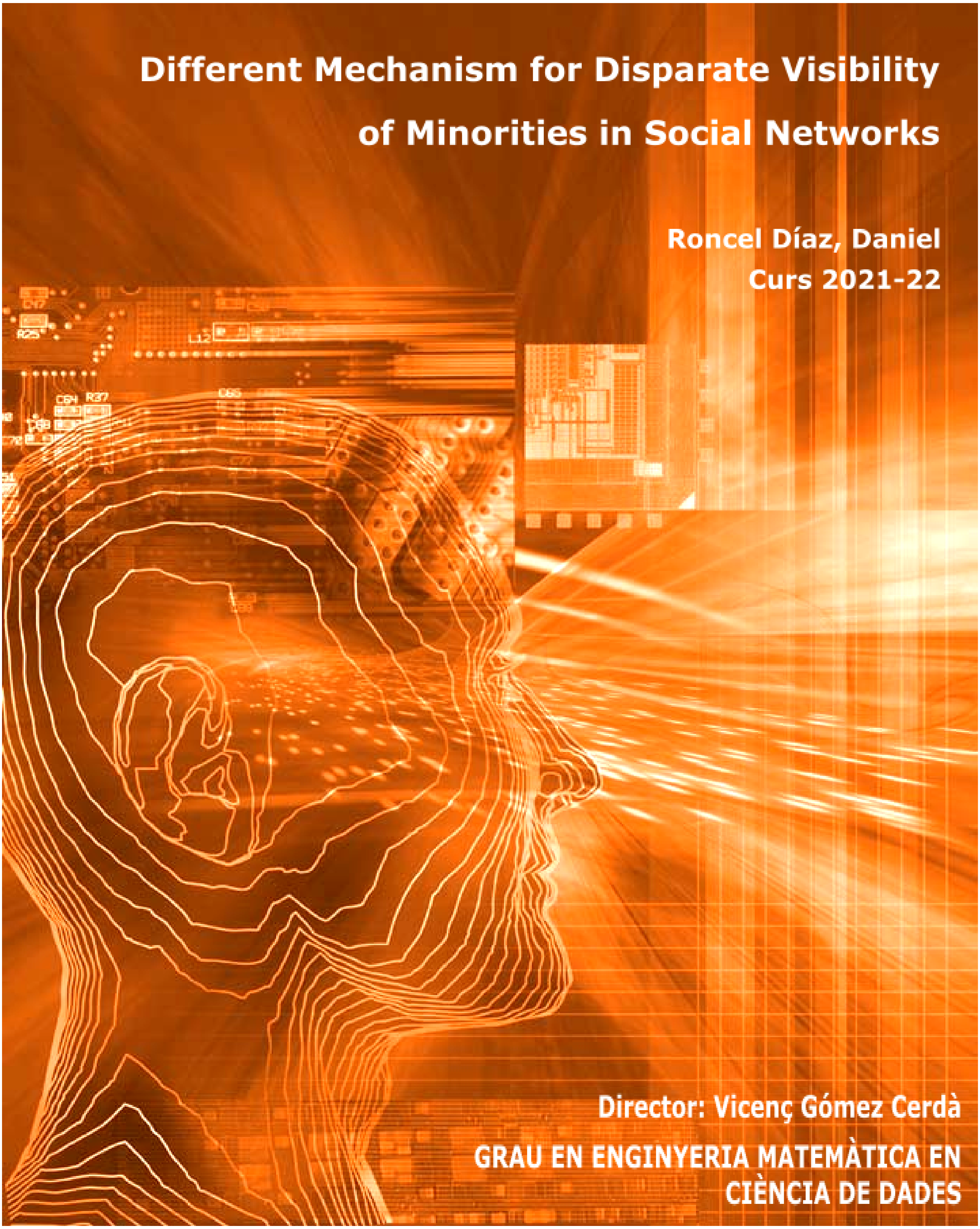


Different Mechanism for Disparate Visibility of Minorities in Social Networks

Roncel Díaz, Daniel

Curs 2021-22

Director: Vicenç Gómez Cerdà

GRAU EN ENGINYERIA MATEMÀTICA EN CIÈNCIA DE DADES

Universitat Pompeu Fabra Barcelona
Escola d'Enginyeria

Treball de Fi de Grau

Different Mechanism for Disparate Visibility of Minorities in Social Networks

Daniel Roncel Díaz

Undergraduate Thesis / Year: 2021 - 2022

Mathematical Engineering in Data Science

SUPERVISOR
Vicenç Gómez Cerdà

Universitat Pompeu Fabra Barcelona

Abstract

Recent studies have shown the existence of disparate visibility of groups in rankings. In [Germano et al., 2019] it is shown that in popularity-based rankings, the smaller the number of items of a particular group, the larger the share of traffic they can attract. This phenomenon has been coined by the term *few-get-richer*. Alternatively, [Karimi et al., 2018] show that, in social networks, the amplification of the traffic can be explained by means of homophily.

In this TFG, we analyse whether the reported phenomena have a qualitatively different nature or they are consequences of the same mechanism. In particular, we test through simulations (1) whether the disparate visibility observed in [Karimi et al., 2018] is a result of an amplification due to a network formation model; (2) whether in the absence of homophily in a network formation model, the disparate visibility of groups can also occur; (3) whether the *few-get-richer* effect can be observed in networks which do not depend on the homophily mechanism.

Our results are that (1) the network formation model used in [Karimi et al., 2018] amplify the disparate visibility of groups; (2) the disparate visibility of groups can be explained by models that do not depend on the homophily mechanism; (3) the *few-get-richer* effect can also appear in models which do not depend on the homophily mechanism.

Resumen

Estudios recientes han demostrado la existencia en rankings de grupos con una visibilidad desproporcionada a su tamaño. En [Germano et al., 2019] se demuestra que, en rankings basados en popularidad, cuanto menor sea el número de integrantes de un grupo, mayor puede llegar a ser el tráfico que estos atraigan, un fenómeno que ha sido denominado *few-get-richer*. Por otro lado, en [Karimi et al., 2018] se demuestra que, en redes sociales, la amplificación del tráfico que atrae un grupo puede ser explicada a partir del mecanismo de homofilia.

En este TFG testeamos mediante simulaciones (1) si el fenómeno observado en [Karimi et al., 2018] ha sido amplificado por el mecanismo de formación de red; (2) si, en redes que no dependan de un mecanismo de homofilia, pueden surgir grupos con una visibilidad desproporcionada; (3) si el fenómeno *few-get-richer* puede darse en redes que no dependen del mecanismo de homofilia.

Nuestros experimentos evidencian que (1) el mecanismo de formación de red usado en [Karimi et al., 2018] amplifica la visibilidad desproporcionada de los grupos; (2) que la visibilidad desproporcionada puede ser explicada por modelos que no dependen de un mecanismo de homofilia; (3) que el fenómeno *few-get-richer* puede darse en redes cuya formación no dependa del mecanismo de homofilia.

Resum

Estudis recents han demostrat l'existència a rànquings de grups amb una visibilitat desproporcionada a la seva mida. En [Germano et al., 2019] es demostra que, en rànquings basats en popularitat, com menor sigui el número d'integrants d'un grup, major pot arribar a ser el tràfic que aquests atreguin, un fenomen que ha estat anomenat *few-get-richer*. Per altra banda, a [Karimi et al., 2018] es demostra com, en xarxes socials, l'amplificació del tràfic que atrau un grup pot ser explicada a partir del mecanisme d'homofília.

En aquest TFG testegem mitjançant simulacions (1) si els fenòmens observats a [Karimi et al., 2018] han estat amplificats pel model de formació de xarxa; (2) si, en xarxes que es creïn sense un mecanisme d'homofília, es poden crear grups amb una visibilitat desproporcionada; (3) si el fenomen *few-get-richer* pot donar-se en xarxes que no depenguin del mecanisme d'homofília.

Els nostres experiments porten evidència que (1) el mecanisme de formació de xarxa emprat a [Karimi et al., 2018] amplifica la visibilitat desproporcionada dels grups, i que (2) aquesta visibilitat desproporcionada pot ser explicada per altres models que no depenen del mecanisme d'homofília; (3) que el fenomen *few-get-richer* pot donar-se en xarxes la formació de la qual no depengui del mecanisme d'homofília.

Extended abstract

Given the current importance of search engines, social networks and ranking systems, several researches have been done on the visibility of groups, such as individuals of a particular gender, political orientation, religion, etc. These researches have shown how some social mechanisms and the algorithms behind the platforms can amplify the disparate visibility of some groups, generating phenomena such as the *majority illusion*, the over-representation of some groups in high ranked positions, etc. Moreover, various measures have been proposed to mitigate these phenomena when the representation of a group does not correspond to the desired visibility. This Final Degree Project is inspired by two works in this field.

The first work which inspired this TFG is [Germano et al., 2019], in which the visibility of minority groups in popularity-based rankings is studied. The model described in [Germano et al., 2019] corresponds to a ranking in which websites are ranked based on popularity, i.e., the number of clicks they receive, in which websites are partitioned into two classes of different sizes. Moreover, users have heterogeneous preferences for the classes, so that some of them will only click on websites of the majority class, others will only click on websites of the minority class, and some others can click on websites of both classes. Using this model, the researchers showed that the disparate visibility of minorities in rankings can be explained by the preference distribution among users, the class sizes and the bias to click on popular items. The most surprising phenomenon, which they coined as *few-get-richer*, is that reducing the number of items of the minority class can lead to increasing the overall number of clicks of the class.

The second work which inspired this TFG is [Karimi et al., 2018], in which the visibility of minority groups in social networks is studied. The network formation model used in [Karimi et al., 2018] is a variant of the well-known Barabási-Albert model. In this variant, nodes have been partitioned into two classes, and the homophily mechanism regulates how frequent contacts between similar and dissimilar nodes, i.e., nodes of the same or different classes, occur. The experiments in [Karimi et al., 2018] showed that the minority class is on average over-represented in the heterophilic regime, i.e., when nodes tend to connect with nodes of the other class. However, the minority class is under-represented in the homophilic regime, i.e., when nodes tend to connect with nodes of the same class.

Interestingly, the fewer the number of nodes reporting a given signal, i.e., the fewer the nodes of the minority type, the higher the average degree of these nodes. We will name this phenomenon as the *few-get-richer* effect in networks.

There is a clear relationship between the model described in both papers: the most popular items (websites in [Germano et al., 2019] and nodes in [Karimi et al., 2018]) have a larger probability of becoming more popular, and the larger the affinity

(between a user and a website in [Germano et al., 2019] and between nodes in [Karimi et al., 2018]), the larger the probability of creating a connection. Finding whether the disparate visibility in [Germano et al., 2019] and [Karimi et al., 2018] are consequences of the same mechanism is the first objective of this TFG. To do so, we will test whether the preferential attachment mechanism, i.e., the bias to connect with the more popular nodes, is indispensable to generate the disparate visibility of groups in [Karimi et al., 2018]. If it is not the case, then we have evidence that the phenomena observed in [Germano et al., 2019] and [Karimi et al., 2018] are not consequences of the same mechanism. This is because the preferential attachment mechanism is necessary to observe this phenomenon in [Germano et al., 2019].

The second and third objectives consist in studying whether, in a model which does not depend on the homophily mechanism, the disparate visibility of groups and the *few-get-richer* effect can also occur.

To achieve the first objective, we define a network formation model which, in essence, only differs with the model used in [Karimi et al., 2018] in the absence of the preferential attachment mechanism: the nodes are partitioned into two classes of different sizes, and the homophily mechanism regulates the frequency of contacts between nodes of the same and different classes. At each time step, one node generates edges with some other nodes. Because of the absence of the preferential attachment mechanism, all the nodes of a particular class have the same probability of attracting a new edge.

Using the model described above, we study the representation of the minority class from two points of view: the overall degree of the nodes of the minority class and the percentage of nodes of this class among the most popular nodes of the graph.

Our result is that, while the disparate visibility of groups still occurs without preferential attachment, this mechanism is amplifying the disparate visibility. Actually, preferential attachment mainly benefits the minority class in the heterophilic regime, i.e., when contacts tend to occur between nodes of different classes. Moreover, the lower the number of nodes of the minority class, the larger the amplification. Then, because preferential attachment is not necessary to generate the disparate visibility in [Karimi et al., 2018], the disparate visibility observed in [Germano et al., 2019] and [Karimi et al., 2018] cannot be consequences of the same mechanism.

To achieve the second and third objectives, we define a network formation model inspired by [Germano et al., 2019]. Analogously to [Germano et al., 2019], nodes are partitioned into two classes of different sizes, and nodes have heterogeneous preferences for the classes, so that some of them will only generate edges with nodes of the majority class, others will only generate edges with nodes of the minority class, and some others can generate edges with nodes of both classes. From an initial graph, at each time step, one node is added to the graph and generates edges with some other existing nodes. Moreover, the probability of attracting an edge grows exponentially with respect to the degree of the nodes, instead of linearly as in the recurrent Barabási-Albert model.

Using this model, we study the effect of the class sizes, the preference distribution and the number of edges generated in the representation of each class. Particularly, we study the effect of these parameters in the overall degree of the minority and majority class nodes, their degree distribution and the advantage obtained with respect to scenarios in which their representation is proportional to their size.

Our main conclusions are that both the disparate visibility and the *few-get-richer* effect can be observed in networks generated by this model. Thus, both phenomena can appear in the absence of homophily.

Contents

List of figures **xii**

1 INTRODUCTION **1**
1.1 Background . . . 1
1.1.1 The *Few-get-richer* Effect . . . 1
1.1.2 Influence of Homophily in the Ranking of Minorities in Social Networks . . . 3
1.2 Objectives . . . 4

2 DISPARATE VISIBILITY IN THE ABSENCE OF PREFERENTIAL ATTACHMENT **5**
2.1 Network Formation Model Definition . . . 5
2.2 Experiments and Conclusions . . . 6
2.2.1 Results . . . 6

3 DISPARATE VISIBILITY IN THE ABSENCE OF THE HOMOPHILY MECHANISM **11**
3.1 Network Formation Model Definition . . . 11
3.2 Experiments and Conclusions . . . 12
3.2.1 Analysis without indifferent nodes . . . 13
3.2.1.1 Effect of the Preference Distribution . . . 13
3.2.1.2 Effect of the Size of the Classes . . . 14
3.2.1.3 The Degree Distribution . . . 14
3.2.2 Analysis with indifferent nodes . . . 16
3.2.2.1 Analysis for $P_{0,i} = P_{1,i}$ . . . 16
3.2.2.2 Analysis for $P_{0,i} > P_{1,i}$ . . . 18
3.2.2.3 Analysis for $P_{0,i} < P_{1,i}$ . . . 20
3.2.2.4 The *Few-get-richer* Effect . . . 23

4 CONCLUSIONS **27**
4.1 Further Work . . . 28

A FURTHER EXPLANATION OF PAP MODEL **29**
A.1 Numerical example of classes sizes and preference distribution . . 29
A.2 The Preference Distribution 30
A.3 Equivalent Model to PAP 31
A.4 Anomalies in the Degree Distribution 32
A.5 Degree Growth 34
A.6 Degree Distribution 35
A.7 Experiments to Study the *Few-get-richer* Effect 36

List of Figures

2.1 Schematic of the network topology 7
2.2 Representation of the minority class 8

3.1 Effect of the preference distribution and the size of the classes on the minority class degree when $P_{0,i} = P_{1,i} = 0$ 13
3.2 Representation of the classes for several group sizes when $P_{0,i} = P_{1,i} = 0$ 15
3.3 Minority class degree for $P_0 = C_0$, $P_1 = C_1$ when $P_{0,i} = P_{1,i}$ . . . 17
3.4 Schematic of the network topology for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 1$ 17
3.5 Minority class degree for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} > P_{1,i}$ 18
3.6 Representation of classes for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$ 19
3.7 Relative over-representation of the minority class for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$ 20
3.8 Majority class degree for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} < P_{1,i}$ 21
3.9 Representation of classes for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$ 22
3.10 Relative over-representation of the majority class for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$ 23
3.11 The *few-get-richer* effect when $m = 2$, $P_0 = 0.8$, $P_{0,i} = 0.8$ and $P_{1,i} = 0$ 24
3.12 The *Few-get-richer* effect when $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 0.6$ 25

