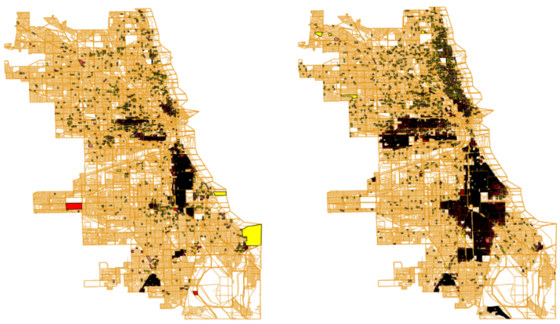
Schelling's Dynamic Model of Segregation
Complex Networks, Instituto Superior Técnico
Pedro Ascensão, 78961 and João Rodrigues, 78672

1 Introduction

Segregation is the separation of humans into groups in the daily life, and has been a constant and hard to deal with social problem. Segregation can occur in various manners: sometimes it results from the practices of organizations; other times it is deliberately organized. Some of it results from specialized communication systems, like different languages [1]. For example, if black people exclude white people from their church, or whites exclude blacks, the segregation is organized and deliberated. However, if black people happen to follow a different religion than white people, the two groups will be segregated Sunday morning whether they intend to be or not. Furthermore, if the church bulletin board is where people advertise rooms for rent, black will rent rooms from blacks and whites from whites because of a communication system that is correlated with churches that are correlated with color. Therefore, it is possible to conclude that segregation can happen even if it's not intended. Although people associate the term segregation with racial segregation (separation into racial or ethnic group) the term is so abstract that it can work for any division into two groups (boys and girls, students and faculty staff, teenagers and grownups, etc.)

The effect of segregation is striking when seen on a real map as shown on figure 1. This figure comes from a study from the authors Möbus and Rosenblat, on where they study the process of ghetto formation [2]. It is possible to say that th biggest reason for this effect, is that people like to live near others like them, and as consequence they open shops, restaurants, and other businesses oriented toward the populations of their respective neighborhoods, bringing even more people with the same characteristics.



(a) *Chicago, 1940* (b) *Chicago, 1960*

Figure 1: Maps of Chicago from 1940 and 1960. In yellow nd orange blocks the percentage of African-Americans is below 25, while in blocks colored brown and black the percentage is above 75. Figure from [2]

In 1971, Thomas Schelling created an agent-based model explaining why segregation is so difficult to combat. As referred above, there are many factors that contribute to segregation in real life, however Schelling's model focuses on a simplified mechanism to illustrate how the forces leading to segregation are robust and can even operate when no one individual explicitly wants a segregated outcome [1]. Therefore, and in conclusion, this model will be focus of the group's work.

1.1 Goal

The goal of this report is to present a computer simulation of the Schelling's model. Furthermore, the group wanted to extend the work of Schelling and study how the model worked in different settings, for example, in a scale-free network context.

1.2 Document outline

The rest of the report is structured as follows. Section 2. explains theoretically the Schelling's model. Section 3. presents two different implementations in Python, and their results and implications. Section 4., explains future work that can improve this project. Finally, Section 5. compares the results from both implementations and presents some conclusions.

2 Schelling Model

The general formulation of the model is as follows. There is a population of individuals, who will be called *agents* from now on, which can one of two types, **blue** or **red** agent. These two types represents some characteristic that is the reason for the difference between agents, for example race, ethnicity, country of origin, or native language. According to the model, these agents reside in the cells of a grid, which serves as a model of the two-dimensional geography of a city, furthermore, this grid can be either seen as a matrix or a graph. As illustrated in the Figure 2, it is assumed that some cells of the grid contain agents while others are unpopulated. A cell's neighbors are the cells that touch it, including diagonal contact; therefore, a cells that is not a boundary of the grid has eight neighbors. [3]

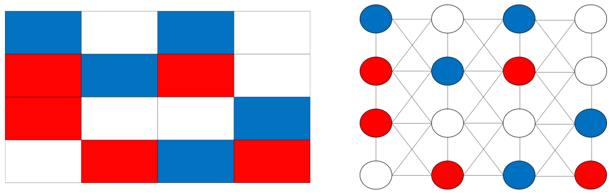


Figure 2: Agents of two different types (reds and blues) occupy cells on grid. The neighbor relationships among the cells can be represented simply as a matrix or a graph.

