Article Review

(Graph theory)

Network effects in Schelling's model of segregation: new evidences from agent-based simulation

Date of report: 1/12/2010

Date of exercise: 8/12/2010

Author: Elsa Guillot

Instructors: Eric Thierry
Christophe Crespelle

Contents

1 Introduction

Though graph theory has only recently been studied such as it is now, it has long been a very intuitive tool for some problems. Among these problems are transport problems and urban planning. In this case, graphs are a natural representation of the problem (roads as egdes, crossroads as nodes). We can easily figure properties in such cases. For example a weight on the edge can represent the capacity of the road (on a highway more people can travel within the same time). With urban planning, graphs are a good way to illustrate the heterogeneity of access of different areas. The ghettos of a city are isolated areas, densely intraconnected, but related to other parts only by a few lanes. If we represent a city as a graph, ghettos will then form a clique. Hence, attempts have been made to explain through graph modelling urban phenoma such as segregation. Urban segregation is the exclusion, mutual or not, of a part of the population, leading to a partitioning of the city. Each population inhabitates one area while mixing is a rare phenomenon. This have been mostly observed in large U.S. city, though it also appears in Europe. In New York for example we can observe black and white quarters, but also, nowadays, hispanic or chinese quarters. Scientists have tried to model the segregation phenomon, based on the tolerance level of people in order to find the critical parameter that would change this reality.
In a first time we are going to present Schelling's model of segregation, then the work of A. Banos putting it into a graph structural problem. Finally we will study the agent based simulations that have been done to comprehend the model.

2 Models of segregation

2.1 Schelling's model of segragtion

In 1969, Thomas C. Schelling produced a paper on segration model ([5] cited 432 times). By its simplicity and its **desarment** results, it soon became a reference in segregation works, and its properties have since been explored. This model is not yet a graph problem as it is an unidimensionnal reflexion.
The model is based on a twofold, exhaustive and recognizable distinction in the population. That could be black and white, but also boys and girls, catholics and protestants represented as + or O for more visibility (instead of 0 and 1 for exmple). Finally the individuals must have some preferences about living among their own kind and must be able to move to satisfy these preferences. The basic example is a line of people (for example as during a dinner) :

+O+O+O+O+O+O
This is the kind of configuration that often happens at table with men and women. Here we can see that if the individuals only look at their closest neighbours, and if they prefer being among their own kind, then they are necesserally unhappy. However if they look at not only the closest neighbours, but also the second-closest, then they are among a group of 50 % of their own kind. This simple example, extracted from [5] already allows us a few remarks :

- The happiness of the individual is a decisive citeria that must be properly defined
- The definition of neighbour is essential when talking about individual happiness and choice
- The unidimensionnal representation is strongly limited when it comes to representing a urban segregation phenomenon

In his model, Schelling also introduces two limitations to this kind of twofold representations :

- The small numbers contraints : if you only have one individual, it can only be white or black. If you have a couple (married for example) there are only two possibilities : a 50% mixture (black and white) or a total segregation (black and black, white and white).
- The simple numerical constraint : in a given group, both population cannot be majoritary. For the same obvious numerical reasons, both group can't ask for respectively 60% and 45% minimum occupation. Obvious as it can seem, it has been underlined, for this is one of the main segregation problem based on indidual choice.

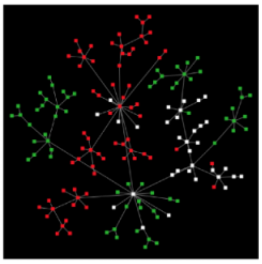
In a later paper (1971 [6] cited 1336 times), Schelling looked into his model with greater care. He notably started talking about problems linked to the extension of the model to more than one dimension, and introduced a concept of movement. If an individual is not happy with his neighbours, he is going to try to move to a better place. It is from this later article that A. Banos produced his work with graph theory tools.
In both cases Schelling's conclusion was very pessimistic : segregation is very often the only stable equilibrium that the system reaches. After publication, Schelling model became very popular, by its simplicity and its striking results. Many people discussed its social implications. Clark [3](2008) in particular tried to develop its relevance regarding sociological segregation, including some mathematical studies of the phenomenon.