A.1 Numerical example of class sizes and preference distribution . . . 30
A.2 Example of anomalies in the degree distribution 33
A.3 Degree growth for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 0$ 34
A.4 Degree growth for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 0.5$ 34
A.5 Degree growth for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 1$ 35
A.6 Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$ . . . 35
A.7 Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0.5$ . . 36

A.8 Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 0$ . . 36
A.9 Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$ . . . 37
A.10 Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 1$ . . 37
A.11 Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 0.5$ . . 38
A.12 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 0$. 38
A.13 Degree distribution for $P_0 = C_0$, $P_1 = C_1$ $P_{0,i} = 0.5$ and $P_{1,i} = 0.5$ 39
A.14 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 1$. 39
A.15 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$. 40
A.16 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0.5$ 40
A.17 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 0$ 41
A.18 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$. 41
A.19 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 1$ 42
A.20 Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 0.5$ 42
A.21 *Few-get-richer* experiment for $m = 5$, $P_0 = 0.8$, $P_{0,i} = 1$ and $P_{1,i} = 0$ 43
A.22 *Few-get-richer* effect experiment for $m = 9$, $P_0 = 0.8$, $P_{0,i} = 1$ and $P_{1,i} = 0$ 43
A.23 *Few-get-richer* effect experiment for $m = 2$, $P_0 = 0.9$, $P_{0,i} = 1$ and $P_{1,i} = 0$ 44
A.24 *Few-get-richer* effect experiment for $m = 2$, $P_0 = 0.8$, $P_{0,i} = 0.6$ and $P_{1,i} = 0$ 44
A.25 *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 1$ 45
A.26 *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 0.8$ 45
A.27 *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 0.6$ 46
A.28 *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0.5$ and $P_{1,i} = 0.5$ 46
A.29 *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.2$, $P_{0,i} = 0.5$ and $P_{1,i} = 0.5$ 47

1

INTRODUCTION

1.1 Background

Given the current importance of search engines, social networks and ranking systems, several researches have been done on the visibility of groups ([Gómez et al., 2021]; [Espín-Noboa et al., 2022]; [Fabbri et al., 2021]), such as individuals of a particular gender ([Chen et al., 2018]). These researches have shown how some social mechanisms and the algorithms behind the platforms can amplify the disparate exposure of some groups, generating phenomena such as the *majority illusion*, the over-representation of some groups in high ranked positions, etc. Moreover, other researchers have proposed measures to mitigate these phenomena when the representation of a group does not correspond to the desired visibility ([Singh and Joachims, 2018]).

This Final Degree Project is inspired by two works in this field that explain the dynamic popularity of items. This two works will be briefly explained in the two next subsections.

1.1.1 The *Few-get-richer* Effect

The first work which inspired this TFG is [Germano et al., 2019], in which the visibility of minority groups in popularity-based rankings is studied. The model described in [Germano et al., 2019] corresponds to a ranking in which websites are ranked based on popularity, i.e., the number of clicks they receive. Thus, the most clicked websites appear in the first positions. Moreover, users tend to click on the websites in the top of the ranking, which corresponds to the preferential attachment (or *rich-get-richer*) effect. Also, websites are partitioned into two classes of different sizes (e.g., left-leaning and right-leaning news sources). Furthermore, users have heterogeneous preferences for the classes. Thus, in the simplified version of the model, some of them will only click on websites of the majority class (group 0), others will only click on websites of the minority class (group 1), and some users are indifferent to the class of the websites, so they click only based in their position in the ranking (group 2). These three groups are represented with parameters γ_0, γ_1, γ_2, so that γ_i is the probability that an individual of group i clicks on an item of the majority class absent ranking. By the above explanation, $\gamma_0 = 1$, $\gamma_1 = 0$ and $\gamma_2 = 0.5$.

The propensity $\varphi_{n,m}$ with which individual n of class i clicks on item m is

$$\varphi_{n,m} = \begin{cases} \dfrac{\gamma_i}{|M_0|} & \text{if} \quad m \in M_0 \\ \\ \dfrac{1-\gamma_i}{|M_1|} & \text{if} \quad m \in M_1 \end{cases} \tag{1.1}$$

where M_0 and M_1 are the set of websites of the majority and minority classes, respectively.

In this model, at each time step, a user observes the ranking and, based on the user preference and the position of the websites in the ranking, the user clicks on a website. The probability individual n clicks on website m is

$$\rho_{n,m} = \frac{\beta^{(|M|-r_{n,m})}\varphi_{n,m}}{\sum_{m' \in M} \beta^{(|M|-r_{n,m'})}\varphi_{n,m'}} \tag{1.2}$$

where M is the set of all websites in the ranking and $r_{n,m}$ is the position of item m when user n observes the ranking. Thus, $r_{n,m} = 1$ if website m appears in the top of the ranking and $r_{n,m} = |M|$ if it appears in the last position.

Using this model, the researchers showed that the disparate visibility of minorities in rankings can be explained by the preference distribution among users, the classes sizes, and the bias to click on popular items. The most surprising phenomenon, which they coined as *few-get-richer*, is that reducing the number of websites of a class can lead to increasing the overall number of clicks of the class. Simplistically, this phenomenon can be explained as follows: if there exists a group of users which always click on websites of a particular class, and the number of websites of that class is reduced, then each website of this class receives more clicks. Because they receive more clicks, they appear in higher positions in the ranking. Because websites in the top positions of the ranking have larger visibility, they can attract more clicks from indifferent users. This phenomenon clearly leads to the over-representation in high-ranked positions of items of the minority class.

According to the above explanation, this phenomenon only can appear if there are neutral users. Actually, the larger the number of users without preference for any class, the larger can be the amplification of the disparate visibility of the group. The bias to click on popular websites (preferential attachment) is also necessary to generate this phenomenon.

1.1.2 Influence of Homophily in the Ranking of Minorities in Social Networks

The second work which inspired this TFG is [Karimi et al., 2018], in which the visibility of minority groups in social networks is studied. The network formation model used in [Karimi et al., 2018] is a variant of the well-known Barabási-Albert model ([Barabási and Pósfai, 2016, Chapter 5]). Thus, this model uses the preferential attachment mechanism (PA): the more popular nodes have a larger probability of increasing its popularity. In this variant, nodes have been partitioned into two classes, and the homophily mechanism regulates how frequent contacts between similar and dissimilar nodes, i.e., nodes of the same or different classes, occur ([Mcpherson et al., 2001]; [Avin et al., 2015]). The homophily level of the network can be adjusted by parameter $h \in [0, 1]$, so that when $h = 0$ nodes only connect with nodes of the other class (completely heterophilic networks) and when $h = 1$ nodes only connect with other nodes of the same class (completely homophilic networks). Because of the presence of PA and the homophily mechanism, we will name this model as PAH.

Starting from an initial graph, at each time step, a node j is added to the graph and attached to m different nodes select at random. The probability of the node j to connect to the pre-existing node i is

$$\Pi_{i,j} = \frac{h_{i,j} k_i}{\sum_l h_{l,j} k_l} \tag{1.3}$$

where k_i is the degree of node i and $h_{i,j}$ is the homophily level between nodes i and j. Thus, if node i and j belong to the same class, then $h_{i,j} = h$. Otherwise, $h_{i,j} = 1 - h$.

The experiments in [Karimi et al., 2018] showed that the minority class is on average over-represented in the heterophilic regime, in the sense that the sum of the degree of the nodes of the minority class is on average larger than when edges are created at random. However, the majority class is over-represented in the homophilic regime, except in the completely homophilic scenario, in which the minority class is adequately represented due to what the researchers named full in-group support.

Interestingly, the *few-get-richer* effect is also present in this model: the fewer the items reporting a given signal, the higher the relative number of attached nodes. In other words, the lower the number of nodes in the minority class, the larger the average degree of these nodes[1].

1.2 Objectives

Both [Germano et al., 2019] and [Karimi et al., 2018] describe models that explain the dynamic popularity of items. Moreover, the mechanisms of these models are very similar: the larger the popularity of an item, the more probability to increase its popularity, and the larger the affinity (between a user and a website in [Germano et al., 2019] and between nodes in [Karimi et al., 2018]), the larger the probability of creating a connection. Is it possible that the disparate visibility in these two models are consequences of the same mechanism?

Answering this question is the first objective of this TFG, and this question will be addressed in chapter 2. To do so, we will test whether PA is indispensable in PAH to generate the disparate visibility of groups. If it is not the case, then we have evidence that the phenomena observed in [Germano et al., 2019] and [Karimi et al., 2018] are not consequences of the same mechanism. This is because PA actually is necessary to observe this phenomenon in [Germano et al., 2019].

The second objective consists in studying whether, in a model which does not depend on the homophily mechanism, the disparate visibility of groups can also occur. This question will be addressed in chapter 3.

Finally, the third objective consists in studying whether the *few-get-richer* effect can also occur in the absence of the homophily mechanism. This question will be also addressed in chapter 3.

[1]Notice that the definition of *few-get-richer* effect for rankings is slightly different than for networks

2

DISPARATE VISIBILITY IN THE ABSENCE OF PREFERENTIAL ATTACHMENT

In this section, we will show that PA is not necessary in PAH to observe the disparate visibility of groups, but it amplifies this phenomenon. Thus, because PA is necessary in [Germano et al., 2019], the disparate visibility observed in [Karimi et al., 2018] and [Germano et al., 2019] are not consequences of the same mechanism.

To do so, we define a network formation model which, in essence, only differs with PAH in the absence of PA.

The implementation of this model and the code to run experiments can be found in the repository of this project: https://github.com/danielroncel/tfg.

2.1 Network Formation Model Definition

The initial edgeless graph will be formed by n nodes. Nodes have been partitioned into two classes, so that a ratio C_0 of them belong to the majority class, and a ratio $C_1 = 1 - C_0$ belong to the minority class. The homophily mechanism regulates the frequency of contacts between nodes of the same and different classes. The homophily level can be tuned with parameter $h \in [0, 1]$ as described in section 1.1.2.

One node at a time, they choose m different nodes to connect with. The probability of the node j to choose node i to connect with is

$$\Pi_{i,j} = \frac{h_{i,j}}{\sum_l h_{l,j}} \tag{2.1}$$

where $h_{i,j}$ is the homophily level between nodes i and j. Notice that the unique difference between equations 1.3 and 2.1 is that in the second one the probability of attracting an edge is not proportional to the degree of the target node. Thus, PA is not included in our model.

2.2 Experiments and Conclusions

Our results are based on the effect of C_1 and h in the representation of the minority class. The experiment is the same that the one performed in [Karimi et al., 2018] but applied to our model. The results are obtained by comparing the output of the same experiment in the two different models. In all the executions of our model, the results have been computed using parameters $m = 5$, $n = 5000$ and averaged over 20 simulations.

In this experiment, we will be mainly interested in the overall degree of the classes. Let S_0 and S_1 be the sets of nodes of the majority and minority classes, respectively. We define the degree of the majority and minority classes as $K_0 = \sum_{i \in S_0} k_i$ and $K_1 = \sum_{i \in S_1} k_i$, respectively. We also define $K = K_0 + K_1$ as the sum of the degree of all nodes in the graph.

In the absence of the homophily mechanism, the degree of a class is expected to be proportional to the size of the group, because classes would be indistinguishable. Mathematically, in this scenario, it must hold that on average $K_0/K = C_0$ and $K_1/K = C_1$. If $K_1/K > C_1$, we will say that the minority class is over-represented. Analogously, if $K_1/K < C_1$, we will say that it is under-represented.

2.2.1 Results

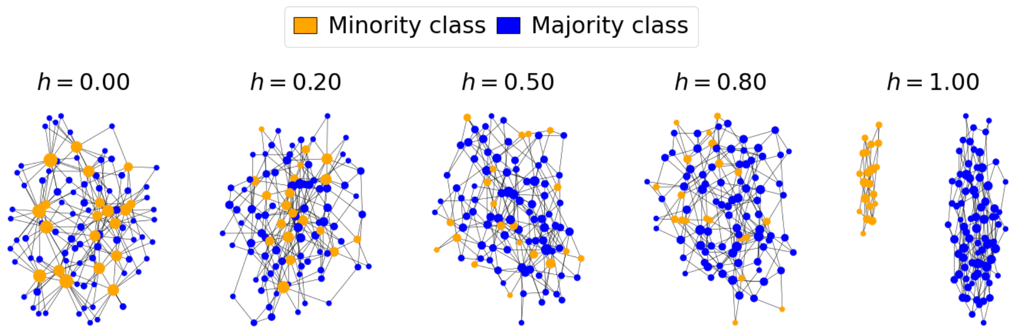
In figure 2.1 we can see the effect of parameter h in the network topology. For $h = 0$ there only exist connections between nodes of different classes, and for $h = 1$ nodes only connect with other nodes of the same class. The spectrum of intermediate situations is generated by the rest of values of h.

We will focus on the results of the experiment shown in figure 2.2. About the top plot, with respect to parameter h, the following observations are valid in our model and in the analogous experiment in [Karimi et al., 2018]:

- In an extreme heterophilic network ($h = 0$), the minority class nodes (formed by the 20% of nodes) are receiving all the edges with source in the majority class (formed by the 80% of the nodes) and vice versa, which leads to the over-representation of the minority class.

- As h increases, the minority class degree decreases non-linearly. For $h = 0.5$, the degree of both classes is on average proportional to the class size, so

Figure 2.1: Schematic of the network topology



Examples of networks for several values of h. In all scenarios $n = 100$, $m = 2$, $C_0 = 0.8$ and $C_1 = 0.2$. The size of a node is proportional to its degree. The darker lines correspond to duplicated edges.

there is no over-representation. This is the expected result since, according to equation 2.1, all nodes, even those of different classes, have the same probability of attracting and edge. Thus, in practice, classes are ignored.

- As h keeps increasing, the minority class is on average under-represented. The minority class degree keeps decreasing until $h \approx 0.8$. Thus, in the homophilic regime, the minority class nodes have difficulties even to attract the edges generated by nodes of the same class.

- For larger values of h, the minority class degree increases until $h = 1$, scenario in which the degree of both classes is again proportional to the class size, so there is no over-representation.

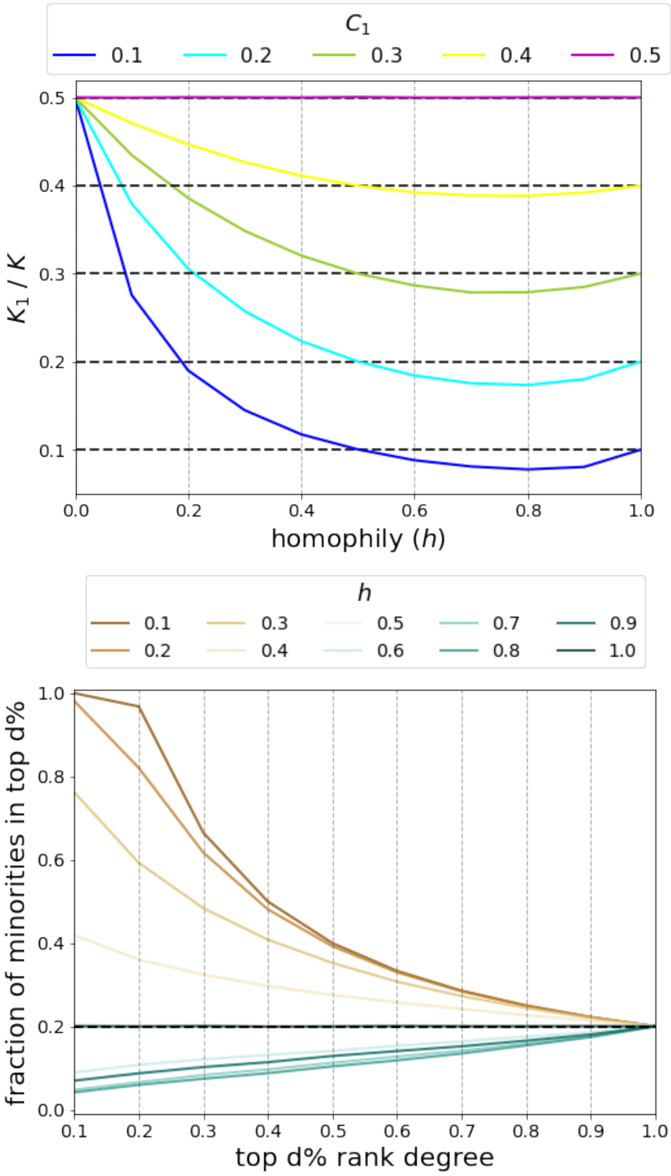
With respect to the class sizes, the larger C_1, the larger is the degree of the minority class. The same phenomenon was observed in PAH.