The fundamental constraint is that each agent wants to have at least some other agents of its own type as neighbors, therefore is assumed that there is a threshold t common to all agents: if an agent has fewer than t% of its neighbors are of the same type as itself, then the agent is *unsatisfied* and has an interest in moving to a new cell. The dynamics of movement are simple to understand. Agents move in a sequence of rounds: in each round, it is considered unsatisfied agents, and for each one in turn, they are moved to an unoccupied cell. The sequence of rounds finishes when all agents are satisfied with their position on the grid. [3]

To compare results in different settings, it was needed a segregation measure. Unfortunately, the original model by Schelling does not present any measure of that kind. After browsing some websites with computer simulations, the group ran into [4] that has graph with a segregation measure alongside the simulation. Although the website does not explain its segregation measure, it was possible to discover it after analyzing the *javascript* of the website. That being said, the segregation measure, presented in the equation below in where N is the total number of agents, is the mean of the *similarity* of each agent. *Similarity*, $similarity_i = \frac{neighbors_i^{type=i}}{neighbors_i}$, is the ratio of neighbors of the same type among all neighbors (unoccupied cells do not count for the total of neighbors). This measure makes sense, since all agents are all surrounded by agents of their type, the segregation will be equal to 1. Furthermore, the more mixed together, lesser will be the value of the measure.

$$segregation = \frac{\sum_{i=0}^{N} similarity(i)}{N} \tag{1}$$

In conclusion, Schelling model is a simple model to understand and implement, however is a good representation of how different characteristics can lead to segregation.

3 Simulation

In the implementation of this model we focused on two different approaches. The first one was the default Schelling Model, where the society is displayed in a grid layout and each grid position can either have one of two races or be an empty space. Then, as an extension to the model, we thought that it would be interesting to analyze the same concepts but in a graph layout. In particular, we modeled the simulation in a Barabási-Albert graph, where each node can also have one of two races or be an empty space. This sounded particularly interesting because this way of displaying the modeled society is closer to a real society and therefore can help comprehend and analyze this phenomenon of racial segregation.

As measures for comparison and analysis of the simulation, we chose the **number of steps** that the model takes to reach 100% satisfiability, (i.e. a state where every node is satisfied), the **slope of the curve** in the plot of unsatisfiability percentage over the number of steps and the **segregation measure**, presented in Section 2.

As for particularities of the implementation, on each step of the simulation the algorithm will first check all nodes and infer the unsatisfied ones. After all unsatisfied nodes are flagged, the algorithm will move them one by one to an empty position until all are moved. The position that a node leaves is immediately considered as an empty node and can be occupied in the same iteration by another node. The way the empty nodes are chosen to move and unsatisfied node is completely random, i.e. doesn't follow any heuristics to choose a preferable empty position.

Different parameters can be altered to reach different simulation results. The parameters are:

1. **Size**: The size of the network. In the matrix layout the size is the side of the matrix (i.e. matrix is actually size $\times$ size) and in the graph layout, the size is the number of nodes.
2. **First race percentage**: The percentage of nodes that belong to the first race.
3. **Second race percentage**: The percentage of nodes that belong to the second race.
4. **Threshold**: The threshold determines what is the minimum percentage of nodes of the same race a node permits before feeling unsatisfied, i.e. if the threshold is 30%, a node is satisfied if at least 30% of its neighbors belong to the same race. Empty positions don't count as neighbors.

The percentage of empty positions, although it isn't explicitly provided, it is what is left of the total number of nodes: $100\% - (First\ race\ percentage + Second\ race\ percentage)$

3.1 Grid Model Analysis

The default Schelling model is composed of two races and empty spaces in a **grid layout**. In this layout, every node has 8 neighbors (the direct nodes surrounding it), except those on the borders or corners of the grid, since it is not a toroidal grid. We will now analyze the behavior of the simulation in the grid layout.