2.2 Graph modelling

The Schelling model can be presented as a graph : each individual is a node, an edge represents a connexion between two individuals. The neighborhood notion we previously mentionned matches with the neighbour notion of graph theory :

In a graph, a vertex u is an adjacent vertex of a vertex v if they are connected by an edge.
In a graph G, the neighbourhood of a vertex v is the subgraph of G induced by the restriction to v and the neighbours adjacent to it.

In this model,the graph is unweigth and non oriented. The edges of the graph will have to represent places, that can be occupied by people, or not. Indeed, we need the people to be able to move, so we also need some free places to go. Here is a figure ([2])showing a simple example of such a graph, the populations are red and green, the free nodes are white:



One can only move to a free place in his neighbourhood. If someone is satisfied with his neighbourhood, he will stay. If he is not, he will try to move to a free place where he will be satisfied.

Here, we shall define P_{ij}, the proportion of people unlike the agent A_i among its neighbours, at the node j.
The individuals prefer to live among people of their own groups with a tolerance criteria. This citeria λ is a number between 0 and 100. People are happy as long as they live among a population with at least $100 - \lambda$ % of their own kind. We thus define a satisfaction criteria, with value 1 is the person is happy with its current location, 0 otherwise:

$$\mathrm{U}_i = \begin{cases} 0 \text{ if } \mathrm{P}_{ij} > \lambda \\ 1 \text{ if } \mathrm{P}_{ij} \leq \lambda \end{cases}$$

Finally we define a mixity index, being simply the mean proportion of contacts between unlike neighbors in the all graph.

$$M = \frac{1}{|A|} \sum_{i=1}^{|A|} P_{ij}$$

In the present study 80% of the nodes are inhabited by persons in order to allow enough space for movement. Half of the population is white, the other half is black. Schelling showed that the proportion of each type matters, and that the more we keep away from the 50-50 proportion, the more segregation occurs.
Banos's theory [1] is that the graph structure is very important in the segregation phenomenon. He defines four graph structures that he will compare with his simulations : Regular (chessboard kind), random, scale free and fractal. The scale free sctructure is actually the more accurate representation of reality. Fractal pattern was the first intuitive shot in trials for representation, although it can't win against scale free, it remains very close in its property.

In order to be able to make some comparisons between the graphs, we must enforce that each node is connected to its n closest neighbors, as defined by Floyd routing algorithm [4].

Shortest path algorithm - R. Floyd
procedure shortest path(m,n); value n; integer m; array m;
comment Initially m[i,j] is the length of a direct link from point i of a network to point j. If no direct link exists m[i,j] is initially 10^{10}. At completion, m[i,j] is the length of the shortest path from i to j. If none exists, m[i,j] is 10^{10}.
begin
integer i,j,k; **real** inf, s; inf:=10^{10}
for i=1:n **do**
for j=1:n **do**
if m[j,i]<inf **then**
for k=1:n **do**
if m[i,k]<inf **then**
begin s=m[j,i]+m[i,k];
if s<m[j,k] **then** m[j,k]=s
end
end shortest path

This simple algorithm enables us to find the shortest path between two points i and j, with a $0(n^3)$ complexity. The algorithm compares all possible paths between each pair of vertices in the graph. This very powerful algorithm is commonly used in graph theory.

