ScodeVerbatimfontshape= StatainputVerbatim fontsize=, fontfamily=courier,frame=single,framesep=2mm, label=Stata code,labelposition=topline RinputVerbatim fontsize=, fontfamily=courier,frame=single,framesep=2mm, label=R code,labelposition=topline PjkinputVerbatim fontsize=, fontfamily=courier,frame=single,framesep=2mm, label=Pajek code,labelposition=topline

Segregation with Social Linkages: Evaluating Schelling's Model with Networked Individuals

Roy Cerqueti[b], Luca De Benedictis[a,*]

[a] *Rossi-Doria Center, University Roma Tre and DED, University of Macerata, Macerata (MC), Italy*
[b] *DED, University of Macerata, Macerata (MC), Italy*

Abstract

To be done

JEL Classification

Keywords: Schelling's Segregation Model, Networks.

Acknowledgements We are grateful to ·.

February 2016 - Preliminary version

*Corresponding author
Email addresses: roy.cerqueti@unimc.it (Roy Cerqueti), luca.debenedictis@unimc.it (Luca De Benedictis)

1. Introduction

Notes:

Georg Simmel (1984) for instance has contributed to our understanding of cross-cutting social circles and the spatial nature of these. In this essay on the modern metropolis he argues that spatial proximity in cities does not automatically result in social closeness. On the contrary, in our modern mobile society it seems that social networks of strong ties have increasingly 'stretched' across greater geographical distances, so that the lack of proximity does not necessarily mean ties cease to exist.

2. The model

We consider a graph $G = (V, E)$, with $|V| = N$ and $|E| = M$. We also introduce two different populations: the red people and the blue ones. The set collecting the reds and the blues are denoted by $\mathcal{R}$ and $\mathcal{B}$, with $|\mathcal{R}| = R$ and $|\mathcal{B}| = B$. We assume that $R + B < N$. The generic elements of $\mathcal{R}$ and $\mathcal{B}$ will be denoted by r and b, respectively.

Each vertex of the graph corresponds to a site, which can be occupied by an element of $\mathcal{R}$, of $\mathcal{B}$ or it can be empty.

Reds and blues differ not only for the color. More specifically, beyond the color, each agent exhibits a multidimensional attribute $\mathbf{v} = (v_1, \dots, v_k, \dots, v_K) \in \mathbb{R}^K$, where v_k is the quantitative translation of a qualitative trait, for each $k = 1, \dots, K$.

Furthermore, the relationship between each couple of agents i, j is captured by H parameters, collected in a vector $\mathbf{a}_{ij} = (a_{ij}^{(1)}, \dots, a_{ij}^{(h)}, \dots, a_{ij}^{(H)}) \in \mathbb{R}^H$. As for the v's, also the parameters a's represent the quantitative view of qualitative characteristics of the links.

The attributes of the agents and of the links are in relation also to the geographical collocation of the agent in the graph G. Thus, a variation of position of an agent could lead to a variation of her/his attributes. For the sake of simplicity, we omit a special reference to the dependence on the site of the attributes, and include such a dependence into the definition of the v's and the a's. We will come back to this relevant aspect below, when further details will be given.

By conventional agreement, we assume that the absence of the k-th attribute for an agent (of the h-th parameter of the link between agents i and j) means that $v_k = 0$ ($a_{ij}^{(h)} = 0$). Under this assumption, we can redefine the set of the edges of the graph as

$$E = \{(i, j) \in V^2 \,|\, \mathbf{a}_{ij} \neq \mathbf{0}\}.$$

For what concerns the links, we can then construct a squared weighted adjacent matrix $\mathbf{A}$ of order $R+B$ such that

$$\mathbf{A} = (\mathbf{a}_{ij})_{i,j=1,\dots,R+B} \,. \tag{1}$$

The principal diagonal of $\mathbf{A}$ is assumed to be filled with null H-dimensional vectors, in order to exclude self-connections. Thus, $\mathbf{A}$ is a table containing $(R+B)(R+B-1)$ nonnull vectors or less. We exclude the presence of direct links of anything with the empty sites, so that $M \leq (R+B)(R+B-1)$.