3.1.1 Grid simulation

In the beginning of the simulation, a random grid is created with the provided parameters. The following grid was created with a size of 100 by 100 positions, a first race and second race percentage of 40%, empty positions are therefore 20% and a threshold of 60%. The initial matrix in represented in Figure 3

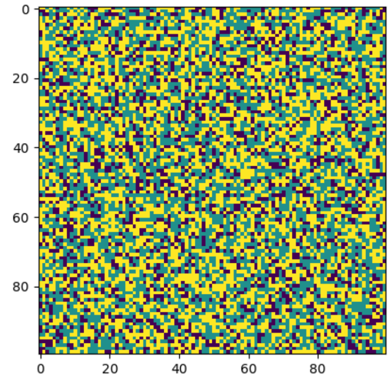


Figure 3: The initial matrix

After 38 steps, the simulation reaches the end, where every node is satisfied and we get the following results

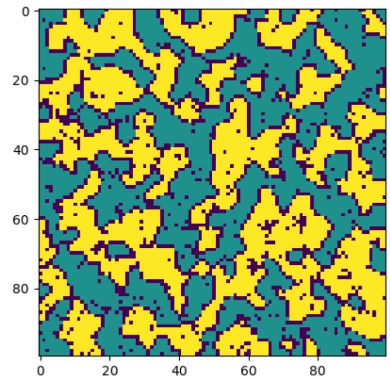


Figure 4: The final matrix

As we can observe, a lot of clusters are formed and the segregation measure is close to 1, being that 1 is the maximum. If we change the parameters around a bit we can compare results. Let's change the threshold to only 30% and run another simulation. We get the following results:

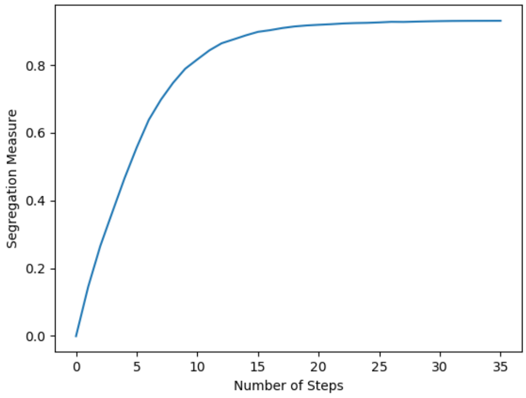


Figure 5: Segregation Measure over Steps

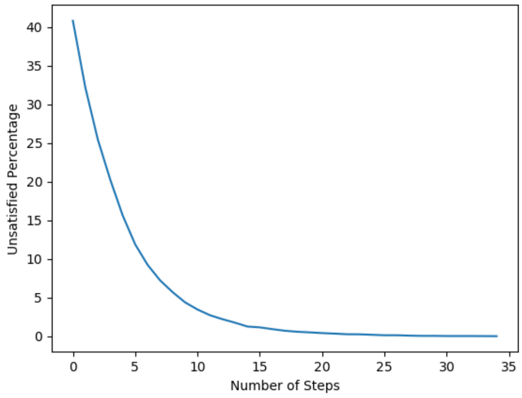


Figure 6: Unsatisfied Percentage over Steps

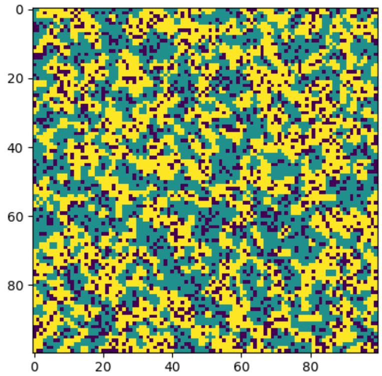


Figure 7: The final matrix with lower threshold

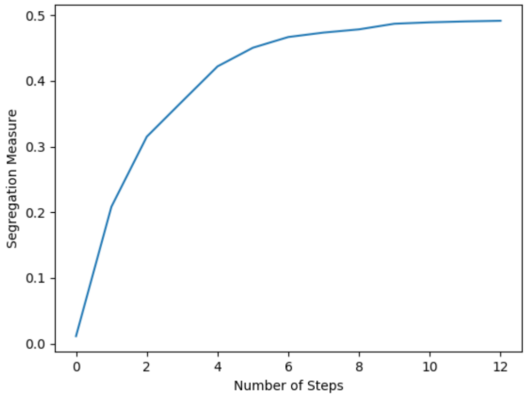


Figure 8: Segregation Measure over Steps

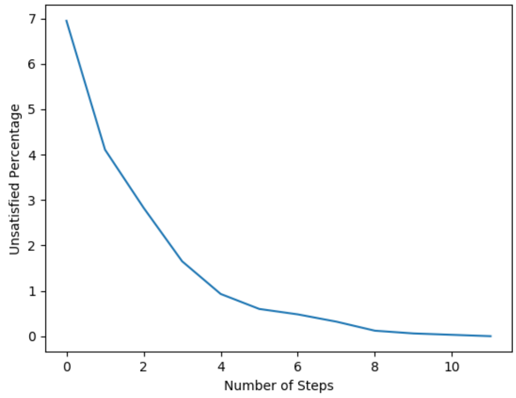


Figure 9: Unsatisfied Percentage over Steps

We can now compare the results. With a lower threshold, a lot less clusters are formed, it is even somewhat hard to identify the clusters, since the nodes are a lot more mixed up. The number of steps was lowered to almost half, from 38 steps to only 12. The segregation measure, as expected, was also a lot lower, from almost 1 to around 0.5. The unsatisfied percentage also starts at a lower value (around 7% vs around 40%)

3.2 Graph Simulation

To create the graph society we used the `NetworkX`[1] Python library. With this library, we were able to create a Barabási-Albert graph, where each new node starts with 4 connections, to match the mean number of neighbors of the grid layout, which tends to 8. It's harder to analyze the graph visually than it is to analyze the grid, we can only analyze it by looking at the generated measures of steps, segregation and unsatisfied percentage.

Let's first run a simulation with 2500 nodes, a first race and second race percentage of 35%, an empty nodes percentage of 30% and a threshold of 60%. We start with a random graph and after 112 steps reach 100% satisfiability. The results are as follows:

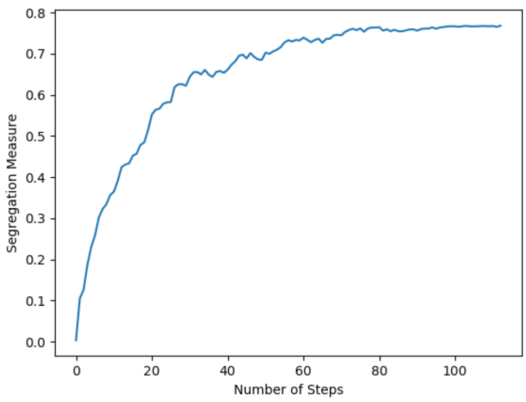


Figure 10: Segregation Measure over Steps

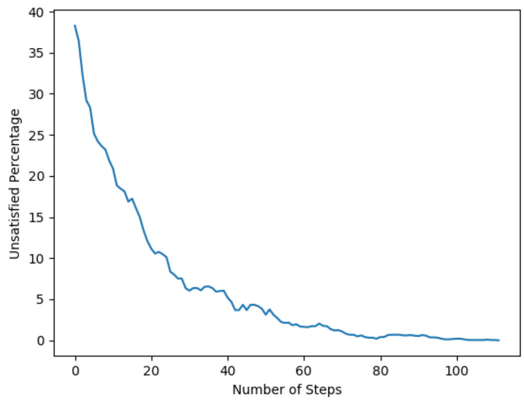


Figure 11: Unsatisfied Percentage over Steps

As we can observe, there are is a somewhat irregular behavior between step 20 and step 60, where there were some fluctuations on the slope of the curves, the process is not as smooth as we observed on the matrix layout. The segregation is close to 0.8. If we change the parameters around a bit we get some interesting results. For example, for the same number of nodes, same threshold and same number of empty spaces, if we create a minority situation, with the first race representing 60%, the second only 10%, the unsatisfied rate never (for the time that we tested, with over 10000 steps) reaches 0, getting stuck at around 8%. This might be due to the fact that it is very hard for the minority to find a suitable neighborhood. To make it possible to reach a stable state, we had to lower the threshold down to 20%.

If we lower the threshold from 60% to 30% and run the simulation with the parameters of the first simulation we get the following results, after only 6 steps:

As we can see, the number of steps are only a fraction of those observed before. We can then conclude that in this layout, which better represents a society, it might be very hard to reach stable states with a minority situation. The society must be willing to accept the minority and vice-versa. Furthermore, without minorities, if the society is willing to accept each other, a stable state is promptly reached.