Since the over-representation of classes also occurs in our model, we have shown that PA is not necessary to generate disparate visibility. The main difference between the results in our model and in PAH is the amplification of the disparate visibility: in all the scenarios with over-representation of the minority class, its average degree is larger in PAH than in our model. Analogously, in all the scenarios with under-representation of the minority class, its average degree is smaller in PAH. Therefore, PA is not necessary to generate this effect, but amplifies it[1].

It is interesting to notice that the effect of PA is stronger in the heterophilic regime, giving a high advantage to the minority class. For example, for $h = 0.2$

[1]We highly encourage reading the results of the analogous experiment in [Karimi et al., 2018] to validate these observations.

Figure 2.2: Representation of the minority class



Representation of the minority class for several class sizes and homophily levels. First plot: degree of the minority class computed for various combinations of values of h and C_1. The minority class is over-represented if K_1/K is larger than C_1. If it is lower than C_1, it is under-represented. Second plot: fraction of nodes of the minority class in the top d% of nodes with largest degree for $C_1 = 0.2$ and various values of h. The minority class is over-represented if its score is larger than $C_1 = 0.2$. If it is lower than C_1, it is under-represented.

and $C_1 = 0.2$, with PA $K_1/K \approx 0.4$, but without PA $K_1/K \approx 0.3$. Thus, PA can significantly increase the degree of minorities. However, in the homophilic regime, the difference of the minority class degree between both models is clearly much smaller. Thus, the effect of PA is smaller in the homophilic regime.

With respect to the second plot in 2.2, the results are again very similar to the ones of PAH: For $h < 0.5$ the minority class is over-represented, for $h > 0.5$ it is under-represented, and for $h = 0.5$ and $h = 1$ its representation is proportional to its size. Again, the main difference compared with PAH is that in PAH this effect has been amplified.

All in all, we have concluded that PA is amplifying the disparate visibility phenomenon in PAH, especially in the scenarios favourable to the minority class. However, PA is not necessary to generate this phenomenon. Thus, the phenomena observed in [Germano et al., 2019] and [Karimi et al., 2018] are not consequences of the same mechanism.

3

DISPARATE VISIBILITY IN THE ABSENCE OF THE HOMOPHILY MECHANISM

In this section we will study whether the disparate visibility and the *few-get-richer* effect can occur in networks which do not depend on the homophily mechanism. To do so, we define a network formation model inspired by [Germano et al., 2019] and we will study the disparate visibility of groups in the networks generated by this model. Analogously to [Germano et al., 2019], in our model, nodes are partitioned into two classes of different sizes and they have heterogeneous preferences. Our preference mechanism differs from the homophily mechanism in the sense that the preference of a node is independent of its class. Because this model is ruled by PA and the preference mechanism, we will name this model as PAP.

The implementation of the model and the code to run experiments can be found in https://github.com/danielroncel/tfg.

3.1 Network Formation Model Definition

The initial graph will be formed by m_0 nodes, and it will grow until containing n nodes. Nodes are partitioned into two classes: a ratio C_0 of them belong to the majority class, and a ratio $C_1 = 1 - C_0$ belong to the minority class.

One node at a time, they are added to the graph and they connect to m different pre-existing nodes that satisfy the preference of the new node. Nodes are divided in three groups according to their preference: those that, when added to the graph, only connect with nodes of the majority class (only nodes of the majority class satisfy their preference), those that only connect with nodes of the minority class, and those that are indifferent to the class of the target node, and thus all the pre-existing nodes satisfy their preference. There is a ratio P_0 of nodes with a prior preference for the majority class. From all these nodes with prior preference for the majority class, at the end a ratio $P_{0,i}$ of them will be indifferent, and the remaining ratio $1 - P_{0,i}$ of nodes will keep their preference for the majority group. Mathematically, if Q_0 is the set of nodes with prior preference for the majority class, and $Q_{0,i}$ is the subset of them that at the end are indifferent, then

$$P_{0,i} = \frac{|Q_{0,i}|}{|Q_0|} \tag{3.1}$$

Analogously, we define $P_1 = 1 - P_0$ as the ratio of nodes with prior preference for the minority class, and $P_{1,i}$ as the ratio of them that at the end has no preference for any class.[1]. The probability of the new node j to connect to the pre-existing node i is

$$\Pi_{i,j} = \frac{\beta^{k_i} g_{i,j}}{\sum_l \beta^{k_l} g_{l,j}} \tag{3.2}$$

where $\beta > 1$ is a constant, k_i is the degree of node i and $g_{i,j}$ indicates whether node i satisfy the preference of node j. More specifically, $g_{i,j} = 1$ if node i satisfy the preference of j. Otherwise, $g_{i,j} = 0$.

As it can be deduced from equation 3.2, popular nodes have a larger probability of attracting new edges. In other words, our model is also ruled by PA.

Notice that when node j is added to the graph, only the preference of node j is considered; the preferences of the pre-existing nodes are ignored.

3.2 Experiments and Conclusions

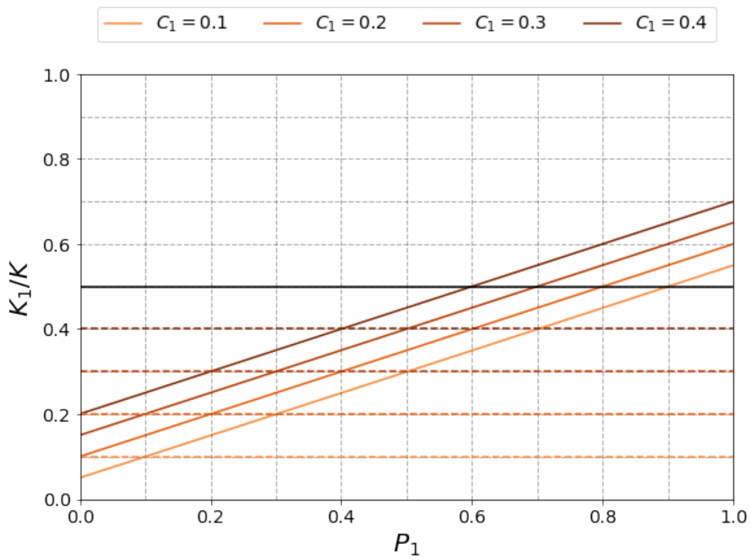
In this section, we will study how the parameters defined in section 3.1 affect the average degree of each class, its degree distribution and the network topology. The degree of the majority and minority classes are defined as in section 2.2, denoted as K_0 and K_1, respectively.

As in the previous chapter, the minority class is over-represented if $K_1/K > C_1$, under-represented if $K_1/K < C_1$, and its degree is proportional to its size if $K_1/K = C_1$.

In all the experiments, $n = 5000$, $m_0 = 10$ and $\beta = 1.15$. The results have been averaged for 20 executions.

[1] For a better understanding of parameters $P_{0,i}$ and $P_{1,i}$, consider reading sections A.1, A.2 and A.3.

Figure 3.1: Effect of the preference distribution and the size of the classes on the minority class degree when $P_{0,i} = P_{1,i} = 0$



Degree of the minority class for several values of P_1, and C_1. In all cases, $P_{0,i}$ = $P_{1,i} = 0$ and $m = 5$. The black line indicates the point at which $K_0 = K_1$, i.e., when both groups have the same average degree. The orange dashed lines indicate the degree of the minority class in the scenario in which its degree is proportional to its size.

3.2.1 Analysis without indifferent nodes

In this section we will study the representation of classes when there are no indifferent nodes. Since there are no indifferent nodes, $P_{0,i} = P_{1,i} = 0$.

3.2.1.1 Effect of the Preference Distribution

In figure 3.1 we can see how the preference distribution and the classes sizes affect the minority class degree. We can see that when $P_1 > C_1$ the minority class is over-represented, i.e., $K_1/K > C_1$. This is the expected phenomenon, since the minority class is attracting a larger fraction of the edges than the fraction of nodes that it contains. Analogously, the minority class is under-represented when $P_1 < C_1$. When $P_1 = C_1$ (and therefore $P_0 = C_0$), the degree of each class is proportional to the number of nodes of the class, i.e., $K_0/K = C_0$ and $K_1/K = C_1$. Thus, if $C_1 = P_1$, then no class is over-represented. It is also interesting to notice that if $P_1 = C_0$ and $P_0 = C_1$, then $K_0 = K_1$. Despite the edges in undirected graphs have no directionality, we can understand this phenomenon by thinking that the in-degree of the minority class (the number of edges attracted by the minority class nodes) equals the out-degree of the majority class (the number of edges generated by the majority class nodes) and vice versa. Therefore, both classes must have the same degree.

3.2.1.2 Effect of the Size of the Classes

In figure 3.2 we explore with more detail a particular case of figure 3.1.

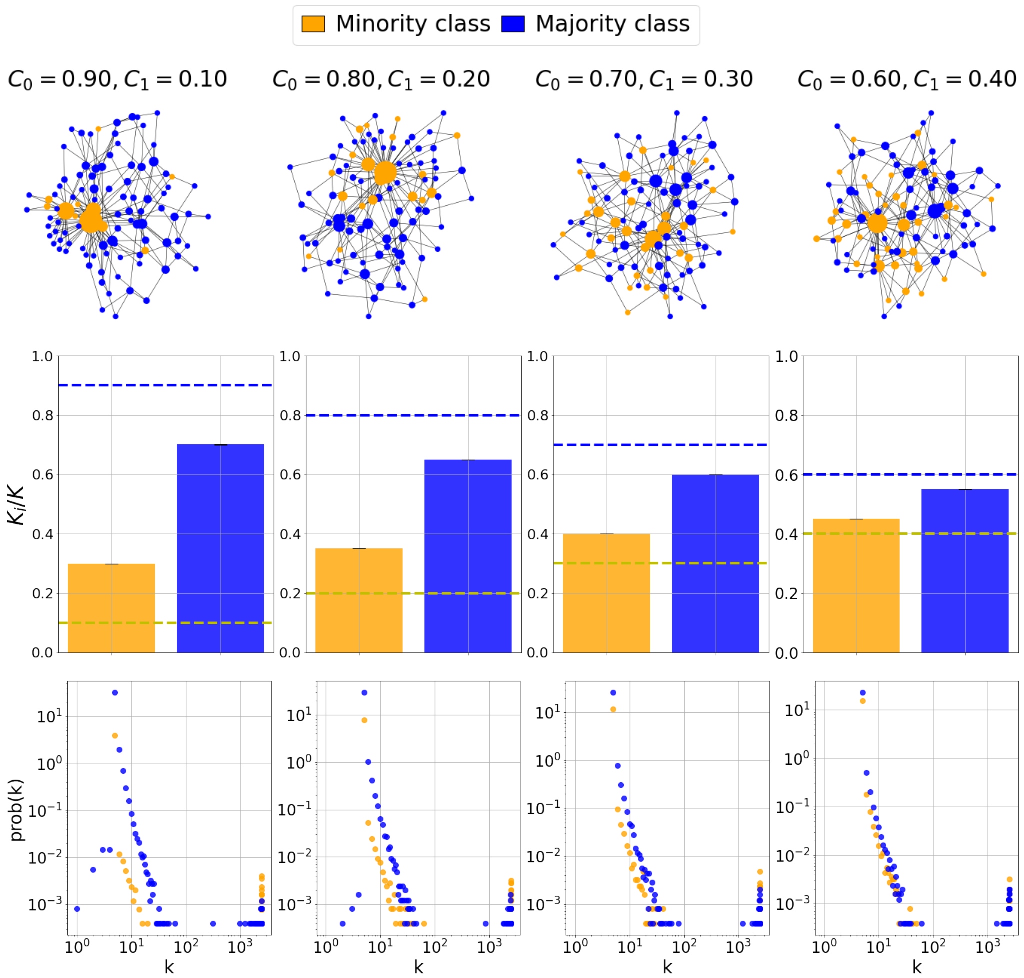
As we can see in both figures, the degree of the minorities decreases monotonously as C_1 decreases. This is because the minority class nodes always attract a ratio P_1 of the edges, and therefore the indegree of the class remains constant when C_1 is modified. Since the outdegree decreases as C_1 decreases, a class with fewer nodes must have a lower degree.

Also, in the average degree of classes in 3.2, it is interesting to notice that the lower the number of nodes in the minority class, the larger is the amplification of its over-representation. In other words, the lower the number of nodes in the minority class, the larger is the advantage that the minority class obtains with respect to the scenario with no over-representation (which correspond to the orange dashed lines in the figure). This phenomenon can be explained as follows: given a fixed number of nodes with preference for the minority class, if the number of nodes in the minority class is reduced, then the same number of edges will be distributed among fewer nodes, and therefore, the average degree of these nodes will also increase. This explanation is reinforced by the degree distribution in figure 3.2: the lower C_1, the larger seems to be the average degree of the nodes of the minority class, since its degree distribution is skewed to larger values. Clearly, this phenomenon corresponds to the *few-get-richer* effect as defined in section 1.1.2.

3.2.1.3 The Degree Distribution

The degree distribution shown in figure 3.2 gives interesting insights about this model. Notice that the degree distribution seems to deviate from a pure power law ([Barabási and Pósfai, 2016, Chapter 4]). This is because the probability of attracting an edge is not proportional to the node degree as in the Barabási-Albert model but increases exponentially with respect to its degree. Thus, there are a few hubs, i.e, very popular nodes, which are attracting almost all the edges generated by new nodes. This phenomenon, named *winner-takes-all*, suggests that this model leads to *hub-and-spoke* networks, in which the first nodes added to the graph receive a huge advantage and can attract almost all the edges generated by new nodes ([Barabási and Pósfai, 2016, Chapter 5]).

Figure 3.2: Representation of the classes for several group sizes when $P_{0,i} = P_{1,i} = 0$



Representation of the classes for several group sizes. For all these experiments, $P_0 = P_1 = 0.5$ and $P_{0,i} = P_{1,i} = 0$. The classes sizes are defined at the top of the columns. The top row corresponds to the schematic of the network topology for $n = 100$ and $m = 2$, where the size of each node is proportional to its degree. The second row show the average degree of the classes for $m = 5000$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value. The third row show the degree distribution of the classes for $m = 5000$ and $m = 5$.

3.2.2 Analysis with indifferent nodes

To understand better how indifferent nodes amplify the over-representation of a class, in all the experiments of this section we will set $C_0 = P_0$ and $C_1 = P_1$. This is because, according to section 3.2.1, without indifferent nodes, if $C_0 = P_0$ and $C_1 = P_1$, then no class is over-represented. Therefore, adding indifferent nodes to this setting we can understand better their role in the disparate visibility effect.

3.2.2.1 Analysis for $P_{0,i} = P_{1,i}$

In this section, we will consider the experiments in which $P_{0,i} = P_{1,i}$. The results of these experiments can be seen in figure 3.3.

