Application of a New Clustering Method in Analyzing Homicide Patterns in Chicago

Fahui Wang
Department of Geography
Northern Illinois University
fwang@niu.edu

MAPS 2007
Pittsburgh

Acknowledgement

- Financial support from the U.S. Department of Housing and Urban Development (HUD) through Grant G2A62172
- Collaboration with Dr. Lan Mu at Department of Geography, University of Illinois at Urbana-Champaign

Dual objectives

- to implement a modified scale-space clustering method (MSSC) in GIS that accounts for both attribute homogeneity and spatial contiguity
- to demonstrate its values in a case study: homicide patterns in Chicago

Implications in spatial analysis

(1) to construct geographic areas with sufficiently large base population to mitigate the *small population problem*

(2) to mitigate the *modifiable areal unit problem* (MAUP)

(3) to risk less model-building error in using OLS regression as the clustered zones exhibit less *spatial autocorrelation*

Literature: Existing Approaches

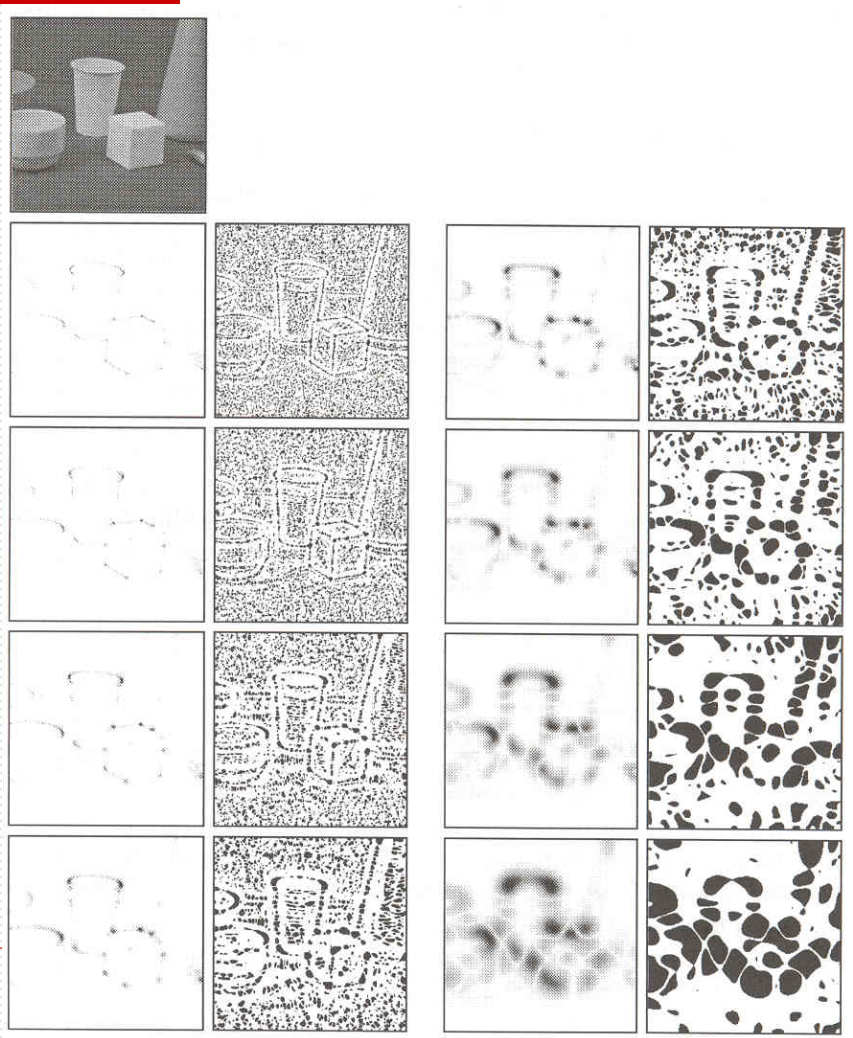
Approach	Examples
Use homicide counts instead of per capita rates	**Morenoff & Sampson (1997)**
Delete samples of small population	**Harrell & Gouvis (1994); Morenoff & Sampson (1997)**
Aggregate over more years or to a high geographic level	**Messner et al. (1999); Land et al. (1990)**
Poisson-based regressions	**Osgood (2000); Osgood & Chambers (2000)**
Construct geographic areas with large enough populations	**Haining et al. (1994); Black et al. (1996); Sampson et al (1997)**

Major challenge

- Accounting for both spatial and nonspatial factors: maintaining spatial compactness while preserving within-area attribute homogeneity

Scale-Space Theory

- In image analysis: small pixels are melt into large cells at a smaller scale



Implementation MSSC Method: Step 1

- Establishing a link between an object and its most similar adjacent object
  - minimum-distance criterion:

$$D_{ik} = \min_{j \in \mathrm{J}} \sum_{t} (x_{it} - x_{jt})^2$$

Implementation MSSC Method: Step 2

- Determining the link's direction
  - using an aggregate attribute score Q
  - i→k if $Q_i < Q_k$