[1]https://networkx.github.io/

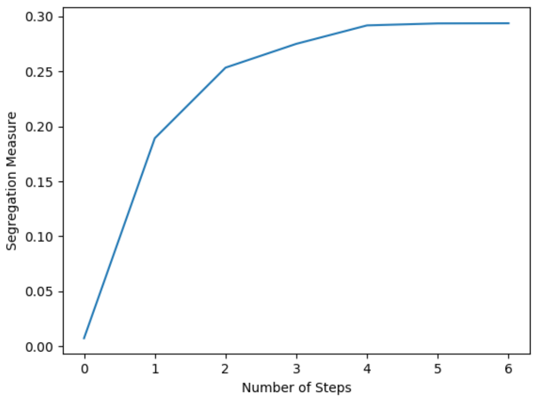


Figure 12: Segregation Measure over Steps

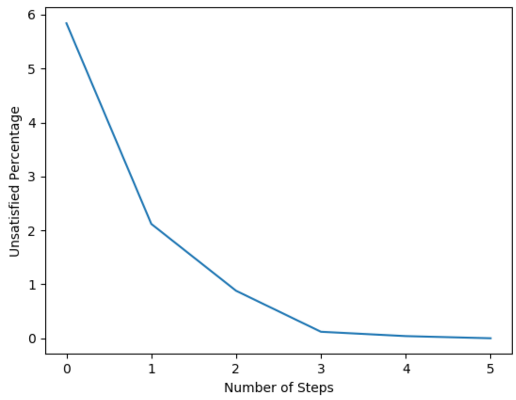


Figure 13: Unsatisfied Percentage over Steps

4 Future Work

In this section is presented two topics for future work that can improve the project. These topics are: *maximum threshold* and *visualization*.

4.1 Maximum threshold

In [4], it was identified a big implication of the Schelling Model, which is, if all agents are satisfied, lowering the threshold has no effect. This means that although people start to become more willing to live in minority, clusters/communities will stay the same.

The only way to dissolve these clusters, is to add another threshold, representing how much neighbors of the same type an agent wants. In reality, this is related to the increasing feeling of people to live in a diverse environment of people.

That being said, this measure has another implication in society; this means that communities are only dissolved if people make the effort to only settle in diverse environments. Therefore, although the original threshold represented a passive characteristic of the agents, this new threshold represents an active behavior of the agents.

4.2 Visualization

For future work, it would be nice if the user could visualize what was happening in each iteration, like how it is done in [4]. Although not difficult to implement for the matrix representation, using web-based languages, it is hard to implement for graphs. Representing static graphs with a lot of nodes in a way that the user can understand in a easy manner is by itself a hard task; representing graphs where in each iteration nodes change of its characteristic is exhausting task and out of the scope of the project.

Although this topic is not that important as the previous one, it would make the project more user-friendly and it would make possible to understand the effects of each iteration.

5 Comparison and Conclusion

As means of comparison, we can look at both layouts put its results side to side.

Let's consider a simulation with 2500 nodes (a 50 by 50 matrix), with 35% of each race and a threshold of 60%. Under these conditions, the grid society took 25 steps to reach a stable state while the Barabási-Albert society took 112 steps. Regarding the segregation measure, the matrix society always reaches a higher value, close to 1, while the graph society segregation only reaches around 0.8 of segregation. One can theorize that, in a closer to reality scenario it takes a greater amount of effort to get a satisfied society, when the threshold is relatively high. However, since on the graph layout, there is a high variability of number of neighbors, the segregation is consistently lower, since on most cases, a node is never surrounded exclusively by nodes of its race.

On the other hand, if we lower the threshold by half, the matrix society takes 11 steps to reach a stable state while the graph society takes only 6 steps to reach said state. This shows that, if from the beginning, the society is more willing to accept different races, a satisfied and less segregated society is possible to achieve.

As we've seen, a lower threshold results in a much more well mixed society. It's important to notice that we can only achieve a well mixed society if we start from a new random society. If we were to start from a segregated one, lowering the threshold wont produce a less segregated society since all the nodes would already be satisfied to start with and would never change its position. However, the opposite is not true. If we start from a segregated or semi-segregated society and increase the threshold, some of the nodes will get unsatisfied and will be forced to move, creating an even more segregated society.

In conclusion, although a simple model, Shelling model shows how segregation is such a difficult problem to eradicate

References

[1] Thomas C. Schelling. Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2):143–186, 1971.

[2] Markus M Mbius and Tanya S Rosenblat. The process of ghetto formation: evidence from chicago. 12 2017.

[3] Easley David and Kleinberg Jon. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[4] Vi Hart and Nicky Case. Parable of the polygons: A playable post on the shape of society.