First, we will consider the scenario in which $P_{0,i} = P_{1,i} = 1$. In other words, all nodes are indifferent. For this setting we can see that no class was clearly over-represented. This is because, given that all nodes are indifferent, classes are irrelevant, and nodes of the minority and majority classes must behave indistinguishably (the examples of networks generated for this setting in figure 3.4 motivate the idea). Therefore, the degree of each class is expected to be proportional to the number of nodes in the class.

For $P_{0,i} = P_{1,i} < 1$, we can also see that no class in on average clearly over-represented. A reasonable explanation is that it is because the number of nodes with preference for each class is proportional to the size of each class. To express this idea mathematically, let's divide the number of nodes with preference for the majority class by the number of nodes of the majority class:

$$\frac{n \cdot P_0 \cdot (1 - P_{0,i})}{n \cdot C_0} = \frac{n \cdot C_0 \cdot (1 - P_{0,i})}{n \cdot C_0} = 1 - P_{0,i} \tag{3.3}$$

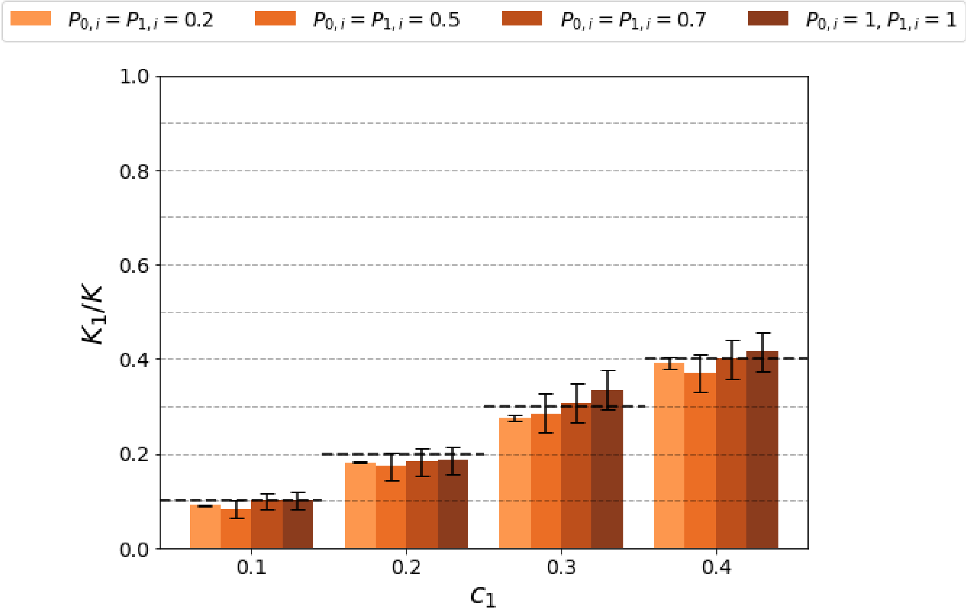
Now, let's divide the number of nodes with preference for the minority class by the number of nodes of the minority class:

$$\frac{n \cdot P_1 \cdot (1 - P_{1,i})}{n \cdot C_1} = \frac{n \cdot C_1 \cdot (1 - P_{1,i})}{n \cdot C_1} = 1 - P_{1,i} \tag{3.4}$$

As a recap, the above equations are valid because we assumed that $C_0 = P_0$ and $C_1 = P_1$. Since we also assumed that $P_{0,i} = P_{1,i}$, then equations 3.3 and 3.4 are the same. Thus, ignoring the edges generated by indifferent nodes, the average degree of nodes of both classes are the same. Consequently, the nodes of both classes are equally competitive to attract the edges generated by indifferent nodes. Taking all this into account, it is coherent to think that no class should be on average clearly over-represented.

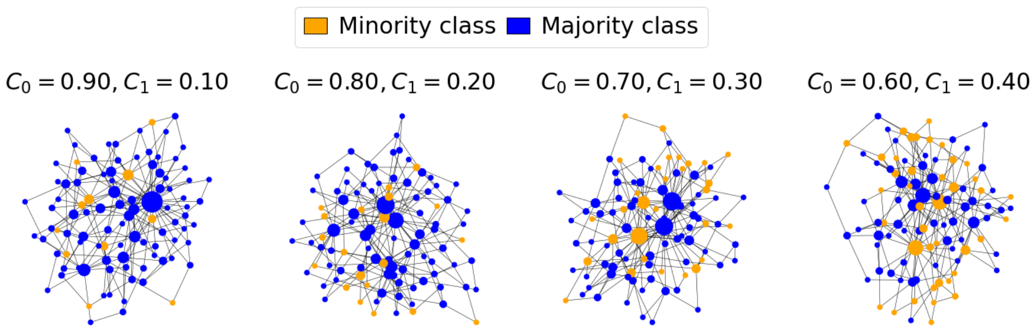
All in all, the evidence suggests that if $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i}$, no class is expected to be clearly over-represented.

Figure 3.3: Minority class degree for $P_0 = C_0$, $P_1 = C_1$ when $P_{0,i} = P_{1,i}$



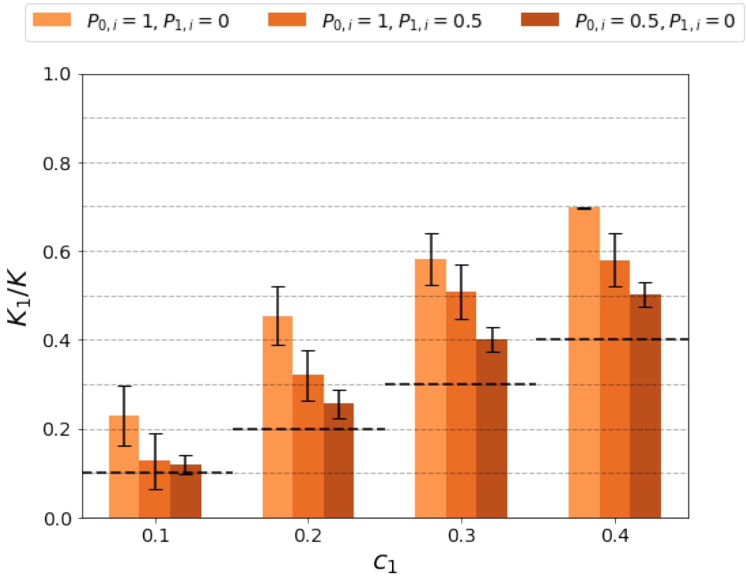
Degree of the minority class for several values of P_0, P_1, $P_{0,i}$ and $P_{1,i}$. In all cases, $C_0 = P_0$, $C_1 = P_1$ and $m = 5$. Each black dashed line indicates which would be the degree of the minority class if its representation were proportional to its size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value. The normalized degree of the majority class can be computed as 1 - K_1/K.

Figure 3.4: Schematic of the network topology for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 1$



Examples of networks generated for several values of C_0 and C_1 when all nodes are indifferent to the class of the other nodes. In all scenarios, $n = 100$, $m = 2$, $m_0 = 5$, $P_0 = C_0$, $P_1 = C_0$ and $P_{0,i} = P_{1,i} = 1$

Figure 3.5: Minority class degree for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} > P_{1,i}$



Degree of the minority class for several values of P_0, P_1, $P_{0,i}$ and $P_{1,i}$. Each black dashed line indicates which would be the degree of the minority class if its representation were proportional to its size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value. In all cases, $P_0 = C_0$, P_1=C_1 and $m = 5$.

3.2.2.2 Analysis for $P_{0,i} > P_{1,i}$

The scenarios explored in this subsection are those in which $P_{0,i} > P_{1,i}$. The result of the experiments for this setting appear in figure 3.5.

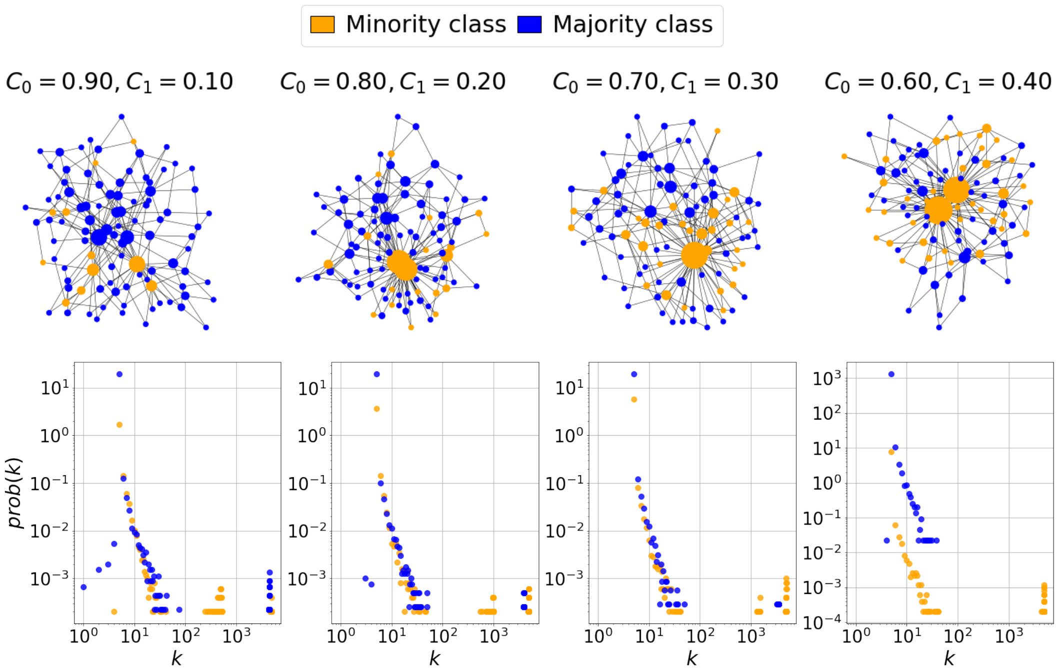
As we can see, the minority class has been on average clearly over-represented in all the scenarios of figure 3.5. The main result is that, the larger $P_{0,i}$ and the lower $P_{1,i}$, the larger is the average degree of the minority class. This phenomenon can also be explained by equations 3.3 and 3.4: since $P_{0,i} > P_{1,i}$, then the minority class nodes receive on average more edges of nodes with a fixed preference, which also makes them more competitive to attract the edges generated by indifferent nodes.

Now we will analyse how much the parameter m influences the degree of the minority class. To do so, we define the relative over-representation of the minority class as

$$\Delta_1 = \frac{avg(\frac{K_1}{K}) - C_1}{avg(\frac{K_1}{K})} \tag{3.5}$$

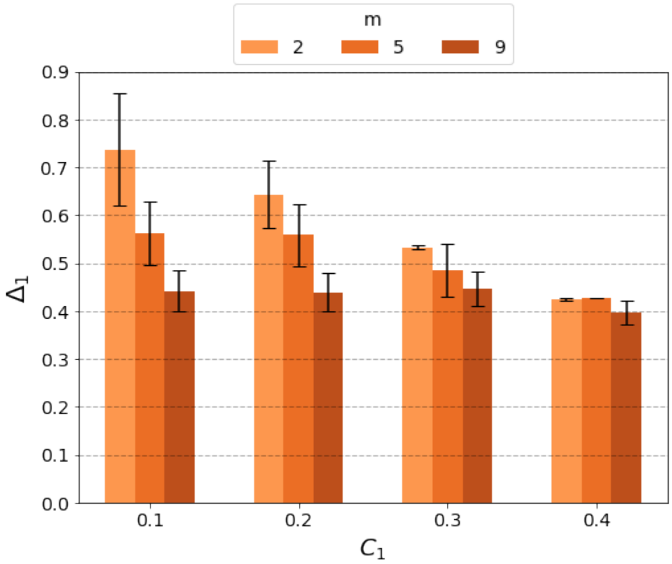
so that $avg(\frac{K_1}{K})$ is the average across all executions of the degree of the minority class normalized by the sum of the degree of all nodes. Intuitively, we are measuring

Figure 3.6: Representation of classes for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$



Representation of the classes for several group sizes. For all these results, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$. The classes sizes are defined at the top of the columns. The top row corresponds to the schematic of the network topology for $n = 100$ and $m = 2$, and the size of each node is proportional to its degree. The third row show the degree distribution of the classes for $m = 5000$ and $m = 5$.

Figure 3.7: Relative over-representation of the minority class for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$



Relative over-representation of the minority class for several values of C_1, P_1 and m. In all cases, $P_{0,i} = 1$, $P_{1,i} = 0$, $C_0 = P_0$ and $C_1 = P_1$. The black vertical line corresponds to the standard deviation centred in the mean value.

which ratio of the degree of the minority class is over the wider dashed line in figure 3.5.

As we can see in figure 3.7, Δ_1 decreases as m increases. This phenomenon can be explained as follows: taking into account that nodes of the minority class are more competitive to attract new edges, if $m = 1$, then most of the edges generated by indifferent nodes will by attracted by a hub of the minority class. For larger values of m, nodes of the majority class have more chances to attract the remaining edges, reducing the advantage of the minority class.

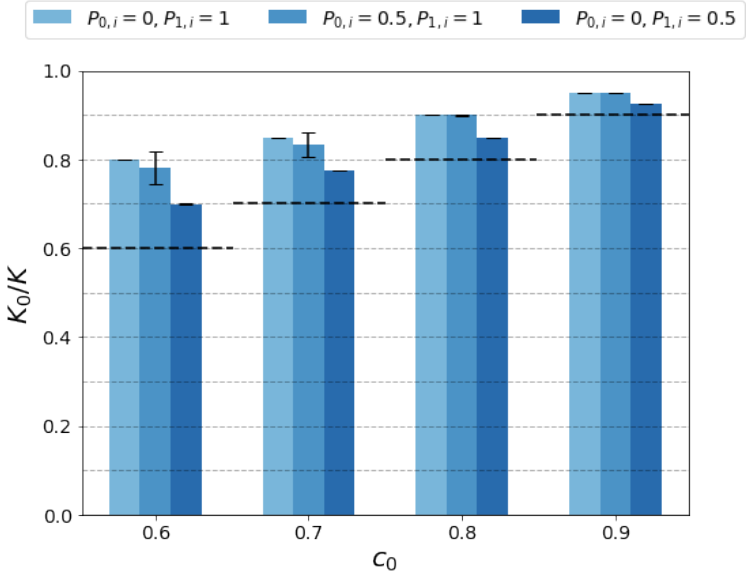
3.2.2.3 Analysis for $P_{0,i} < P_{1,i}$

Now we will explore the scenarios in which $P_{0,i} < P_{1,i}$. The result of the experiments for this setting appear in figure 3.8.

Analogously to the previous section, the larger $P_{1,i}$ and the lower $P_{0,i}$, the larger is the average degree of the majority class. Again, this is because, according to equations 3.3 and 3.4, the nodes of the majority class are receiving on average more edges of nodes with a fixed preference, which it also makes them more competitive to attract the edges generated by indifferent nodes.