Each nonempty site of the graph is then endowed with a color and a K-dimensional weight, and each edge exhibits a nonnull H-dimensional weight.

The weighted graph is then a network on G.

Agents move in the sites of the graph. Without losing of generality, we assume that movements are on a sequential basis, which means that the location process of the agents is a process with discrete time. The way in which agents move is natural: the agent changes her/his location site with an empty site. In so doing, the network changes and evolves in time.

For a fixed time $t \in \mathbb{N}$, we denote the network at time t by $\mathcal{N}_t$. The agents start to move at time 1. Thus, $\mathcal{N}_t$ represents the situation of the sites-links-weights on G at the t-th move, i.e.: after the moves of t agents.

We denote by *initial configuration* the network at time $t = 0$, i.e. $\mathcal{N}_0$.

Each agent selects the empty site where collocating according to a specific criterion. Moreover, the order in which agents move is also driven by a rule. We enter the details of such criteria.

We introduce two normalized utility functions: $u_V : \{\mathcal{R}, \mathcal{B}\} \times \mathbb{R}^K \to [0,1]$ (UTILITA' DELLA POSIZIONE. BISOGNA INSERIRE ANCHE LA CONFIGURAZIONE DEI VICINI, IN MODO DA INSERIRE NELLE CARATTERISTICHE DEL SINGOLO SITO ANCHE COSA C'E' VICINO AD ESSO. QUESTO ASPETTO RIENTRA NELLA DIPENDENZA DEGLI ATTRIBUTI DALLO SPECIFICO SITO) such that $\{\mathcal{R}, \mathcal{B}\} \times \mathbb{R}^K \ni (color, \mathbf{v}) \mapsto u_V(color, \mathbf{v})$. Function u_V represents the level of satisfaction of an agent of a specific color with the multidimensional attribute $\mathbf{v}$ for her/his position in the graph; $\mathbf{u}_\emptyset : \{\mathcal{R}, \mathcal{B}\} \times \mathbb{R}^{K+H} \to [0,1]^{N-R-B}$ (UTILITA' MULTIDIMENSIONALE DELLE CELLE VUOTE) such that $\{\mathcal{R}, \mathcal{B}\} \times \mathbb{R}^{K+H} \ni (color, \mathbf{v}, \mathbf{a}) \mapsto \mathbf{u}_\emptyset(color, \mathbf{v}, \mathbf{a})$. Function $\mathbf{u}_\emptyset$ assigns a level of satisfaction of an agent for all the $N-R-B$ empty vertices in V, so that we write $\mathbf{u}_\emptyset = (u_\emptyset^{(1)}, \dots, u_\emptyset^{(e)}, \dots, u_\emptyset^{(N-R-B)})$.

Utilities have zero (unitary) value at the lowest (highest) level of satisfaction.

The range of the utility functions plays a key role here. In fact, the value of the utility u_V can be compared with the ones of the utilities $u_\emptyset^{(1)}, \dots, u_\emptyset^{(e)}, \dots, u_\emptyset^{(N-R-B)}$, and such a comparison will drive the allocation procedure of the agents in the sites of the graph.

Of course, as the time changes, the value of the utility functions of each agent changes as well. In fact, the changing of configuration leads to new attributes $\mathbf{v}$'s and adjacent matrices $\mathbf{A}$'s. Substantially, agents change their opinions about the appeal of the free vertices of the graph and of their own positions, since – for example – new neighborhoods appear and the old ones disappear.

To avoid a cumbersome notation, we state hereafter that $u_V(j,t)$ and $u_{\emptyset}^{(e)}(j,t)$ are the utilities u_V and $u_{\emptyset}^{(e)}$ for the agent j at time t, respectively, for each $j = 1,\dots,R+B$, $t \in \mathbb{N}$ and $e = 1,\dots,N-R-B$.