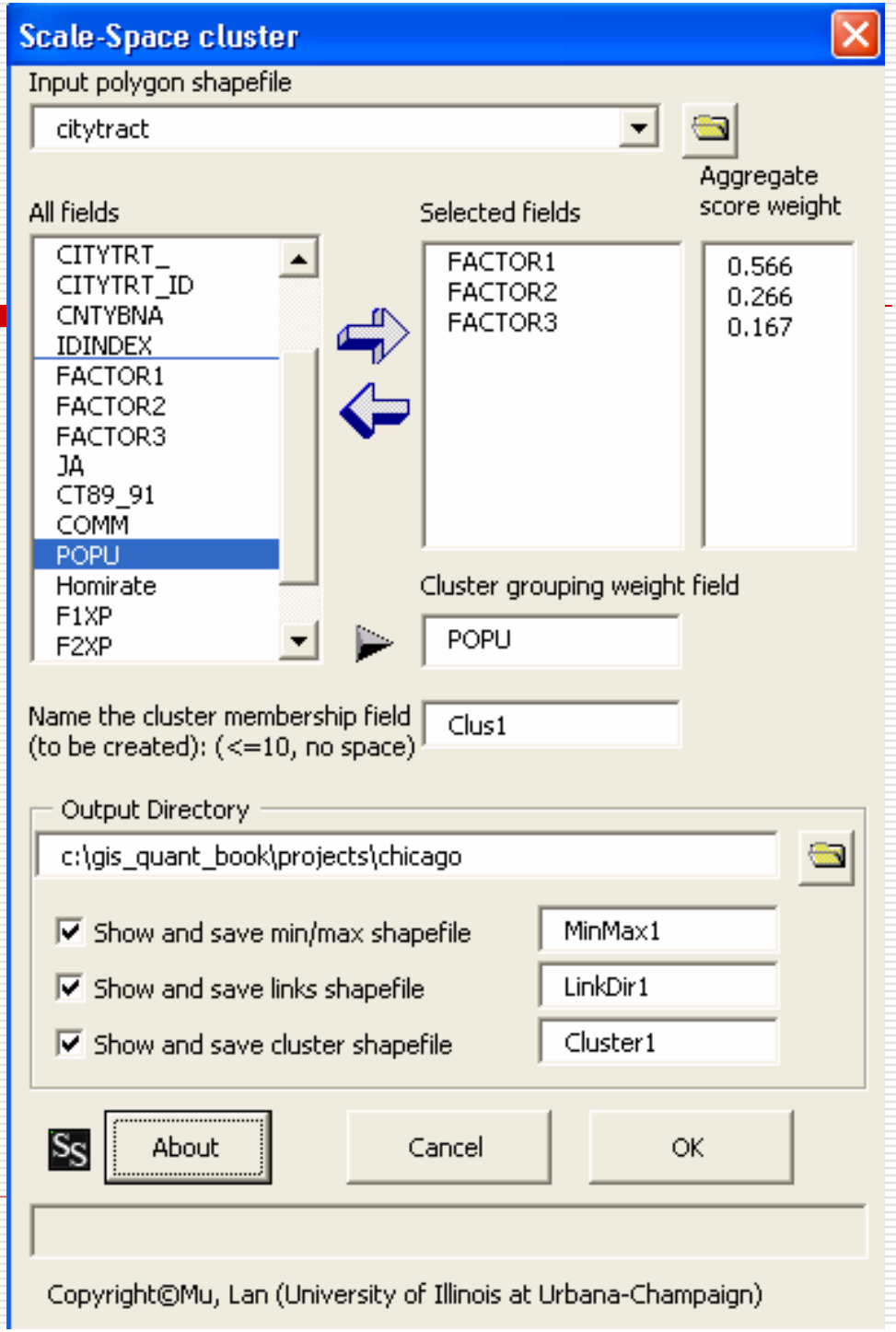
Implementation MSSC Method: Step 3

- Identifying local minima and maxima
  - local min: all directional links pointing towards other objects
  - local max: all directional links pointing towards it