Notice that the variance of the degree of the classes in figure 3.8 has decreased compared with figure 3.5. A possible explanation of this phenomenon is as follows:

Figure 3.8: Majority class degree for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} < P_{1,i}$

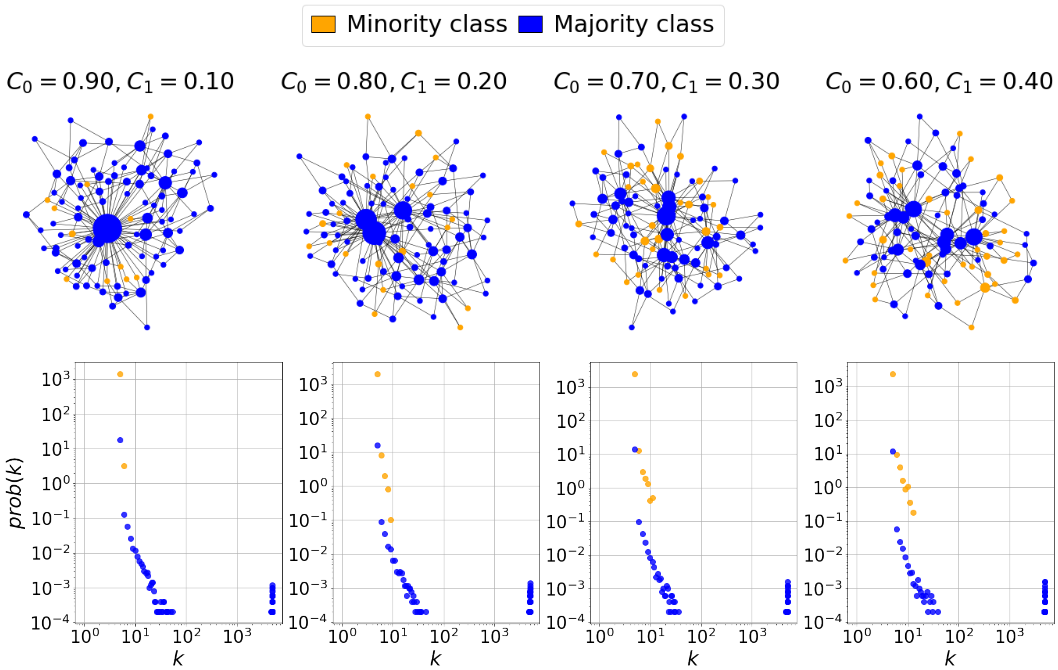


Degree of the majority class for several values of $P_{0,i}$, $P_{1,i}$, P_0 and P_1. Each black dashed line indicates which would be the degree of the majority class if its representation were proportional to its size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value. In all cases, $P_0 = C_0$, P_1=C_1 and $m = 5$.

in the scenarios in which $P_{0,i} = 1$ and $P_{1,i} = 0$ there are more indifferent nodes than in the scenarios in which $P_{0,i} = 0$, $P_{1,i} = 1$, which is easy to see given that the majority class contains more nodes. For example, keeping in mind that $C_0 = P_0$ and $C_1 = P_1$, if $C_0 = 0.9$, $C_1 = 0.1$, $P_{0,i} = 1$ and $P_{1,i} = 0$, the 90% of the nodes will have no preference assigned. However, if $P_{0,i} = 0$ and $P_{1,i} = 1$, only the 10% of nodes have no preference assigned. Therefore, it is not surprising that the variance of the degree of the classes is larger in the first scenario. Although this explanation may fit in the scenario in which C_1 is very small, it may not be plausible for larger values of C_1 such as $C_1 = 0.4$.

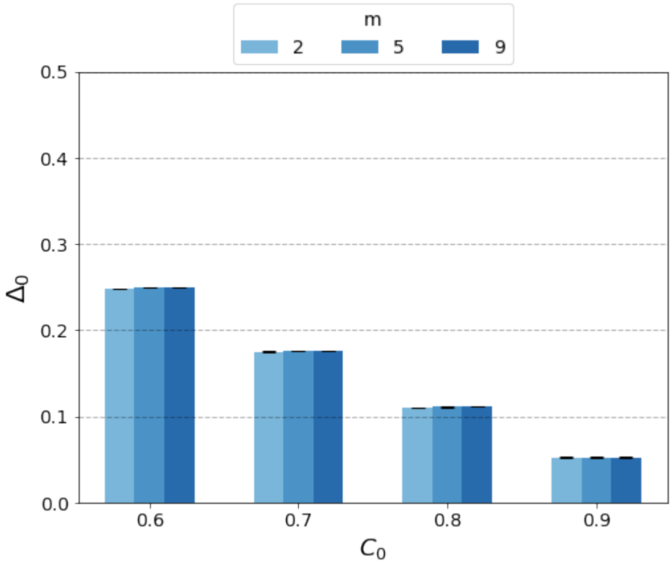
Another possible explanation can be deduced comparing the degree distributions shown in figures 3.6 and 3.9. On the one hand, in the most favourable scenario for the minority class (figure 3.6), both classes can generate very popular nodes. Thus, despite the minority advantage, both classes have hubs and can compete for attracting edges of indifferent nodes, increasing the variance of the final degree. On the other hand, in the most favourable scenario for the majority class (figure 3.9), the minority class cannot generate hubs, so all the high-degree nodes belong to the majority class. Thus, the minority class nodes cannot compete for attracting the edges generated by indifferent nodes. Since these edges are nearly always attracted by nodes of the majority class, the class degree variance must be smaller.

Figure 3.9: Representation of classes for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$



Representation of the classes for several group sizes. For all these results, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$. The classes sizes are defined at the top of the columns. The top row corresponds to the schematic of the network topology for $n = 100$ and $m = 2$, and the size of each node is proportional to its degree. The third row show the degree distribution of the classes for $m = 5000$ and $m = 5$.

Figure 3.10: Relative over-representation of the majority class for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$



Relative over-representation of the majority class for several values of C_0, P_0 and m. In all cases, $P_{0,i} = 0$, $P_{1,i} = 1$, $C_0 = P_0$ and $C_1 = P_1$. The black vertical line corresponds to the standard deviation centred in the mean value.

Now we will see how parameter m influences the degree of the majority class (figure 3.10). To do so, we define the relative over-representation of the majority class as

$$\Delta_0 = \frac{avg(\frac{K_0}{K}) - C_0}{avg(\frac{K_0}{K})} \tag{3.6}$$

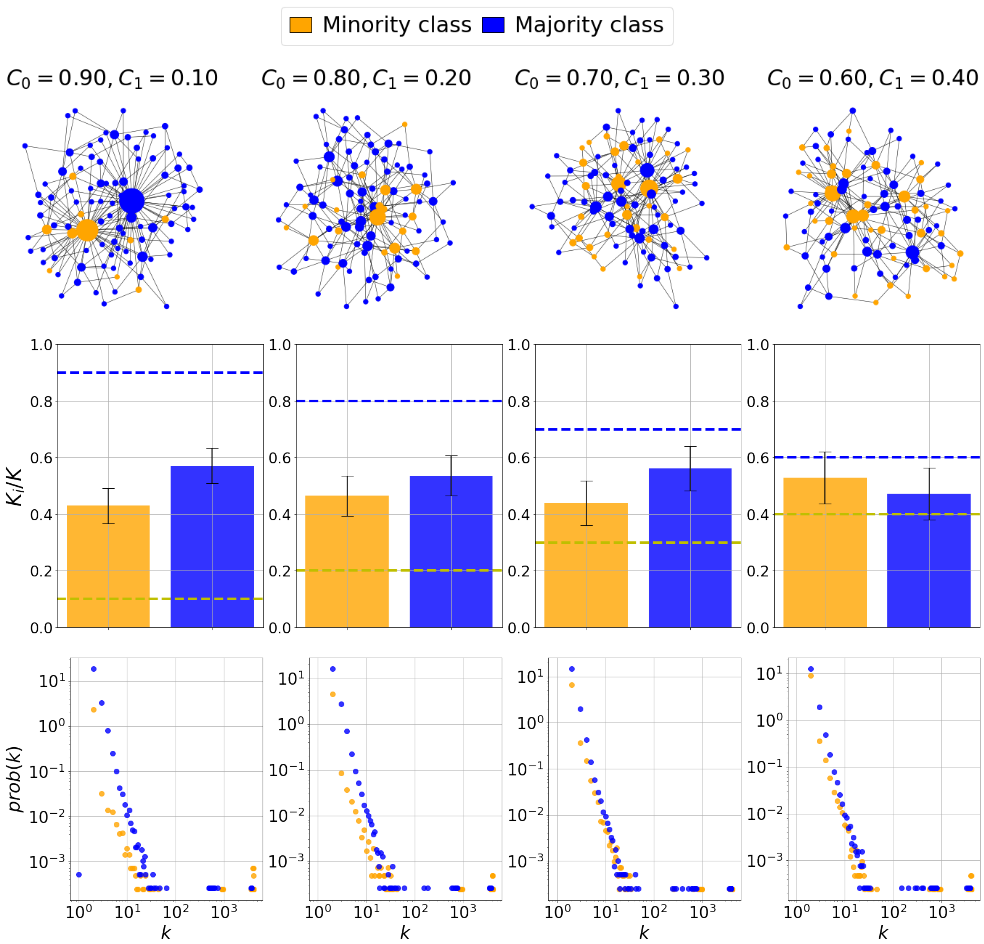
so that $avg(\frac{K_0}{K})$ is the average degree of all executions of the majority class normalized by the sum of the degree of all nodes.

It is interesting to notice that we did not get evidence that the majority class is given advantage by reducing the value of m. Again, it can be explained by the absence of hubs of the minority class (figure 3.9): since there are no hubs of the minority class, they cannot compete for attracting the edges of indifferent nodes even for larger values of m.

3.2.2.4 The *Few-get-richer* Effect

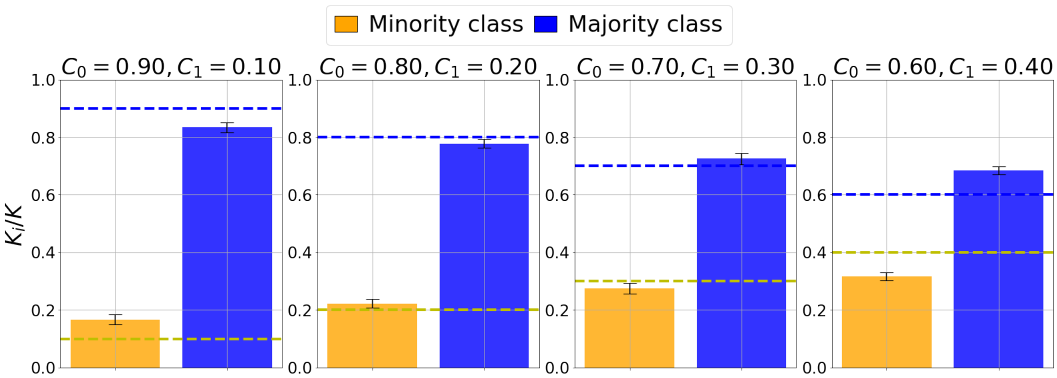
As we have seen in section 3.2.1.2, the *few-get-richer* effect can appear using this model. To obtain better insights about this effect, we have executed several experiments. The results of some of them can be seen in appendix A.7. However,

Figure 3.11: The *few-get-richer* effect when $m = 2$, $P_0 = 0.8$, $P_{0,i} = 0.8$ and $P_{1,i} = 0$



Representation of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0.8$, $P_{1,i} = 0$, $P_0 = 0.8$, $P_1 = 0.2$ and $m = 2$. First plot: schematic of the network topology. Second plot: Average degree of classes. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value. Third plot: Degree distribution of classes.

Figure 3.12: The *Few-get-richer* effect when $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 0.6$



Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0$, $P_{1,i} = 0.6$, $P_0 = 0.5$, $P_1 = 0.5$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

we will only focus on two of them.

Among all the experiments done, the one in which the *few-get-richer* effect is stronger is shown in figure 3.11. This is because when the minority class size decreases, its degree remains quite stable. Actually, the average degree of the minority class is larger for $C_0 = 0.2$ than for $C_1 = 0.3$. This phenomenon can be explained as follows: ignoring the links generated by indifferent nodes, for a fixed preference distribution, if the number of nodes in a class is reduced, the average degree of these nodes increases. Then, the nodes of this class become more competitive to attract the edges generated by indifferent nodes. This reasoning is reinforced by the degree distribution in the same figure: the lower C_1, the number of nodes with few connections in the minority class is reduced, and the number of hubs increases. This idea is also motivated by the examples of networks in figure 3.11: the more popular hubs of the minority class are more frequent for lower values of C_1.

Another interesting experiment is the one shown in figure 3.12. The fewer the nodes of a minority class, the higher the relative share of the overall traffic they collectively attract, i.e., the higher the relative number of attached nodes. We can observe that reducing the minority from a proportion of $C_1 = 40\%$, where the amount of traffic is $\approx 30\%$ (10% less), to a proportion of $C_1 = 10\%$, where the amount of traffic is $\approx 15\%$ (5% more), results in a relative increase of around $5 - (-10) = 15\%$.

Notice that in the experiment corresponding to figure 3.11, when $C_1 = 0.4$ the minority class is already over-represented. However, in the experiment corresponding to figure 3.12, when $C_1 = 0.4$ the minority class is under-represented. Thus, this phenomenon can appear in scenarios already beneficial for the minority class and also can help minorities in scenarios of under-representation.

As a final comment, when the number of nodes in a class is reduced and the average degree of these nodes increases, it may be also increasing the centrality of the nodes. In other words, reducing the size of the class may be increasing their influence, for example, to propagate the ideas of these nodes ([Zhang and Luo, 2017]). However, this analysis is beyond the scope of this TFG.

4

CONCLUSIONS

In chapter 2 we defined a network formation model similar to PAH but removing PA. Thus, the homophily mechanism completely dominates the dynamics of the networks. Our conclusions are:

- First, in PAH, PA amplifies the disparate visibility of the advantaged group (the advantaged group in the homophilic regime is the majority class, and in the heterophilic regime it is the minority class). However, PA is not necessary to observe this phenomenon.
- Second, in PAH, PA is mainly benefiting the minority class in the heterophilic regime. This is because the amplification of the disparate visibility is larger in this regime.
- Third, given that the disparate visibility can be observed in the absence of PA, the phenomena observed in [Germano et al., 2019] and [Karimi et al., 2018] are not consequences of the same mechanism. This is because PA is necessary in [Germano et al., 2019] to generate the disparate visibility of groups.

In chapter 3 we defined a network formation model inspired by [Germano et al., 2019]. The model allows to define how many nodes generate edges only with nodes of the majority class, how many nodes generate edges only with nodes of the minority class, and how many nodes can generate edges with nodes of both classes. PA also rules the dynamics of the networks. Our main conclusions about the model are the following:

- Due to the model definition, the probability of attracting a new edge grows exponentially with respect to the degree of the node. This large advantage for popular nodes leads to the *winner-takes-all* phenomenon and *hub-and-spoke* topologies.

- The disparate visibility of groups can be observed in networks which do not depend on the homophily parameter.

- In favourable scenarios for the minority class, the larger the number of edges generated by a node, the lower the advantage of the minority class. However, in favourable scenarios for the majority class, the degree of this class does not decrease when more edges are generated by each node.

- The *few-get-richer* effect can be observed in networks in which the homophily mechanism is not considered.

4.1 Further Work

This work can be extended by studying information propagation in the model defined in chapter 2 and comparing it with PAH. This would be useful to understand whether PA is also amplifying or reducing the ability of communication between and within classes. To do so, centrality measures and epidemiology models can be used ([Zhang and Luo, 2017]; [Barabási and Pósfai, 2016, Chapter 10]).

The analysis of information propagation can also be useful in the model defined in chapter 3. It can be useful to find whether reducing the number of nodes in a class also increases the ability to propagate information generated by the nodes of this class.

Finally, given the interesting effect of parameter m in the model defined in chapter 3, it would be interesting to study whether the reported phenomenon in our model is also present in PAH.

Appendix A

FURTHER EXPLANATION OF PAP MODEL

A.1 Numerical example of classes sizes and preference distribution

To fully understand parameters C_0, C_1 P_0, P_1, $P_{0,i}$ and $P_{1,i}$, in this section we will see a numerical example, which correspond to figure A.1.

Let's assume that $n = 20$, $C_0 = 3/4$, $C_1 = 1 - C_0 = 1/4$, $P_0 = 0.6$, $P_1 = 1 - P_0 = 0.4$, $P_{0,i} = 1/3$ and $P_{1,i} = 1/2$.

