
  - Focus on focal events/transitions
    - Markov models
    - Event History Analysis
  - Data model
- A holistic perspective
  - Neighborhood trajectories/sequences as meaningful units
    - Sequence Analysis (SA)
    - Optimal matching
  - Data mining or algorithmic model