Implementation MSSC Method: Step 4

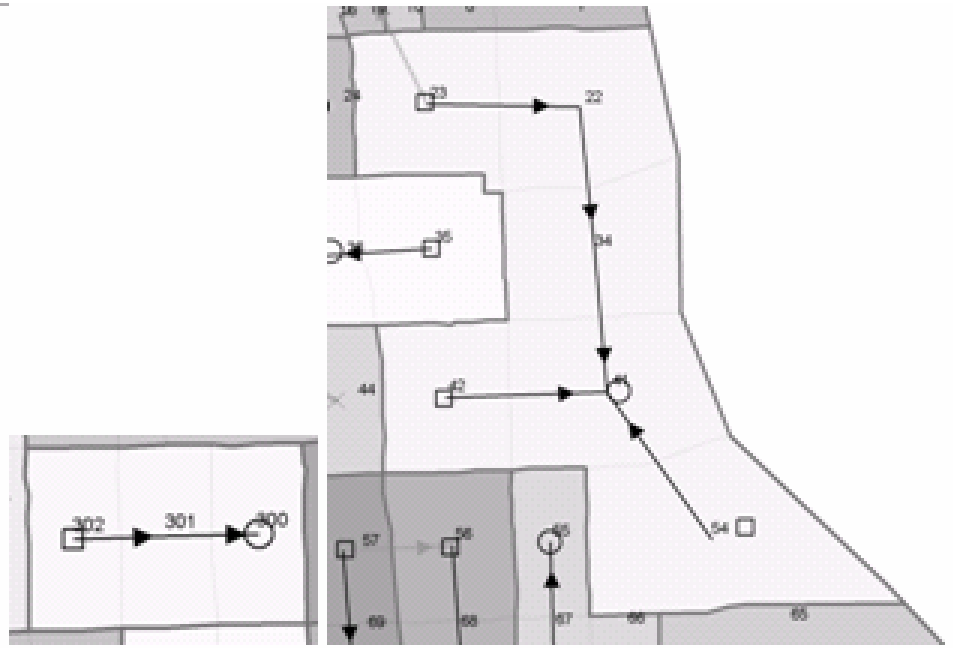
- Grouping around local maxima
  - Beginning with a local minimum, search outwards following link directions until a local maximum is reached

Implementing in VB in ArcGIS

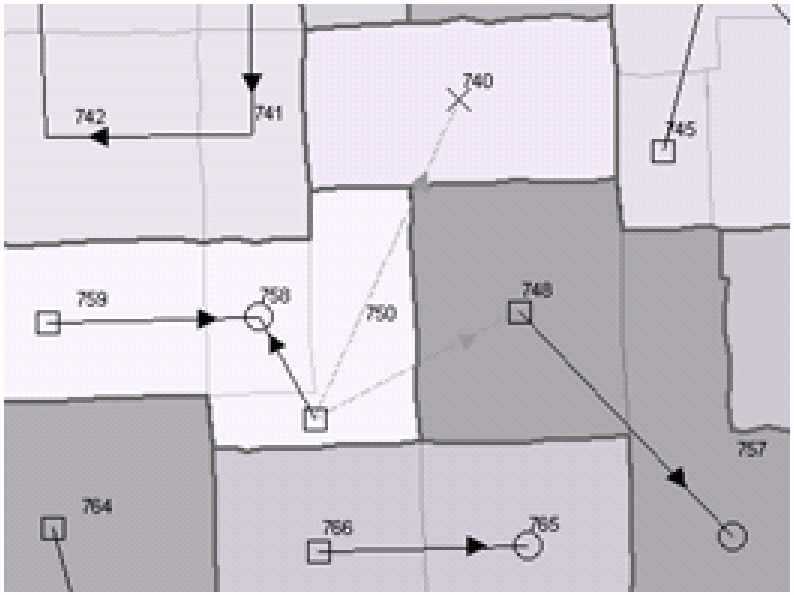


Topology in clustering

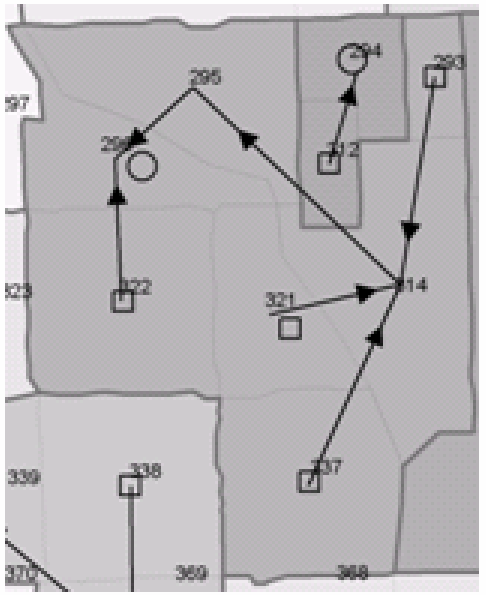
□ local min
○ local max
× orphan (no link)
→ link direction