Finally, the network is characterized by a clustering coefficient :

$$C_i = \frac{2E_i}{k_i(k_i-1)}, 0 \leq C_i \leq 1$$

with E_i the number of connected pairs among neighbours of node i
and k_i the degree of node i.
We compute an average indicator on the graph stucture :

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

This coefficient corresponds to the presence of the intraconnected component. If there were is only a clique for example, it will be 1. The closer to one it is, the more clique-like the structure. Here is a figure presenting the different graph structure, ordered by the Average cluster coefficient $\bar{C}$. The four networks represented here have a degree 10. We can compute there custering values:

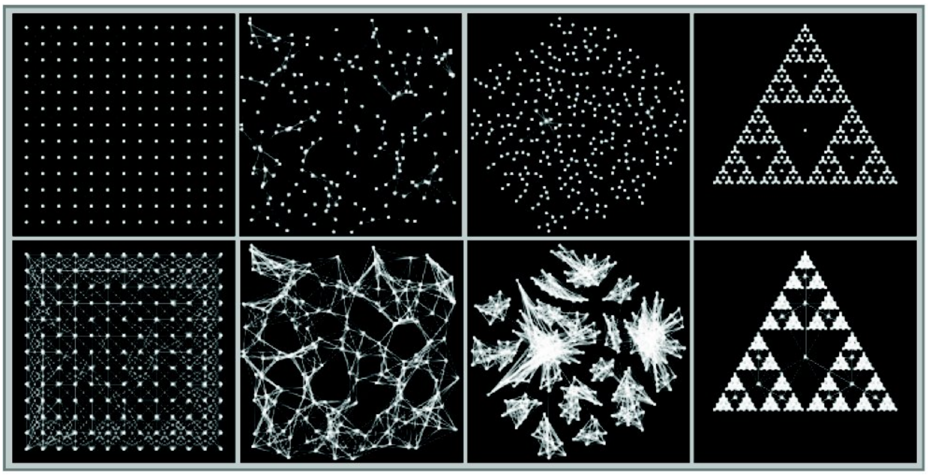


Figure 1: Regular, random, scale free and fractal (Sierpinski) and their corresponding neighbouring graphs (bottom) for a fixed degree (d=10; neighbourhood defined using the shortest path algorithm)

	$Grid$	$Random$	$Scale-free$	$Sierpinski$
$Numberofnodes$	361	361	361	363
$Numberofedges(structuralgraph)$	1332	1448	360	363
$Numberofedge(neighbohringgraph)$	2039	1749	1960	2134
$Clusteringcoefficient$	0.55	0.65	0.99	0.95

3 Agent-based simulation for model of segregation

An agent-based simulation was computed in order to make the graph evolve. It allowed A. Banos [1] to play with the parameters and the structures he wanted to test.
In the agent-based simulation the agents are the people who move according to the description above. The dynamic of this model is asynchronous. A random order is defined at each iteration, while each of them decides, one after the other, to move, or not. The simulation keeps running until there is no movement anymore. This can happen in two cases : each agent is satisfied with its current location, or the graph is in a frozen state. In the last scenario, although they are unhappy, the agents cannot move anymore as no accessible location satisfies them.
At the end of the simulation, the mixity index is computed to evaluate the segregation. This

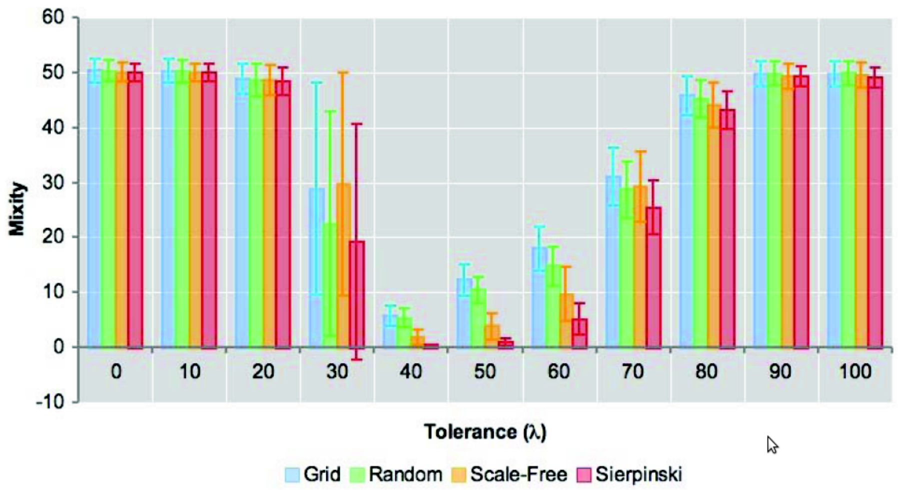


Figure 2. Average mixity index and its standard-deviation (1000 simulations), for each of the four networks