The mechanism of movement can be described as follows. At each step, only one agent moves[1]. An agent moves only if the change of her/his position leads to an improvement of her/his utility. The first moving agent is the one with the lowest level of utility u_V, and she/he moves to one of the empty cells e's giving the maximum levels of utility $u_{\emptyset}^{(e)}$'s.

If more than one empty cell maximizes the utility, then the occupied one is randomly selected among the utility maximizing cells. Analogously, if more than one agent is in the position of moving, then the moving one is selected according to a uniform distribution over the available alternatives.

We formalize this mechanism. At time t, all the agents have a complete information on the sites of the network $\mathcal{N}_{t-1}$ and, specifically, on the distribution of reds, blues and empty cells.

The agent moving at time t will be labeled as $j_t^\star$, whose definition is obtained according to the following two conditions:

$$\begin{cases} j_t^\star \in \operatorname{argmin}\{u_V(j,t)\,|\,j=1,\dots,R+B\}\,; \\ \exists\, e = 1,\dots,N-R-B\,|\,u_V(j_t^\star,t) \le u_{\emptyset}^{(e)}(j_t^\star,t) \end{cases} \tag{2}$$

The two conditions in (2) must be jointly verified: the former one indicates that the agent moving are that whose utility of her/his position is at the lowest level; the latter one specifies that the agent changes her/his position only if he/she will improve her/his utility.

The mechanism admits theoretically simultaneous movements of different agents at time t, in that system (2) can be satisfied by more than one agent. By assuming that Q agents satisfy (2), with $Q = 1,\dots,R+B$, we will label the Q moving agents by $j_{1,t}^\star,\dots,j_{q,t}^\star,\dots,j_{Q,t}^\star$.

A random extraction of one element of the set $\{1,\dots,Q\}$ identifies univocally $\bar{q} \in \{1,\dots,Q\}$ such that the moving agent is $j_t^\star = j_{\bar{q},t}^\star$.

The selection of the empty cell where moving is driven also by utility. In particular, the moving agent will occupy an empty cell labeled by $e_t^\star$ and defined as follows:

[1] The extension to the case of simultaneous movements of agents can be, of course, presented. However, it does not add much to the single-moves case.

$$e_t^\star \in \Phi_t := \text{argmax}\left\{ u_\emptyset^{(e)}(j_t^\star, t) \,|\, e = 1, \dots, N - R - B \right\}. \tag{3}$$

If $|\Phi_t| = C > 1$, then there are C different cells which are indifferent – in terms of the utility function $u_\emptyset$ – for the moving agent.

We will label the C utility maximizer cells by $e^\star_{1,t}, \dots, e^\star_{c,t}, \dots, e^\star_{C,t}$. Also in this case, we extract randomly one element of the set $\{1, \dots, C\}$, and identify accordingly $\bar{c} \in \{1, \dots, C\}$ such that the selected empty cell is $e^\star_t = e^\star_{\bar{c},t}$.

The movement of $j^\star_t$ induces a new configuration of the network, which passes from $\mathcal{N}_{t-1}$ to $\mathcal{N}_t$. The movement process then continues according to this mechanism. Anew moving agent $j^\star_{t+1}$ is identified, which changes her/his original position with the empty cell $e^\star_{t+1}$.

The process stops at time T when all the agents cannot improve their utility when moving from their position to an empty cell. Formally, this means that it does not exists $j^\star_T = 1, \dots, R + B$ such that system (2) is verified. This said, the process can also never stop.

Remark 1. The proposed framework extends the Schelling segregation model. QUI PARLIAMO UN MOMENTO, PER CAPIRE COME E' FATTO SCHELLING. HO UN'IDEA, MA MI CONFRONTEREI CON TE.