A simple cluster B multiple minima merging to one maximum C multiple branches

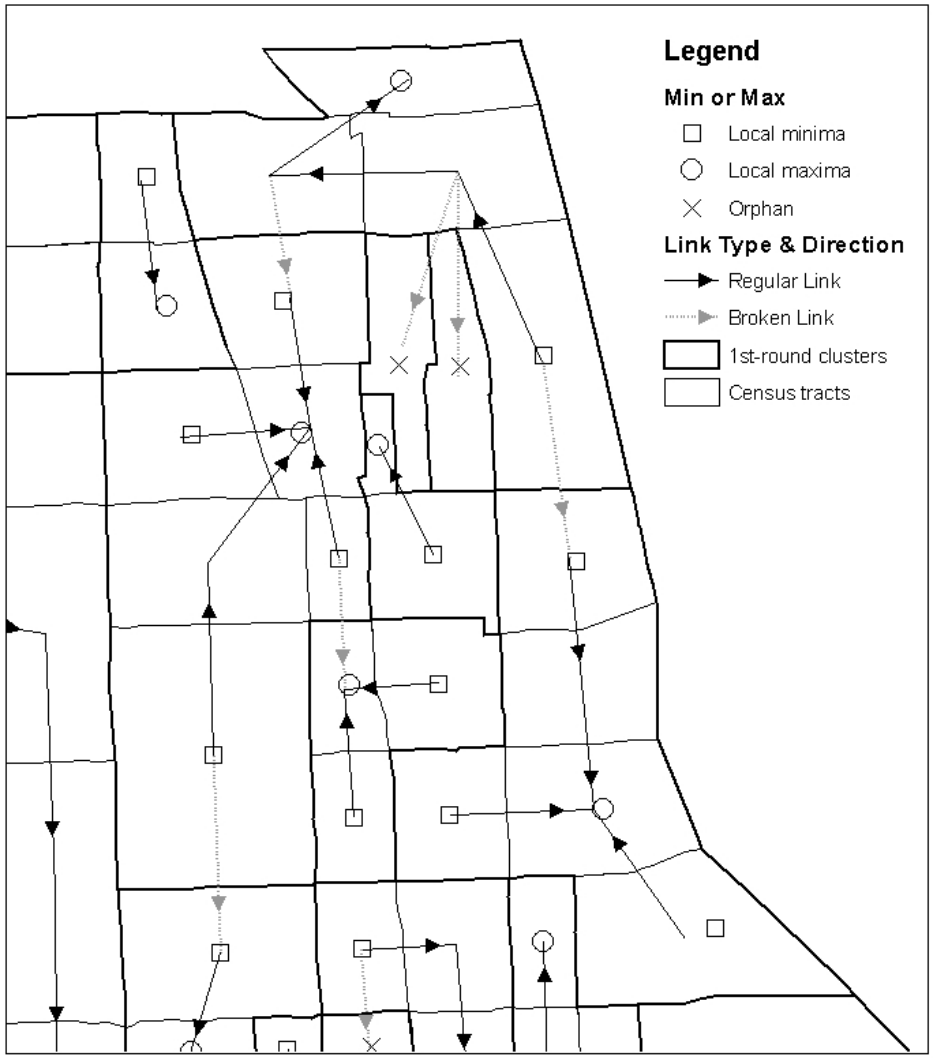


D broken link and orphan



E complicated cluster

Clustering process: a sample area



Case Study: Homicides in Chicago

- Data:
  - ICPSR by Block and Block, 1998: *Homicides in Chicago 1965-1995*.
  - 1990 Census data for defining covariates: socioeconomic factors and job access
- Basic unit: census tracts

Variables

- Homicide rate per 100,000
- Independent variables:
  - factor 1: concentrated disadvantage
  - factor 2: concentrated Latino immigration
  - factor 3: residential instability
  - Job access (JA): convenience of obtaining jobs

Factor analysis on covariates

	Factor1	Factor2	Factor3
Pubasst	0.93120	0.17595	0.01289
Femhead	0.89166	0.15172	-0.16524
Black	0.87403	-0.23226	0.15131
Poverty	0.84072	0.30861	-0.24573
Unempl	0.77234	0.18643	0.06327
Nonhighs	0.40379	0.81162	0.11539
Crowded	0.25111	0.83486	-0.12716
Latino	-0.51488	0.78821	-0.19036
Resid5	0.21224	0.02194	0.91275
Huowner	-0.45399	-0.20098	0.77222

Measure of job accessibility

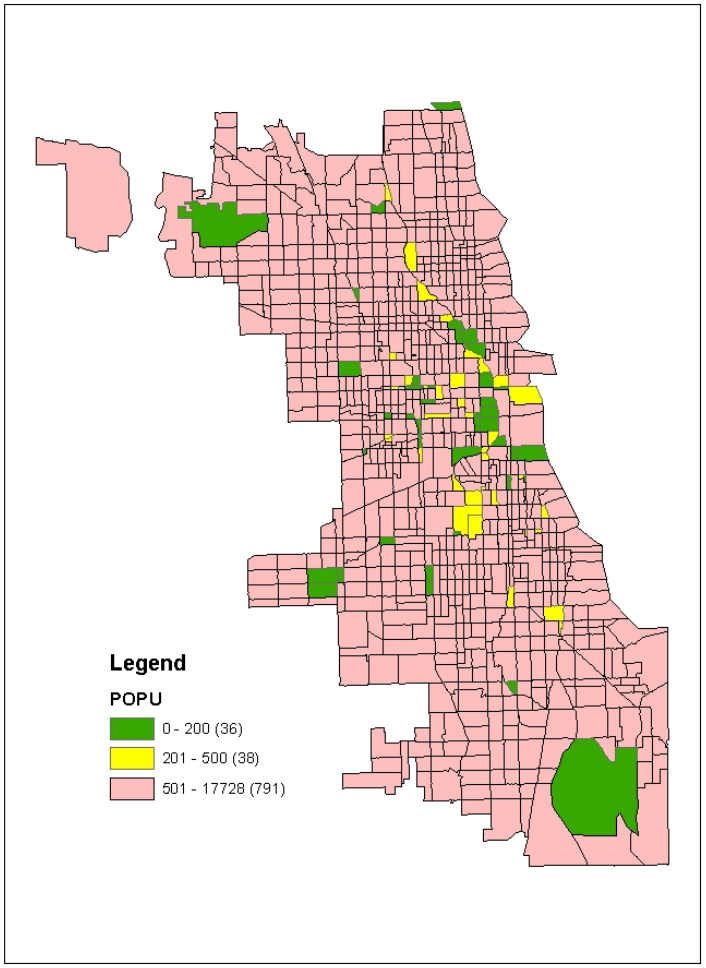
- A higher *Ai* value corresponds to better job access:

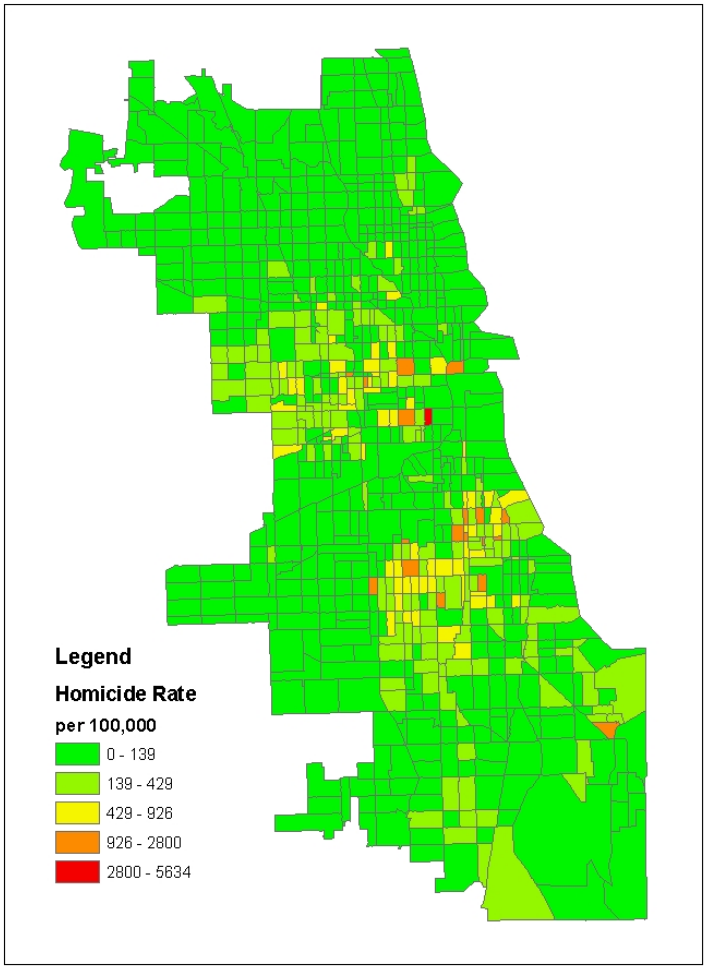
$$A_i = \sum_{j=1}^{n} \frac{J_j d_{ij}^{-\beta}}{V_j}, \quad \text{where} \quad V_j = \sum_{k=1}^{m} W_k d_{kj}^{-\beta}$$

