Measuring Segregation in a World of Peer Effects

Sonia Jaffe
Harvard University and the Becker Friedman Institute

February 2, 2013

Motivation

Why do we care about segregation?

- Neighborhood effects
  - Public goods
  - Opportunities

Motivation

Why do we care about segregation?

- Neighborhood effects
  - Public goods
  - Opportunities
- Peer effects
  - Segregated individuals are influenced by different information and norms.

Motivation

Why do we care about segregation?

- Neighborhood effects
  - Public goods
  - Opportunities
- Peer effects
  - Segregated individuals are influenced by different information and norms.

Want a metric that reflects the latter mechanism

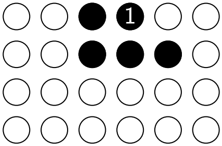
- Segregation increases with
  - the number of one's friends in the group
    - the people by whom one is influenced
  - the Segregation of one's friends
    - how 'in-group' their influence is

This Talk

1. Example
2. Background
3. Model
    - Notation
    - Preview of Metric
    - Information Propagation
4. Conclusion & Questions

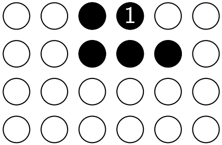
Example

Least Segregated:

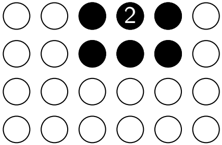


Example

Least Segregated:

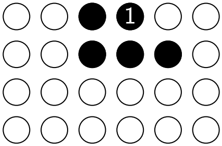


More black friends:

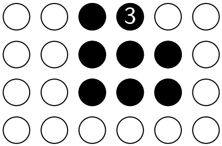


Example

Least Segregated:

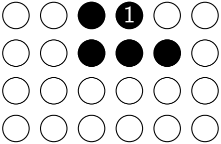


Friends are more segregated:

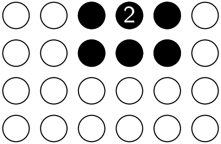


Example

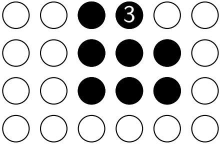
Least Segregated:



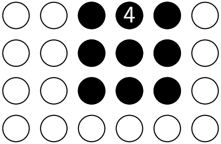
More black friends:



Friends are more segregated:



Most Segregated:



Literature

Most work on segregation uses aggregate measures:

- Empirical work on
  - Mortality (e.g.: Collins and Williams, 1999),
  - Human capital (e.g.: Borjas 1995; Guryan 2004),
  - Employment (e.g.: Kain 1968).
- Theoretical work on:
  - Properties / axioms that generate different metrics (e.g.: Duncan and Duncan 1955; Hutchens 2001),
  - Welfare implications of different metrics (e.g.: Philipson 1993).

This is most closely related to Echenique and Fryer (2007)

- Eigenvector approach gives longrun distribution of weight;
  - Effect doesn't decrease with distance.
- Only uses connections within group.

Notation

- Let n be the number of individuals;

Notation

- Let n be the number of individuals;
- Let R be a $n \times n$ matrix of relations:
  - $r_{ij} > 0$ implies that i has a relationship with j.
  - $\sum_j r_{ij} = 1$, so r_{ij} is j's share of i's relationships.
  - Often we take a matrix R' of dummies for having a relationship and divide each entry by the sum of its row.
- Let e be a vector of direct exposure.

Notation

- Let n be the number of individuals;
- Let R be a $n \times n$ matrix of relations:
  - $r_{ij} > 0$ implies that i has a relationship with j.
  - $\sum_j r_{ij} = 1$, so r_{ij} is j's share of i's relationships.
  - Often we take a matrix R' of dummies for having a relationship and divide each entry by the sum of its row.
- Let e be a vector of direct exposure.
  - Normalize to be between 0 and 1.
  - Often use $R \cdot c$ where c is some characteristic
    - Direct exposure is just the average of one's friends characteristic.
- Let δ a weight between 0 and 1.

Notation

- Let n be the number of individuals;
- Let R be a $n \times n$ matrix of relations:
  - $r_{ij} > 0$ implies that i has a relationship with j.
  - $\sum_j r_{ij} = 1$, so r_{ij} is j's share of i's relationships.
  - Often we take a matrix R' of dummies for having a relationship and divide each entry by the sum of its row.
- Let e be a vector of direct exposure.
  - Normalize to be between 0 and 1.
  - Often use $R \cdot c$ where c is some characteristic
    - Direct exposure is just the average of one's friends characteristic.
- Let δ a weight between 0 and 1.
  - Can be the 'decay factor,' the ratio of the influence of a friend-of-a-friend to the influence of a friend.