Remark 2. The properties of the model depends on how the networks and the utility functions are defined. A relevant role is also played by the topology of the graph and by the relative number of blues, reds and empty cells.

One can present specific setting where the process does not stop and goes on at the infinity.

Moreover, if the process stops at a certain time T, then the final configuration might be viewed as a segregation one – where, of course, the concept of segregation needs a proper definition.

3. An example

In the framework described in the previous section, we imagine that the nodes of the graph are placed in a square lattice with side n, so that $|V| = N = n \times n$. Thus, each node is identified by its row and column, so that $v_{rc} \in V$ is the node of the grid at the intersection of row r and column c, for each $r, c = 1, \dots, n$. For the sake of simplicity and when needed, we will say that $v_{rc} = i$, where $i = (r - 1)n + c$. Such a definition of i labels with a number the elements of V, and it results $i = 1, \dots, N$.

Hereafter, without loss of generality, we will use equivalently $i \in V$ and $i \in \{1, \dots, N\}$.

All the nodes, also those sharing the same color, differ for the social connections they have. Social connections are nothing but links among the nodes, and form in this context the set E of the edges of the graph.

Specifically, according to the definition of E in (**??** XXXX DA METTERE, ULTIMA FORMULA A PAGINA 3 XXXX), we say that $(i,j) \in E$ if and only if there is a social connection between agent i and agent j, for each $i,j \in \mathcal{B} \cup \mathcal{R}$. This definition includes in a natural way the evidence that an empty cell cannot have any interaction with an agent, and so we do not need to specify that the nodes $i,j \in V$ are occupied by agents. We define

$$E_i = \{j \in V | (i,j) \in E\}, \qquad \forall\, i \in V.$$

The set E_i is the *friendship set* of the node i, and $j \in E_i$ is said to be *friend* of i. We assume that friendship disregards colors and thus, in general, $E_i \cap \mathcal{B} \neq \emptyset$ and $E_i \cap \mathcal{R} \neq \emptyset$, for each $i \in V$.

Links are assumed to be symmetric, so that $j \in E_i$ if and only if $i \in E_j$, for each $i,j \in V$. Briefly, i and j are *friends*.

In the general framework described in the previous Section, we can formalize the N-order squared adjacent matrix $\mathbf{A}$ in (1) by setting

$$a_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are friends;} \\ 0, & \text{otherwise.} \end{cases}$$

The distance between two nodes $i,j \in V$, namely $d(i,j)$ is defined as a standard spatial Euclidean distance and it is conveniently normalized to one, so that

$$d(i,j) = \frac{\sqrt{(c_i - c_j)^2 + (r_i - r_j)^2}}{n\sqrt{2}} \mathbf{1}_{\{j \in E_i\}}, \tag{4}$$

where $c_\star$ ($r_\star$, resp.) is the column (the row, resp.) of the node $\star$, with $\star = i,j$.

Notice that when i and j are friends, then $d(i,j)$ represents the length of the arc (i,j). Moreover, by construction, $d(i,j) \in [0,1]$.

The measure of the friendship set E_i of an agent i, namely $\Delta(E_i)$, can be defined in several ways, any of them bringing information on the role played by the friendship in the preferences of the agents populating the network. We propose here three scenarios:

- $\Delta_m(E_i) = \min\{d(i,j) | j \in E_i\}$;
- $\Delta_M(E_i) = \max\{d(i,j) | j \in E_i\}$;
- $\Delta_f(E_i) = \frac{1}{|E_i|} \sum_{j \in E_i} d(i,j)$.

In a more general sense, we can define the measure of the friendship set E_i when i's friends are viewed from another node e of the graph. Such a concept aims at describing how an agent can change her/his closeness to her/his friends by modifying her/his position in the graph. So, we define the measure of the friendship set E_i by the perspective of a node $e \in V$ as follows:

- $\Delta_m(e, E_i) = \min\{d(e,j)|j \in E_i\}$;
- $\Delta_M(e, E_i) = \max\{d(e,j)|j \in E_i\}$;
- $\Delta_f(e, E_i) = \frac{1}{|E_i|}\sum_{j \in E_i} d(e,j)$.