d_{ij}: travel time between resident worker location *i* and job location *j*

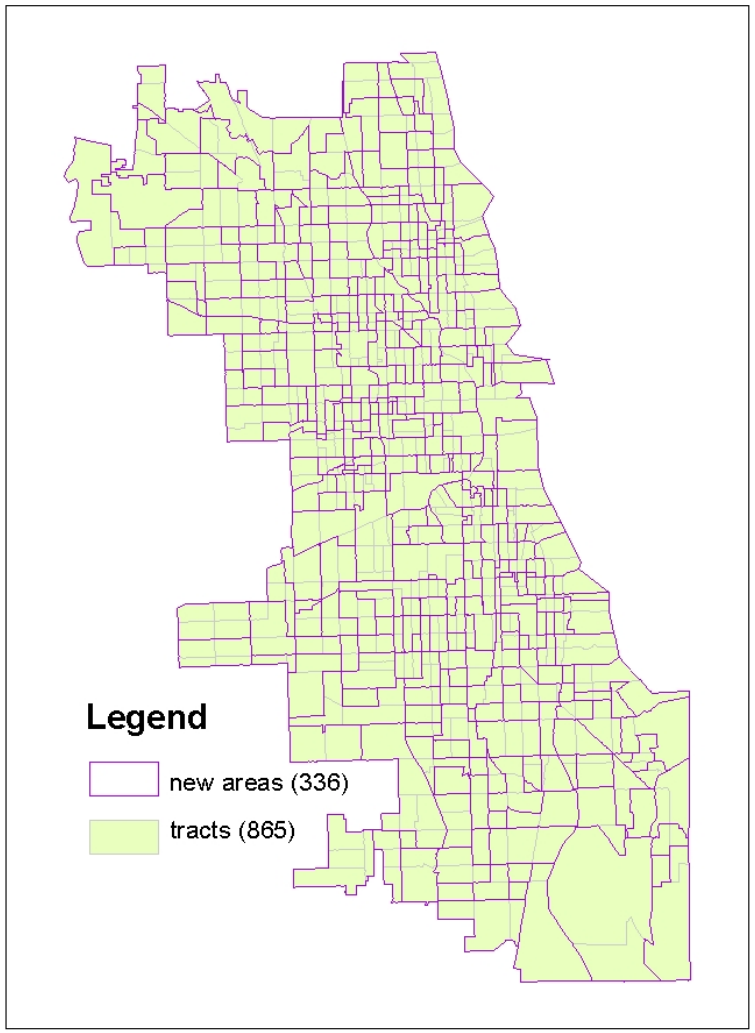
J_j: Number of jobs at j

W_i: Number of resident workers at i

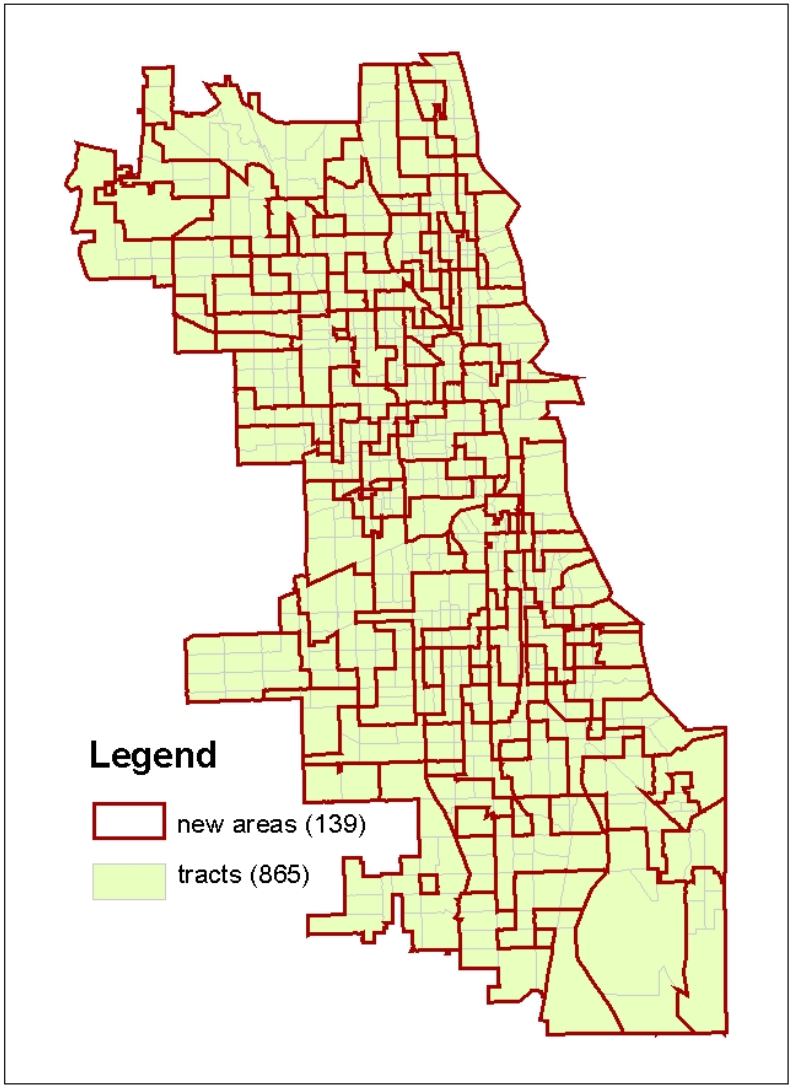

Legend
POPU
0 - 200 (36)
201 - 500 (38)
501 - 17728 (791)


Legend
Homicide Rate
per 100,000
0 - 139
139 - 429
429 - 926
926 - 2800
2800 - 5634