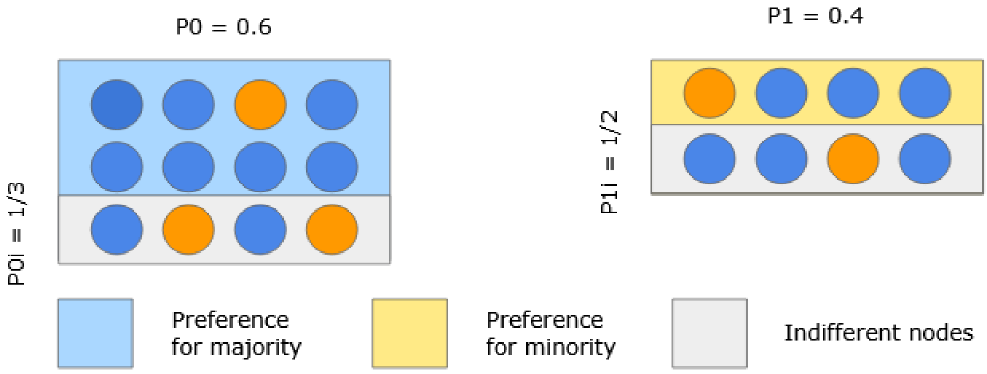
From these parameters, we can deduce that the total number of nodes of the minority class is $n \cdot C_1 = 20 \cdot 1/4 = 5$. This is why in figure A.1 only 5 nodes are orange. Analogously, the number of nodes of the majority class can be computed as $n \cdot C_0 = 20 \cdot 3/4 = 15$.

Given that $P_0 = 0.6$, the 60% of nodes have prior preference for the majority class, i.e., $P_0 \cdot n = 0.6 \cdot 20 = 12$ nodes. These are the 12 nodes in the left block of nodes in figure A.1. However, given that $P_{0,i} = 1/3$, then $12 \cdot P_{0,i} = 12 \cdot 1/3 = 4$ of these nodes at the end are indifferent. The remaining $12 - 4 = 8$ nodes preserve their preference for the majority class.

Analogously, given that $P_1 = 0.4$, then $n \cdot P_1 = 20 \cdot 0.4 = 8$ nodes have a prior preference for the minority class. Since $P_{1,i} = 1/2$, then $8 \cdot P_{1,i} = 8 \cdot 1/2 = 4$ of these nodes at the end are indifferent. The remaining $8 - 4 = 4$ nodes preserve their preference for the minority class.

All in all, there are 8 nodes with preference for the majority class, 4 nodes with preference for the minority class and 8 indifferent nodes.

Figure A.1: Numerical example of class sizes and preference distribution



Classes sizes and preference distribution in a particular example. Blue nodes correspond to those of the majority class, and orange nodes belong to the minority class. Nodes in the left block correspond to those with prior preference for the majority class, and those in the right block correspond to those with prior preference for the minority class. The colour in the background of each nodes indicates its preference: light blue indicates preference for the majority class, light orange indicates preference for minorities, and grey indicates that the node is indifferent.

A.2 The Preference Distribution

In this section we will develop which is the preference distribution given P_0, P_1, $P_{0,i}$ and $P_{1,i}$. Let's define p_i as the preference of node i so that $p_i = 0$ if it has preference for the majority class, $p_i = 1$ if it has preference for the minority class and $p_i = 2$ if it is indifferent to the class of the target node.

As we explained in section 3.1, P_0 is the ratio of nodes with prior preference for the majority class, and $P_{0,i}$ is the ratio of them which at the end have no preference for any class. Let Q_0 be the set of nodes with prior preference for the majority class, and let $Q_{0,i}$ the subset of them that at the end are indifferent. P_1, $P_{1,i}$, Q_1 and $Q_{1,i}$ have an analogous meaning but for the minority class.

Keeping in mind the above concepts, it is easy to see that

$$\begin{aligned} Prob[p_i = 0] &= Prob[i \in Q_0 \wedge i \notin Q_{0,i}] \\ &= Prob[i \in Q_0] \cdot Prob[i \notin Q_{0,i} | i \in Q_0] \\ &= P_0 \cdot (1 - P_{0,i}) \end{aligned} \tag{A.1}$$

With an analogous reasoning it can be deduced that

$$Prob[p_i = 1] = P_1 \cdot (1 - P_{1,i}) \tag{A.2}$$

Also, we can compute the probability of being an indifferent node as

$$\begin{aligned} Prob[p_i = 2] &= Prob[i \in Q_{0,i} \vee i \in Q_{1,i}] \\ &= Prob[i \in Q_{0,i}] + Prob[i \in Q_{1,i}] \\ &= Prob[i \in Q_0] \cdot Prob[i \in Q_{0,i} | i \in Q_0] + Prob[i \in Q_1] \cdot Prob[i \in Q_{1,i} | i \in Q_1] \\ &= P_0 \cdot P_{0,i} + P_1 \cdot P_{1,i} \end{aligned} \tag{A.3}$$

Finally, it can be checked that the sum of these probabilities is 1:

$$\begin{aligned} Prob[p_i = 0] + &Prob[p_i = 1] + Prob[p_i = 2] \\ &= P_0(1 - P_{0,i}) + P_1(1 - P_{1,i}) + P_0 P_{0,i} + P_1 P_{1,i} \\ &= P_0 - P_0 P_{0,i} + P_1 - P_1 P_{1,i} + P_0 P_{0,i} + P_1 P_{1,i} \\ &= P_0 + P_1 = 1 \end{aligned} \tag{A.4}$$

All in all, we have deduced that the ratio of nodes with preference for the majority class is $P_0 \cdot (1 - P_{0,i})$, the ratio of nodes with preference for the minority class is $P_1 \cdot (1 - P_{1,i})$ and the ratio of indifferent nodes is $P_0 \cdot P_{0,i} + P_1 \cdot P_{1,i}$.

A.3 Equivalent Model to PAP

Parameters of the model defined in section 3.1 may look confusing because of how the number of indifferent nodes was defined. It was defined with parameters $P_{0,i}$ and $P_{1,i}$ because it simplified the task of finding in which scenarios the over-representation of each class could happen. However, there is an equivalent definition of the network formation model with a more intuitive definition of node preferences.

In this equivalent network formation model, we define $\pi_0 \in [0, 1]$ as the ratio of the n nodes that have preference for the majority class, and therefore, when added to the graph, they will only be attached with nodes of the majority class. Analogously, we define $\pi_1 \in [0, 1]$ as the ratio of n nodes that have preference for the minority class. Notice that the notion of prior preference has disappeared in this model.

There is a new parameter called $\pi_2 = 1 - \pi_0 - \pi_1$, which replaces parameters $P_{0,i}$ and $P_{1,i}$, which we define as the ratio of the n nodes that are indifferent.

Therefore, when added to the graph, they can be attached to nodes both from the majority and minority classes.

With respect to the rest of the parameters of the model (parameters n, m, m_0 and β) they remain with the same definition as in the network formation model defined in section 3.1.

It is easy to see that given a network formation model as defined in section 3.1 with parameters P_0, P_1, $P_{0,i}$ and $P_{1,i}$, an equivalent model as defined in this section could be built with parameters

$$\pi_0 = P_0 \cdot (1 - P_{0,i}) \tag{A.5}$$

$$\pi_1 = P_1 \cdot (1 - P_{1,i}) \tag{A.6}$$

$$\pi_2 = P_0 \cdot P_{0,i} + P_1 \cdot P_{1,i} \tag{A.7}$$

Notice that these equations follow directly from the results in section A.2.

A.4 Anomalies in the Degree Distribution

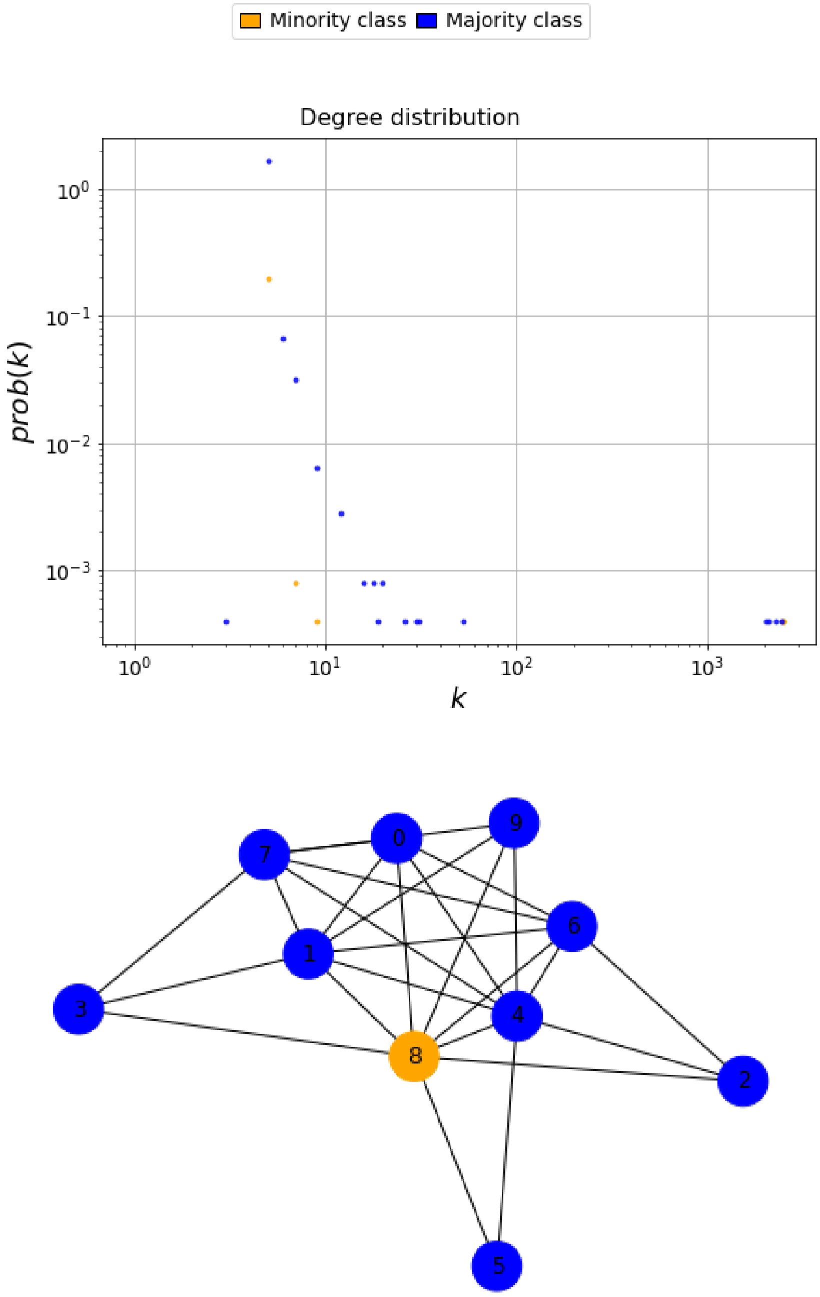
In the degree distributions of figure 3.6, for $C_0 = 0.9$ we can see that some nodes obtained a final degree lower than m, despite m is the number of edges generated by a new node, according to network formation model definition. These anomalies appear because, if it is the case that there are $m' < m$ nodes that satisfy the preference of the node which is going to generate the edges, only m' edges are generated.

To understand why this happens, we will illustrate it with an example, which corresponds to figure A.2. Looking at the degree distribution of the final network, we see that some nodes only obtained degree 3, despite $m = 5$. Actually, this node is node 3. If we check for the initial graph, we can see that in the initial state node 3 has degree 3. This is because node 3 has preference for the minority class, and since the initial graph only has one node of the minority class, node 3 only generated one edge (with node 8); the other two edges attached to node 3 were generated by nodes 1 and 7, which have preference for the majority class.

It is easy to see that these anomalies correspond to nodes which were added to the graph when there were less than m nodes that satisfied their preference. These anomalies should not be considered very important when analysing the degree distribution.

Notice that nodes 2, 5 and 9 also started with a degree lower than m because of the same reason as node 3. However, they were able to capture at least m edges during the whole execution of the model.

Figure A.2: Example of anomalies in the degree distribution



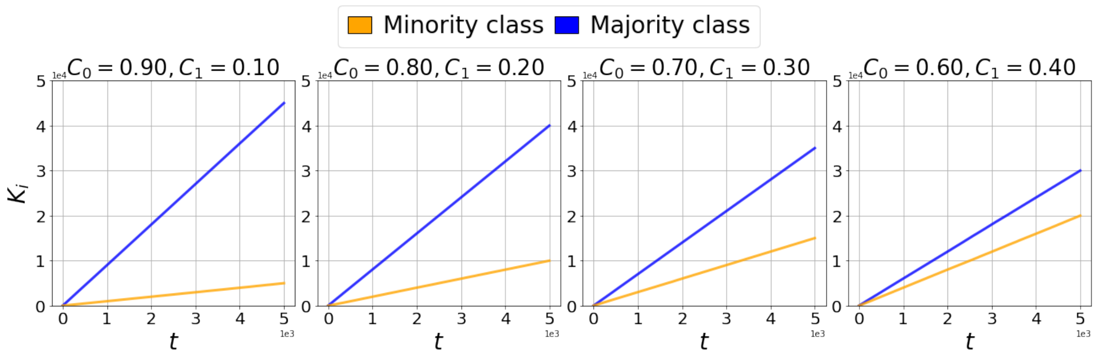
Example of graph in which some nodes did not achieve a degree larger than m. First plot: degree distribution of the minority and majority classes. Second plot: initial graph. For this execution, $n = 5000$, $m = 5$, $m_0 = 10$, $C_0 = 0.9$, $C_1 = 0.1$, $P_0 = P_1 = 0.5$, $P_{0,i} = P_{1,i} = 0$ and $\beta = 1.15$.

A.5 Degree Growth

In this section we can see the figures of the average degree growth of both classes for some of the experiments explored in section 3.2. Figures A.3, A.4 and A.5 correspond to the scenarios in which no class is expected to be over-represented, figures A.6, A.7 and A.8 correspond to scenarios favourable to the minority class, and figures A.9, A.10 and A.11 correspond to scenarios favourable to the majority class.

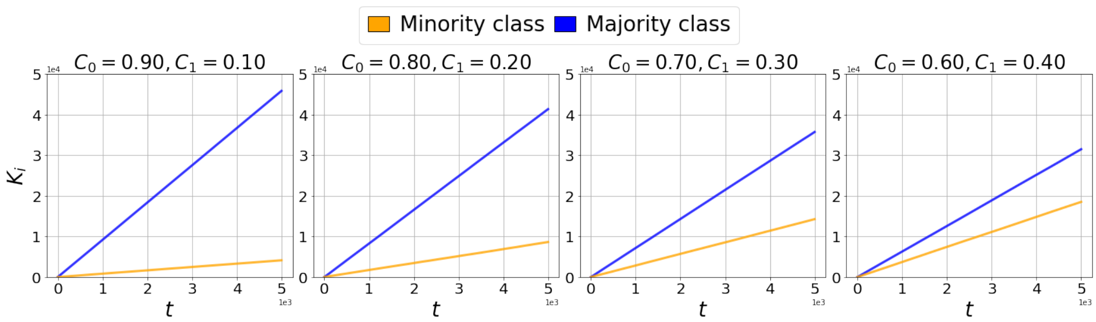
In all cases, the degree of both cases seems to evolve linearly.

Figure A.3: Degree growth for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 0$



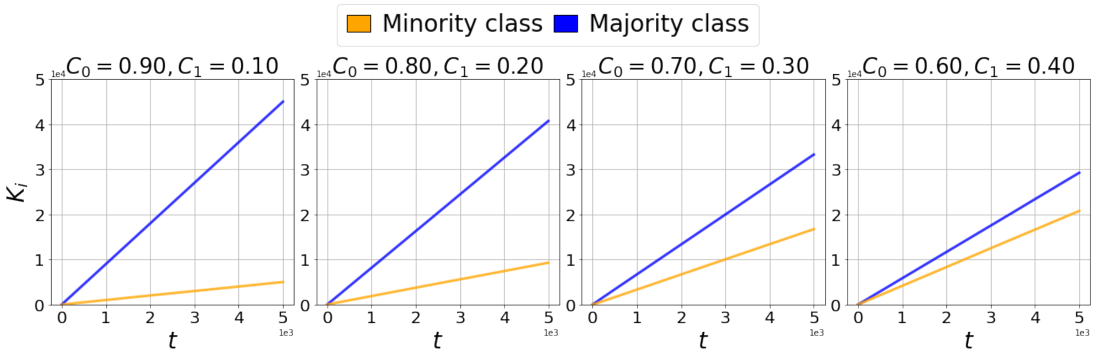
Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = P_{1,i} = 0$ and $\beta = 1.15$.