It is evident that $e = i$ implies that $\Delta(e, E_i) = \Delta(E_i)$, for each $\Delta = \Delta_m, \Delta_M, \Delta_f$.

We naturally assume that agents feel to be more satisfied about their connections (and surrounded by friendship) as the value of the $\Delta(E_i)$'s is lower. Thus, a simple interpretation of the proposed measures can be easily derived.

The minimum case Δ_m describes a situation in which agent i feels that her/his social connections are generally of satisfactory type if at least one of her/his friends is close to her/him. Differently, the maximum case Δ_M requires that all friends are close to the agent to let she/he be satisfied with her/his connections. The fair case Δ_f is in the middle of such extremes, as one intuitively grasps.

Agents are also assumed to be sensitive to the diversity of the neighborhood. Specifically, each agent has a disutility in being surrounded by a majority of agents showing a color different from her/his own one. We formalize this assumption.

For each node $i = v_{r_i c_i} \in V$, which is then at the intersection of row r_i and column c_i, we define the *neighborhood set of* i by

$$N_i = \{v_{rc} \in V \setminus \{i\}|(r,c) \in \{r_i - 1, r_i, r_i + 1\} \times \{c_i - 1, c_i, c_i + 1\}\}, \tag{5}$$

with $r_i, c_i = 1, \ldots, n$, with the conventional agreement that v_{rc} is out of the square lattice, i.e. it is not a node, when $r = 0$ and/or $c = 0$.

By construction, we have that

$$|N_i| = \begin{cases} 3, & \text{if } (r_i, c_i) \in \{1, n\}^2; \\ 5, & \text{if } (r_i, c_i) \in \{1, n\} \times \{2, \ldots, n-1\} \text{ or } (r_i, c_i) \in \{2, \ldots, n-1\} \times \{1, n\}; \\ 8, & \text{otherwise.} \end{cases}$$

For each $i \in V$, we introduce two variables ξ_i^B and ξ_i^R which describe the colors of N_i as follows:

$$\xi_i^B = \frac{|N_i \cap \mathcal{B}|}{|N_i|}, \qquad \xi_i^R = \frac{|N_i \cap \mathcal{R}|}{|N_i|}.$$

Agents move in the graph on the basis of the utility functions of their positions $u_V : \{\mathcal{R}, \mathcal{B}\} \times \{1, \dots, n\}^2 \times \{0,1\}^N \to [0,1]$ and of the empty cells $u_\emptyset : \{\mathcal{R}, \mathcal{B}\} \times \{1, \dots, n\}^2 \times \{0,1\}^N \to [0,1]^{N-R-B}$, defined as follows:

$$u_V(\star_i, r_i, c_i, \mathbf{A}) = \frac{1}{2}\left[\alpha \xi_i^\star \mathbf{1}_{\{\xi_i^\star > 1/2\}} - (1-\alpha)\Delta(E_i) + 1\right]$$

and

$$u_\emptyset^{(e)}(\star_i, r_e, c_e, \mathbf{A}) = \frac{1}{2}\left[\alpha \xi_i^\star \mathbf{1}_{\{\xi_i^\star > 1/2\}} - (1-\alpha)\Delta(e, E_i) + 1\right],$$

where $\star = \mathcal{B}, \mathcal{R}$ and $\star_i$ is the color of agent i, $\Delta = \Delta_m, \Delta_M, \Delta_f$ and $\alpha \in (0,1)$.

Functions u_V and $u_\emptyset^{(e)}$ are defined under the same rationale. They grow with the percentage of neighborhoods of agent's same color as the agent is surrounded by a majority of such individuals, and decreases with the distance of the agent from her/his friends.

4. Experiments