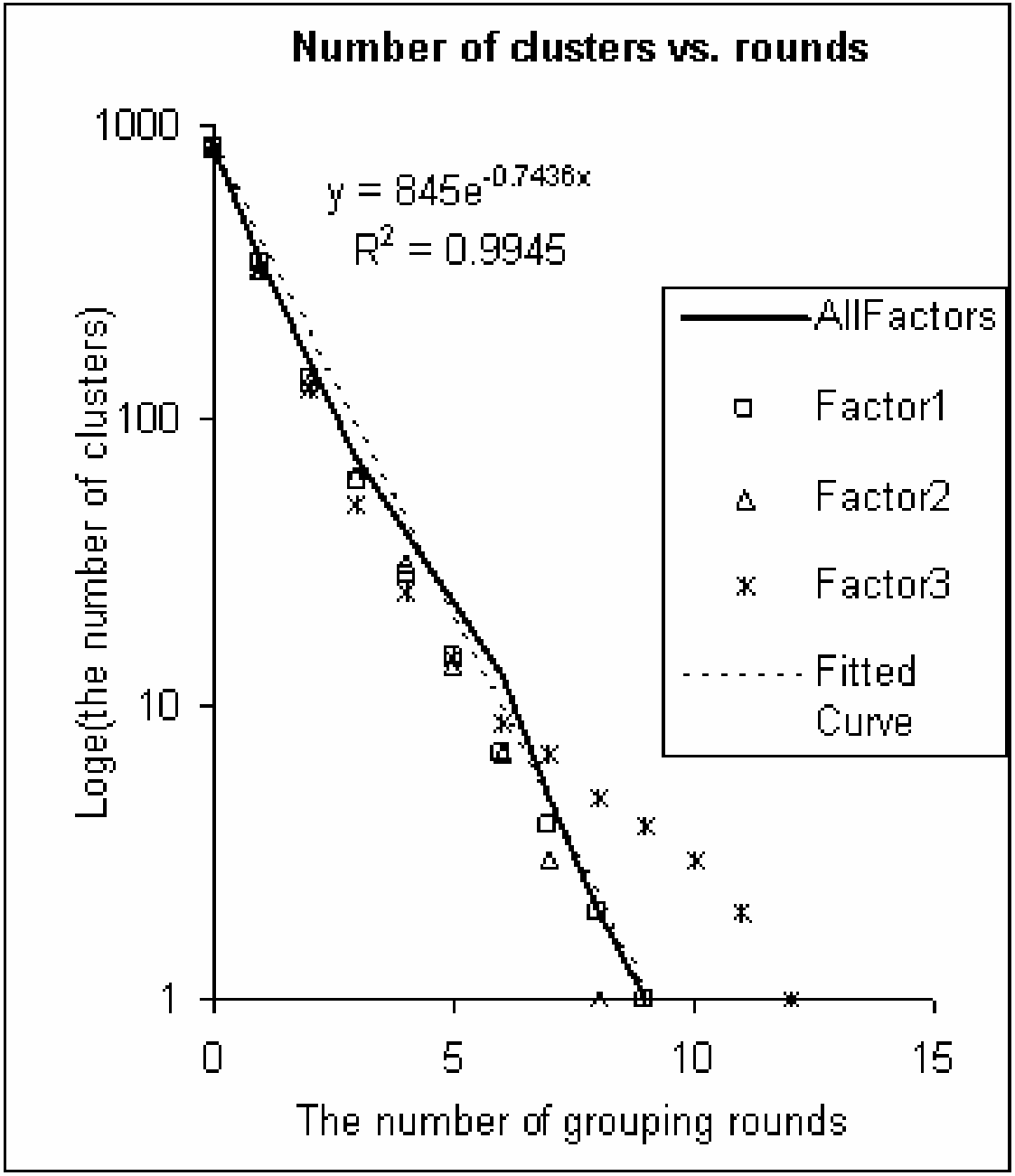
1st Round of Grouping



2nd Round of Grouping



Converging Effect of the MSSC Method



Level of convergence (LC)