figure represents the main results of Banos's simulation on this model [1]. We can see for each structure, and different levels of tolerance of agents, the mixity index represented as a boxplot. Several conclusions can be reached from this plot.
First of all, we can clearly see an effect of the structure on the mixity index when all the other parameters are equals. $\lambda = 30$ and $\lambda = 60$ are the most striking examples.
Second of all we can define this plot in three parts :
- $\lambda < 30$: the mixity is very high. What happens is that the tolerance is too low, the agent can't move to be more satified and stays unsatisfied. Those are representing frozen states.
- $30 < \lambda < 70$: the mixity index is very low, segregation is occuring very strongly
- $\lambda > 70$ the tolerance is very high. There is no more segregation occuring.
So we can see that even if the agent are pretty tolerant (ok to live among 60 % of the other kind)

then segregation stills occurs. In this case the scale free and fractal structure are making the segregation worse. On each *lambda* value we can make an order of which stucture is the worst for segregation. It appears that the order is always the same : regular, random, scale free then fractal. Let's observe the following figure representing the final states : Here we can observe that patterns

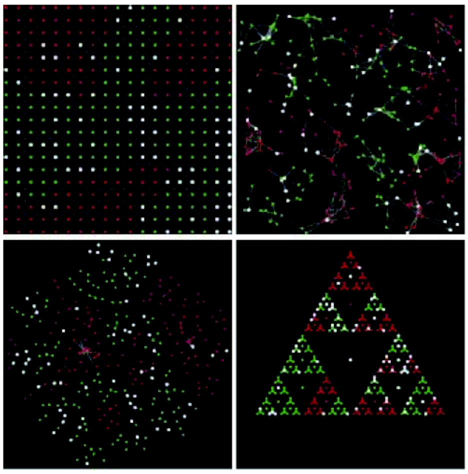


Figure 3. Examples of equilibriums attained for mild preferences (λ = 50, Density = 80%, Degree = 10, empty squares = vacant nodes)

are clearly related to the neighbouring graph. Cliques are acting like segregation traps. Once a color or a kind takes over the majority of a clique, then it becomes completely populated with this color. The more cliques there are, the more segregation occurs.

One of the problem that we see in these simulations is that for a small λ (<30) the system is always in a frozen state. In order to loosen the movement, we introduce a noise in the decision with a parameter N. At each step, if the agent is unsatisfied and cannot move, we generate a random number n from a uniform distribution. If $n > N$ then the agent will move to a randomly chosen vacant node.

Banos's paper [1] shows that a very small amount of noise (N=0.1) loosens the system, enough to get over the frozen state and reach equilibrium in most cases. Here is the new figure presenting the results :

We can see that for scale free and Spierpinski's fractal network, the equilibrium state is always reached. We can also notice that the noise only affects the simulation when < 30 and then becomes a non significant parameter.

So the simulation can be seen as an optimization algorithm to find a pattern, satisfying the whole population, in scale free-like or Spierpinski-like graphs, with a very intolerant λ. This algorithm may have many interesting other apllications, far from segregation problems.