Figure A.4: Degree growth for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 0.5$



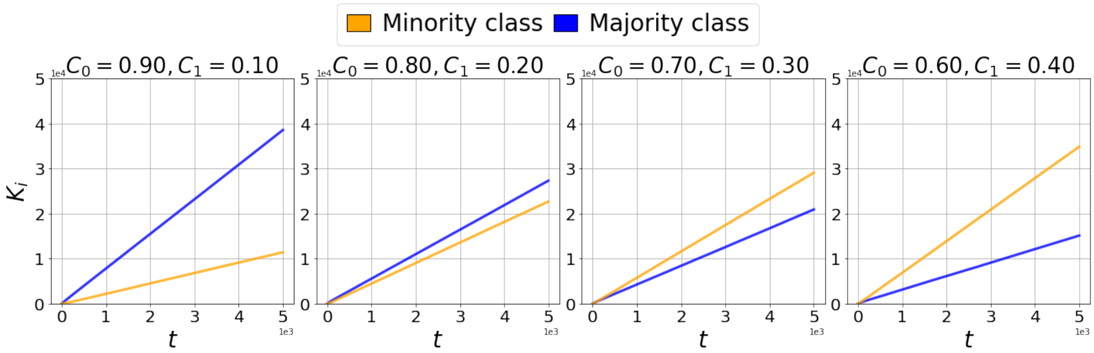
Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = P_{1,i} = 0.5$ and $\beta = 1.15$.

Figure A.5: Degree growth for $P_0 = C_0$, $P_1 = C_1$ and $P_{0,i} = P_{1,i} = 1$



Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = P_{1,i} = 1$ and $\beta = 1.15$.

Figure A.6: Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$



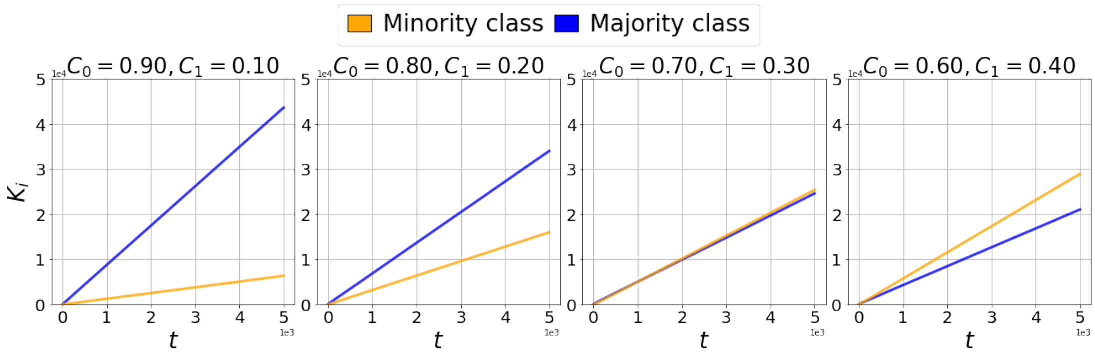
Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$, $P_{1,i} = 0$ and $\beta = 1.15$.

A.6 Degree Distribution

In this section we can see the degree distribution for some of the experiments explored in section 3.2.

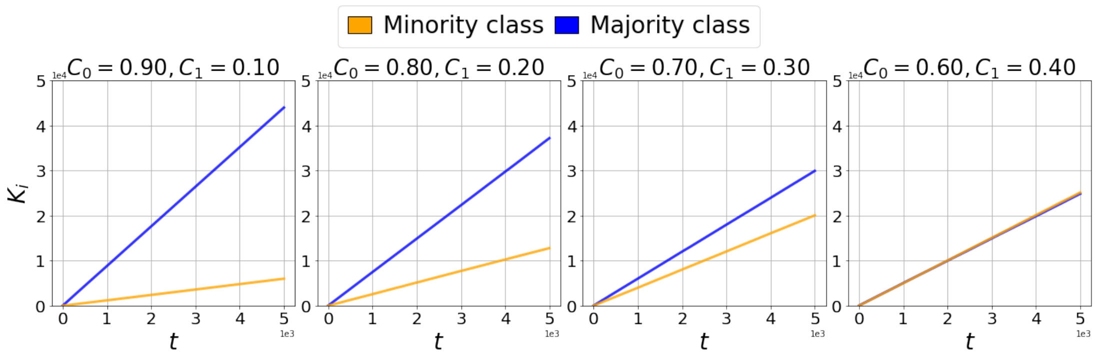
For all the experiment of this section $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$ and $\beta = 1.15$. The results have been averaged across 20 executions. Figures A.12, A.13 and A.14 correspond to the scenarios in which no class is expected to be over-represented, figures A.15, A.16 and A.17 correspond to scenarios favourable to the minority class, and figures A.18, A.19 and A.20 correspond to scenarios favourable to the majority class.

Figure A.7: Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0.5$



Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$, $P_{1,i} = 0.5$ and $\beta = 1.15$.

Figure A.8: Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 0$

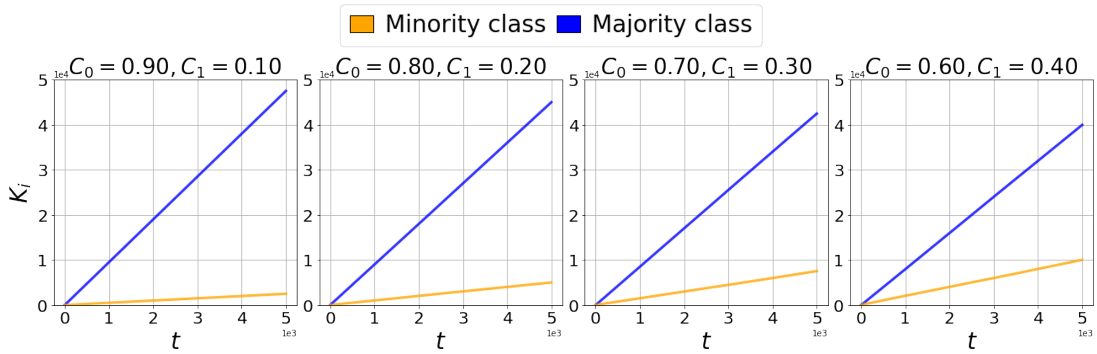


Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$, $P_{1,i} = 0$ and $\beta = 1.15$.

A.7 Experiments to Study the *Few-get-richer* Effect

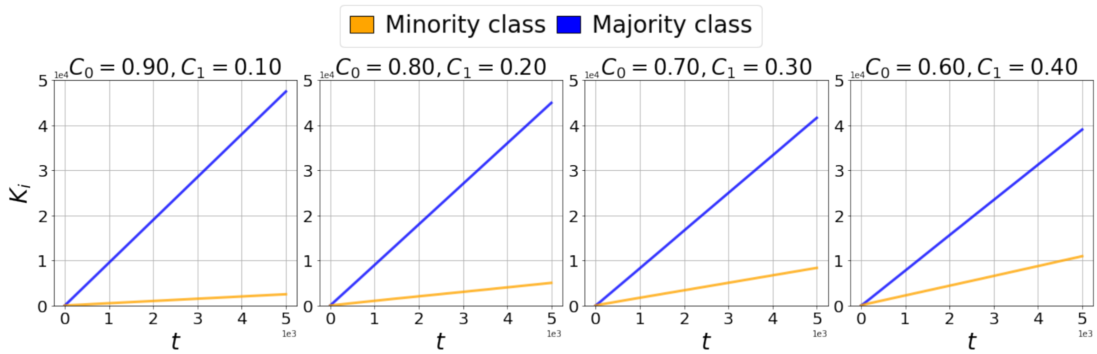
In this subsection we can see the remaining nine attempts to observe the *few-get-richer* effect. The experiments begin at figure A.21 and end at figure A.29.

Figure A.9: Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$



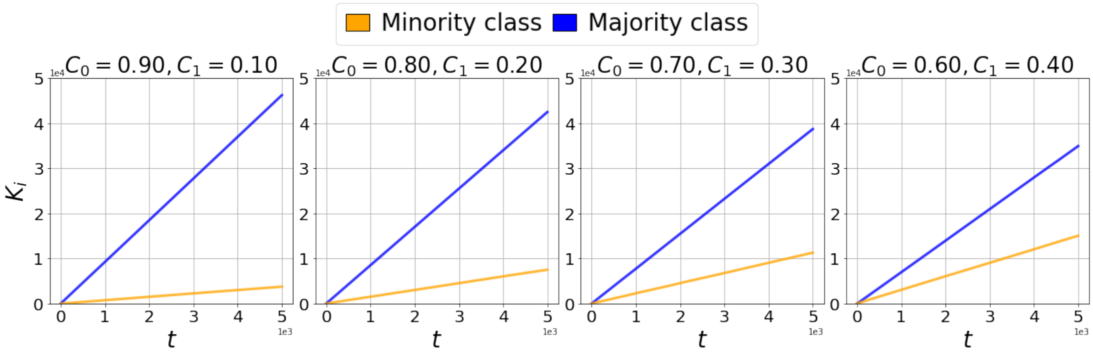
Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$, $P_{1,i} = 1$ and $\beta = 1.15$.

Figure A.10: Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 1$



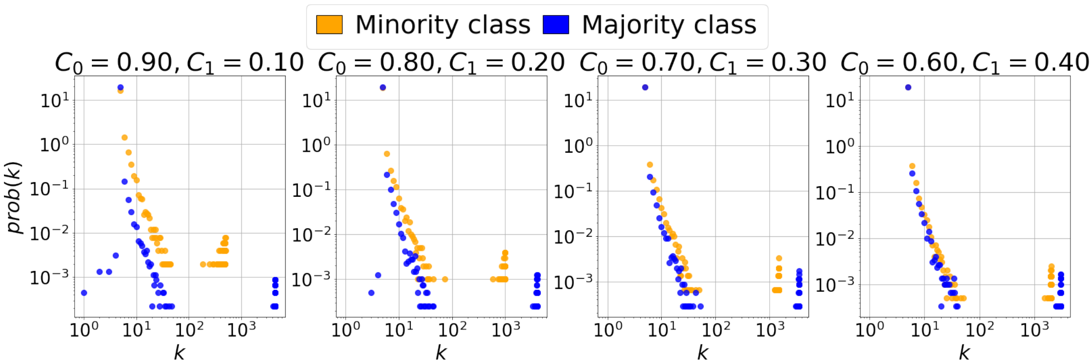
Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$, $P_{1,i} = 1$ and $\beta = 1.15$.

Figure A.11: Degree growth for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 0.5$



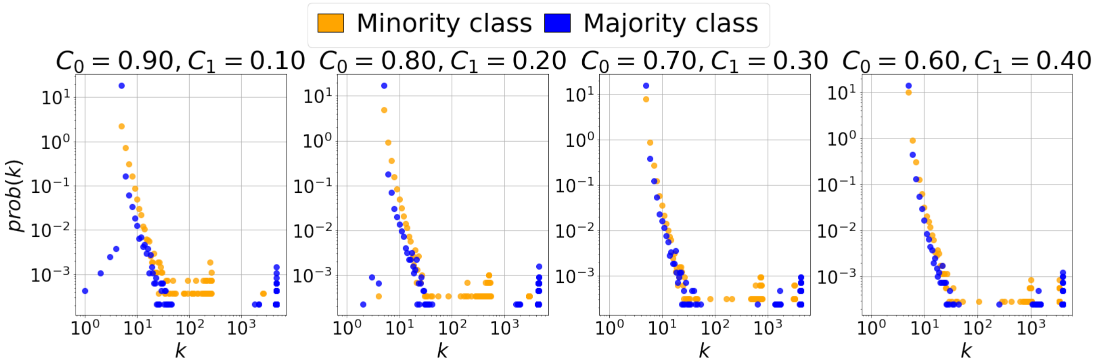
Degree growth of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$, $P_{1,i} = 0.5$ and $\beta = 1.15$.

Figure A.12: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 0$



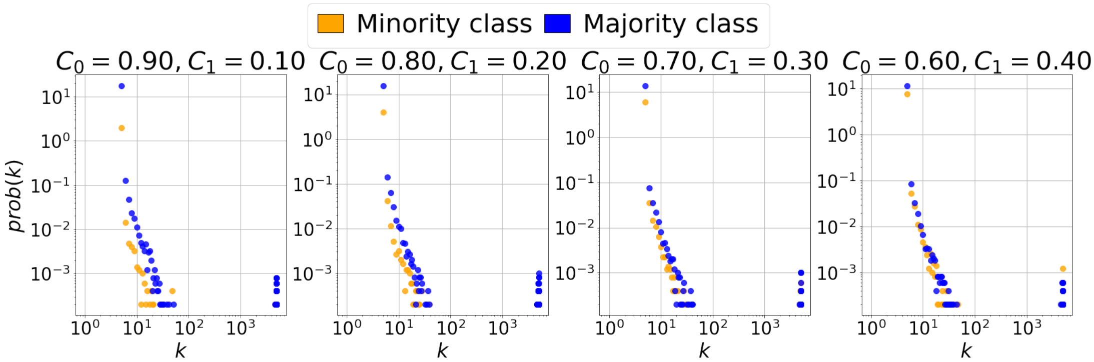
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$, $P_{1,i} = 0$ and $\beta = 1.15$.

Figure A.13: Degree distribution for $P_0 = C_0$, $P_1 = C_1$ $P_{0,i} = 0.5$ and $P_{1,i} = 0.5$



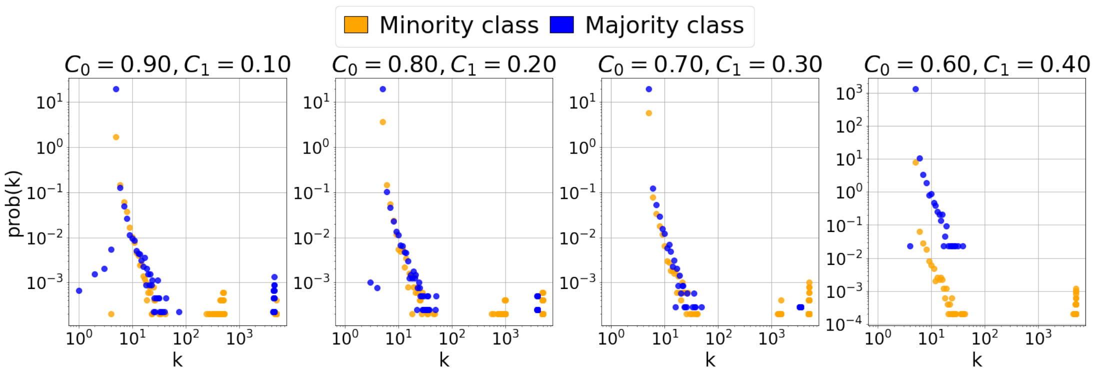
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$, $P_{1,i} = 0.5$ and $\beta = 1.15$.