Metric

Segregation, s, along the characteristic of c, with decay factor δ is:

$$\begin{aligned} s^c =& (1-\delta)(Rc + \delta R^2 c + \delta^2 R^3 c \ldots) \\ =& (1-\delta)Rc + \delta R \cdot s \\ =& (1-\delta)(\mathcal{I} - \delta R)^{-1} Rc \end{aligned}$$

- $R \cdot c$ is the *direct exposure*.
- $R \cdot s$ is the *indirect exposure*.

Metric

Segregation, s, along the characteristic of c, with decay factor δ is:

$$\begin{aligned} s^c =& (1-\delta)(Rc + \delta R^2 c + \delta^2 R^3 c \ldots) && \text{recursive sum of effects} \\ =& (1-\delta)Rc + \delta R \cdot s && \text{weighted avg of direct and indirect} \\ =& (1-\delta)(\mathcal{I} - \delta R)^{-1} Rc && \text{explicit formula} \end{aligned}$$

- $R \cdot c$ is the *direct exposure*.
- $R \cdot s$ is the *indirect exposure*.

Model

A reduced form model of information flows:

- Information dissemination
  - Individual i receives a piece of information;
    (the existence a job opening)
  - He passes it to j with probability r_{ij};
  - With probability δ, j passes it along (using wieghts r_{jk});
    (with probability $(1-\delta)$, j applies for the job)
  - The process continues;
  - Segregation s_i^c is the prob the information ends with someone in group c;
    (someone from c applies for the job).

- Information Search
  - Individual i wants the answer to a question;
    (Is going to college worthwhile?)
  - Each agent has an answer with probability $1 - \delta$;
  - With probability r_{ij}, i asks j;
    - If j has an answer, he tells i,
      ("Yes, my brother's making a fortune on Wall Street")
    - Otherwise, he passes the question along and then passes back whatever answer he receives;
  - Segregation s_i^c is the probability that i gets an answer from someone in group c;

- Information Search
  - Individual i wants the answer to a question;
    (Is going to college worthwhile?)
  - Each agent has an answer with probability $1 - \delta$;
  - With probability r_{ij}, i asks j;
    - If j has an answer, he tells i,
      ("Yes, my brother's making a fortune on Wall Street")
    - Otherwise, he passes the question along and then passes back whatever answer he receives;
  - Segregation s_i^c is the probability that i gets an answer from someone in group c;

For i passively receiving information, use "incoming relationship" matrix, R^T.

- Signal Aggregation
  - Each individual gets a signal e_i
    (Estimate of the return to education)
  - Each agent's opinion gives weight $(1-\delta)$ to their own signal and weight δ to the weighted average of their friends opinions
  - Everyone shares opinions and updates opinions until each opinion converges.
  - The final opinions are

$$\tilde{s}^e = (1-\delta)(\mathcal{I}-\delta R)^{-1}e.$$

If signals vary by groups then

$$\tilde{s}^c = (1-\delta)(\mathcal{I}-\delta R)^{-1}c$$

gives the extent that each agent is affected by the signals of group c.

Can also think about overexposure

- Let $\bar{a}$ be a vector the same length as a with all entries equal to the mean of a.
- The extent that individuals are more exposed than expected to group c is

$$\begin{aligned}\hat{s}^c =&(1-\delta)(\mathcal{I}-\delta R)^{-1}(R\cdot c-\bar{c})\\ =&(1-\delta)(\mathcal{I}-\delta R)^{-1}R\cdot c-\bar{c}.\end{aligned}$$

Extensions

What are other interesting questions

- Correlation across types of segregation
  - Try to reject "racial segregation can be explained by income segregation"

$$E[S_a|R,b] = (1-\delta)(\mathcal{I}-\delta R)^{-1}(RE[a|b]).$$

- Test for homophily

$$S_r - E[S_r|R,I] = \beta_0 + \beta_1 S_I.$$

- Using test scores/income

Conclusion and Questions

- A metric of segregation motivated by peer effects.
- Based on a model of information flows through networks.

Using the metric

- Bringing to data: Add Health /census
- For what effects is this the right metric of segregation?
- What interventions would/should target this type of segregation?
  - If people don't think about information flows when choosing friends, are they over-segregating?
  - What about the informational externalities on their friends?