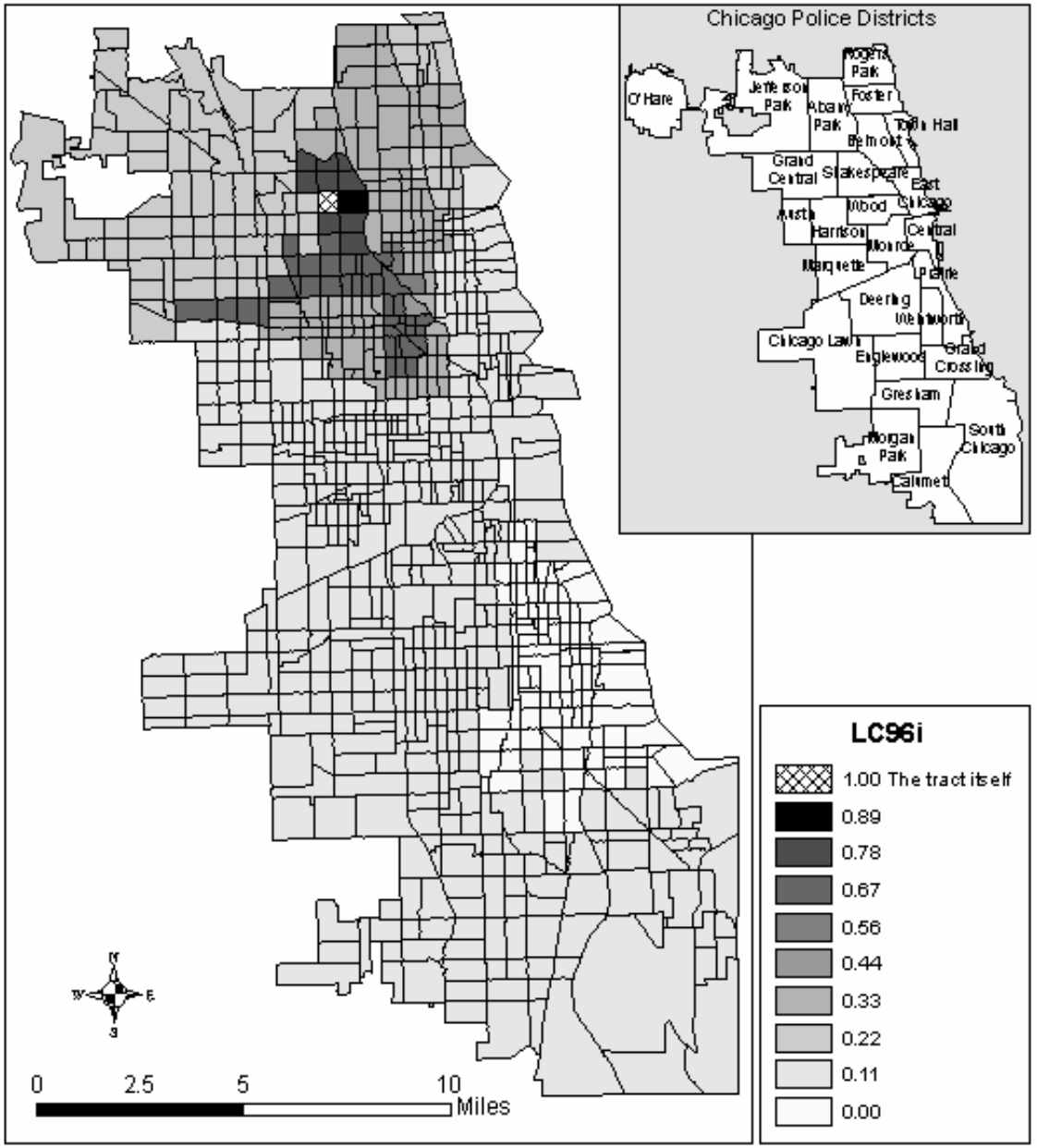
- measuring closeness of objects in terms of *attributive* and *spatial* proximity

$$LC_{ij} = 1 - \frac{r_{ij}}{R}$$

r_{ij}: clustering round when i and j are first melt into one cluster

R: total number of clustering rounds

LC values between one tract and others



Implication of LC

- Are differences between objects, *at the scale*, are minor or major?
- Is within-district socio-demographic homogeneity is achieved at the cost of spatial compactness?
- LC is a quantitative assessment tool in evaluating legitimacy in boundary changes

Spatial lag regression models

Indep't Variables	Census tracts (n=845)	1st-round clusters (n=351)	2nd-round clusters (n=154)	3rd-round clusters (n=72)
Intercept	4.5338 (7.52) ***	4.6685 (5.66) ***	5.9434 (5.53) ***	6.9563 (4.28) ***
Factor1	0.9654 (10.91) ***	0.9994 (8.98) ***	0.9578 (6.92) ***	0.8432 (4.62) ***
Factor2	0.4048 (6.01) ***	0.5259 (6.11) ***	0.6408 (6.07) ***	0.7021 (4.36) ***
Factor3	-0.0993 (-1.53)	-0.1861 (-2.17) *	-0.1291 (-1.23)	-0.1950 (-1.19)
Job Access	-2.2056 (-4.13) ***	-2.3250 (-3.18) **	-3.3508 (-3.54) ***	-4.0053 (-2.82) **
Spatial lag (ρ)	0.2750 (5.90) ***	0.2634 (4.16) ***	0.2304 (2.71) **	0.1093 (0.89)
Sq. corr.	0.424	0.533	0.667	0.640

Observations from regressions

- Consistent across analysis units
  - Poorer JA → higher homicide rates
  - Factors 1 and 2 are significant and have the expected sign (+)
  - Factor 3 is mostly not significant
- Significance level of spatial lag declines from census tracts to higher rounds of clusters
  - Less spatial autocorrelation in clusters

Implications in spatial analysis

- Homicide rates in new geographic areas have large base population, and are more reliable
- MSSC minimizes loss of information in clustering, and mitigates the *MAUP*
- By merging areas of similar attributes, need to control for *spatial autocorrelation* is less pressing

References:

F Wang. 2005. Job Access and Homicide Patterns in Chicago: An Analysis at Multiple Geographic Levels Based on Scale-Space Theory, *Journal of Quantitative Criminology* 21: 195-217.

L Mu & F Wang. A Scale-Space Clustering Method: Mitigating the Effect of Scale in the Analysis of Zone-Based Data. Forthcoming in *Annals of AAG*.

Program available in the CD in:

Wang, F. 2006. *Quantitative Methods and Applications in GIS*. Boca Raton, FL: CRC Press

Questions?