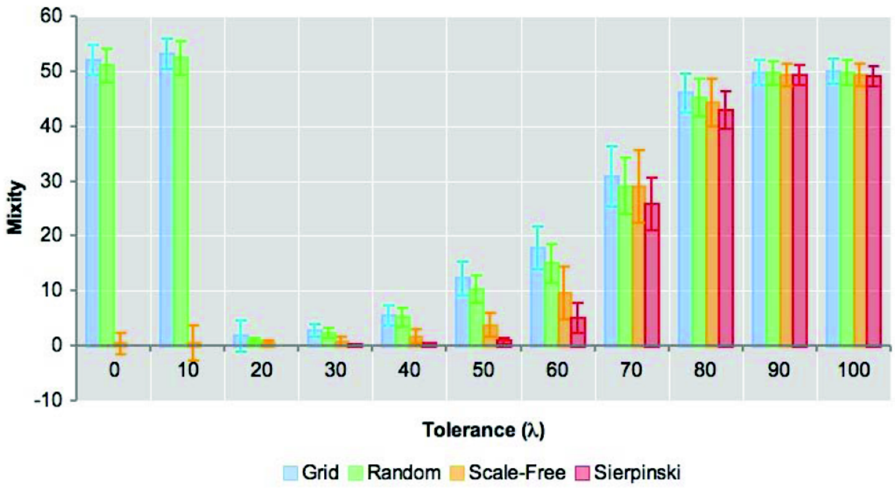


Figure 5. Average mixity index and its standard-deviation (1000 simulations) for each of the four networks, with noise in the system (N = 0.1)

Segregation satisfaction algorithm

input : a graph $G = (V, E)$, A_i a twofold population of $n = 0.8|E|$ agents, d the degree of the graph, λ the tolerance of the agent to be surrounded by a proportion of vertices populated with the other kind, $0 < N < 1$ a noise parameter.

While $\sum_i U_i < |A_i|$ do :

```
for i=1:n:
.        if U_i<1:
.            for j=1:d:
.                if (P_ij > lambda) and (node j is empty):
.                    move to j
.                    break
.        if U_i<1:
.            r=random(0,1)
.            if r<N :
.                for i=1:d:
.                    if node j is empty:
.                        move to j
.                        break
return G = (V, E), A_i
```

4 Conclusion

The segregation problem is very good example of the power yieldied by graph theory. The formalization of the problem with graphs provides lots of mathematical tools for a formal study of the problem (neighbours and mixture definition for example). The model presented by Schelling had a very interesting conclusion about the unavoidability of segregation.
With the graphical formalization made by A. Bannos, and its agent-based simulation, we have identified the critical values of parameters. The tolerance parameter λ has an overriding role. We can distinguish three intervals for this parameter : low $[0-30]$, mild $[30-70]$, and high $[70-100]$. The low and mild values are leading to systematic segregation phenomena. We also saw that the structure of the graph can reinforce the segregation phenomena. The fractal Sierpinski graph has been the one, among those studied, that makes segregation occur the most.
Finally we can conclude that we created an optimization algorithm to satisfy each agent in a given graph when a small noise is added.
Though the sociological conclusion is a bit pessimistic, as the segregation is the stable equilibrium of most system, we also showed that the structure can have an influence. This could be used in urban planning. If we want to avoid segregation, we should avoid cliques (ghettos) and try to extend the graph to give it a clustering coefficient very low : make highway to connect the whole town, and never let a part of the graph be scarcely connected to the rest.

References

[1] A. Banos. Networks effects in schelling's model of segregation : new evidences from agent based simulation. 2010.

[2] Arnaud Banos. A la recherche dâeffets reseaux dans la dynamique du modele de segregation de schelling. *Cybergeo : European Journal of Geography, Systemes, Modelisation, Geostatistiques*, (492), 2010, URL : http://cybergeo.revues.org/index22976.html.

[3] FossetM Clark A. Understanding the social context of the shcelling segregation model. *PNAS*, 105-11:4109-4114, 2008.

[4] Floyd RW. Algorithm 97 - shortest path. *Communication of the ACM*, 5(6: 345), 1962.

[5] Thomas C. Schelling. Model of segregation. *The american economic review*, 59(2), 1969.

[6] Thomas C. Schelling. Dynamic of model of segregation. *Journal of mathematical sociology*, 1, 1971.