Figure A.14: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 1$



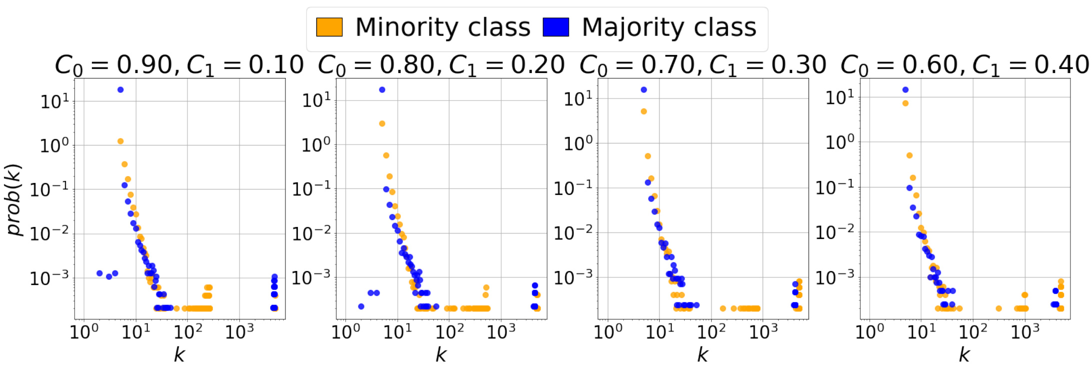
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$, $P_{1,i} = 1$ and $\beta = 1.15$.

Figure A.15: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0$



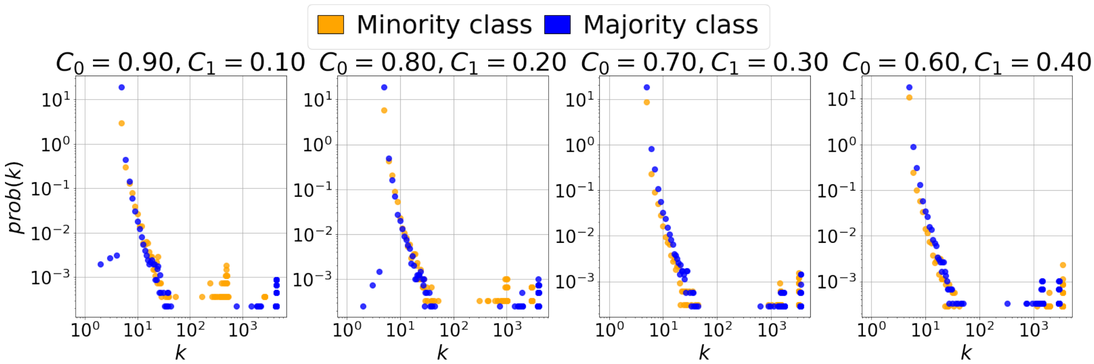
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$, $P_{1,i} = 0$ and $\beta = 1.15$.

Figure A.16: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$ and $P_{1,i} = 0.5$



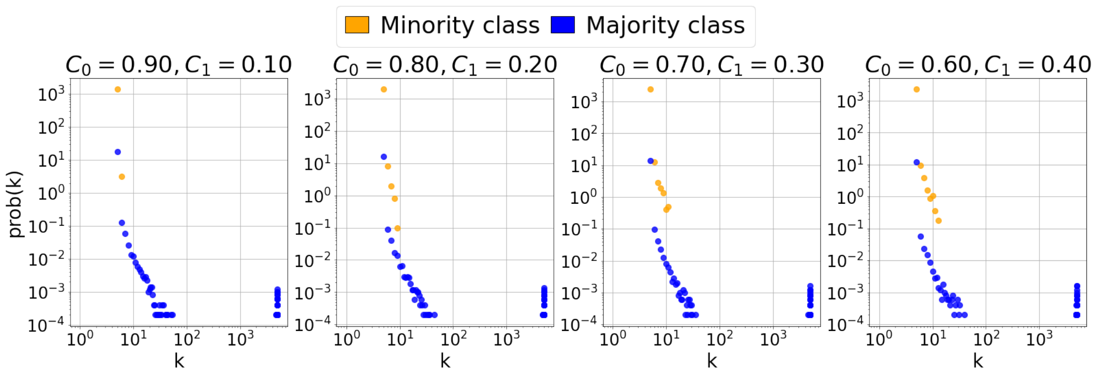
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 1$, $P_{1,i} = 0.5$ and $\beta = 1.15$.

Figure A.17: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 0$



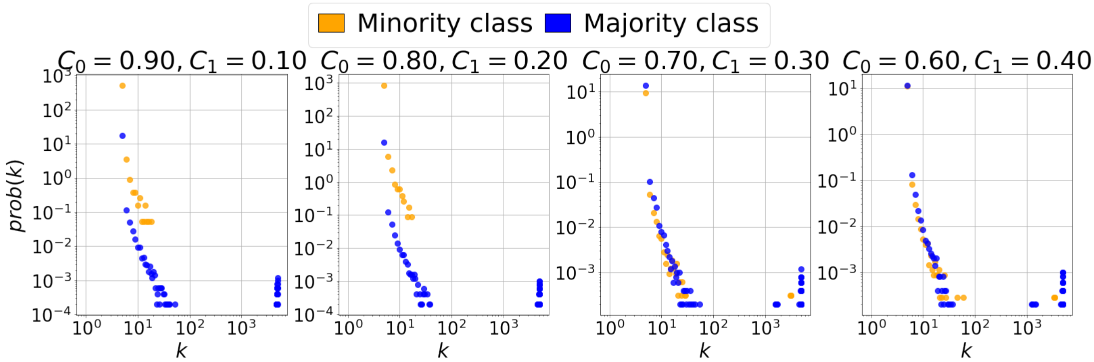
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$, $P_{1,i} = 0$ and $\beta = 1.15$.

Figure A.18: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 1$



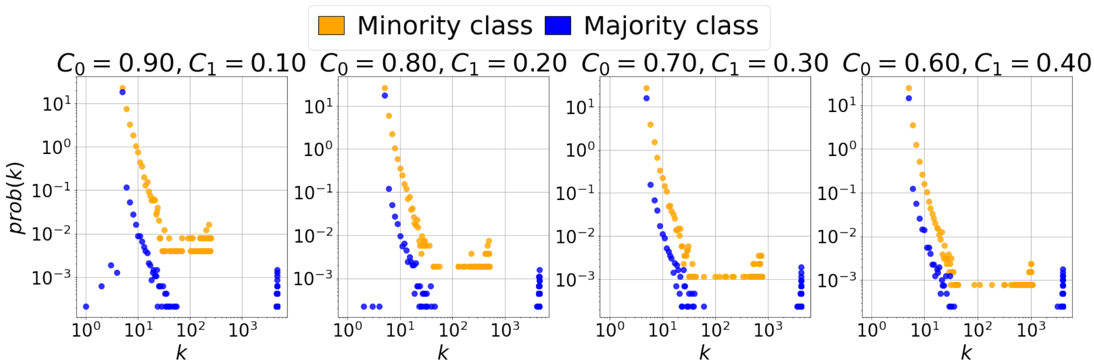
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$, $P_{1,i} = 1$ and $\beta = 1.15$.

Figure A.19: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$ and $P_{1,i} = 1$



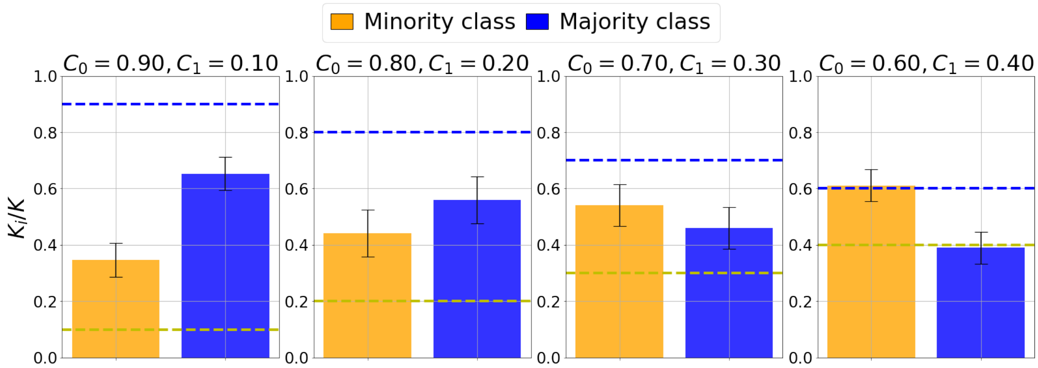
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0.5$, $P_{1,i} = 1$ and $\beta = 1.15$.

Figure A.20: Degree distribution for $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$ and $P_{1,i} = 0.5$



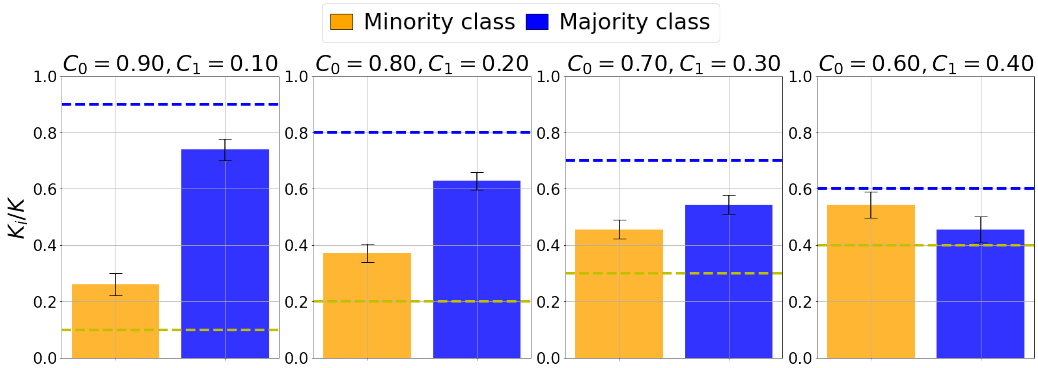
Degree distribution of the majority and minority classes. In all scenarios $m = 5$, $m_0 = 10$, $P_0 = C_0$, $P_1 = C_1$, $P_{0,i} = 0$, $P_{1,i} = 0.5$ and $\beta = 1.15$.

Figure A.21: *Few-get-richer* experiment for $m = 5$, $P_0 = 0.8$, $P_{0,i} = 1$ and $P_{1,i} = 0$



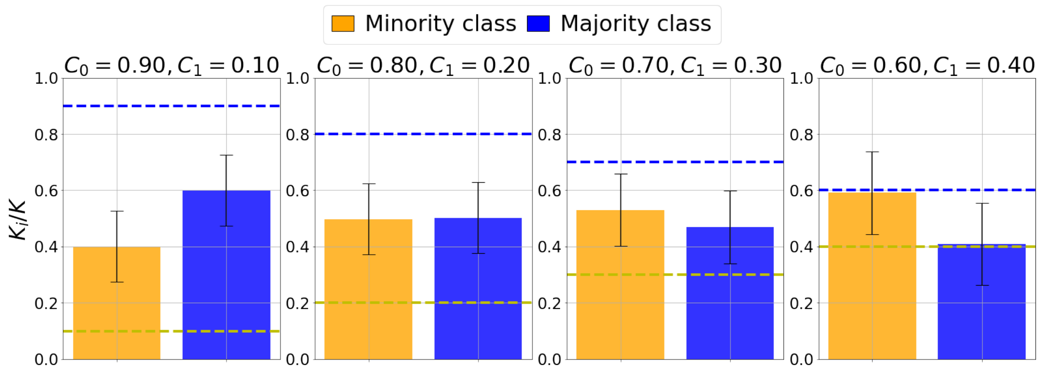
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 1$, $P_{1,i} = 0$, $P_0 = 0.8$, $P_1 = 0.2$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.22: *Few-get-richer* effect experiment for $m = 9$, $P_0 = 0.8$, $P_{0,i} = 1$ and $P_{1,i} = 0$



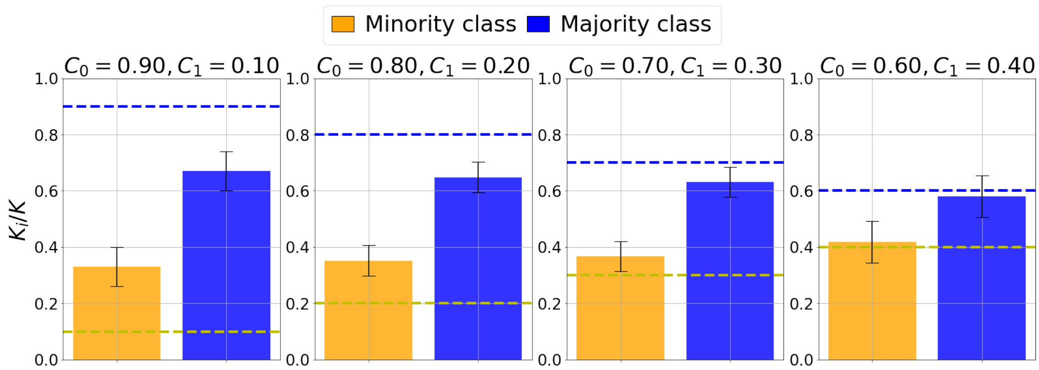
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 1$, $P_{1,i} = 0$, $P_0 = 0.8$, $P_1 = 0.2$ and $m = 9$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.23: *Few-get-richer* effect experiment for $m = 2$, $P_0 = 0.9$, $P_{0,i} = 1$ and $P_{1,i} = 0$



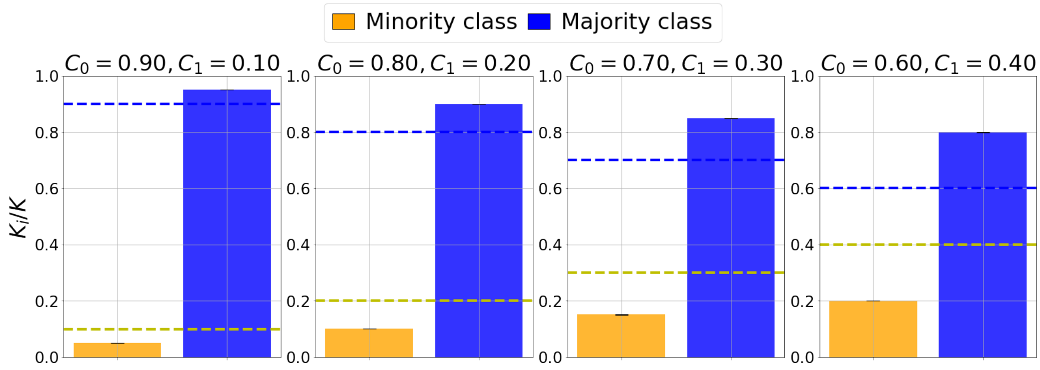
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 1$, $P_{1,i} = 0$, $P_0 = 0.9$, $P_1 = 0.1$ and $m = 2$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.24: *Few-get-richer* effect experiment for $m = 2$, $P_0 = 0.8$, $P_{0,i} = 0.6$ and $P_{1,i} = 0$



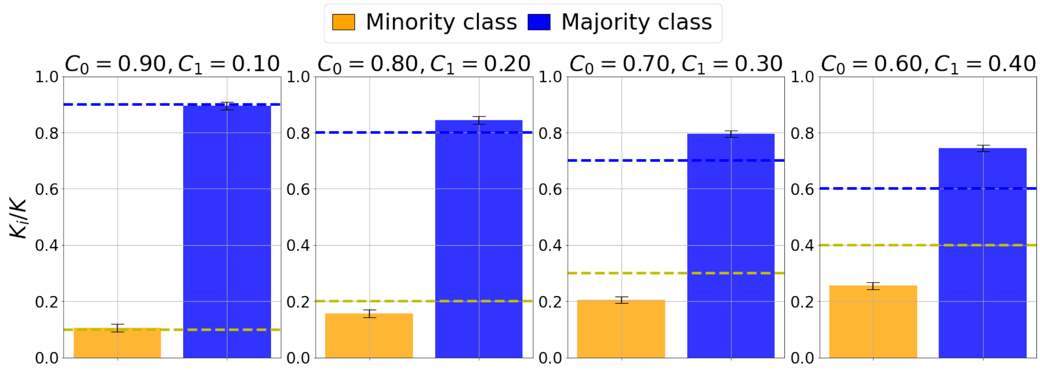
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0.6$, $P_{1,i} = 0$, $P_0 = 0.8$, $P_1 = 0.2$ and $m = 2$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.25: *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 1$



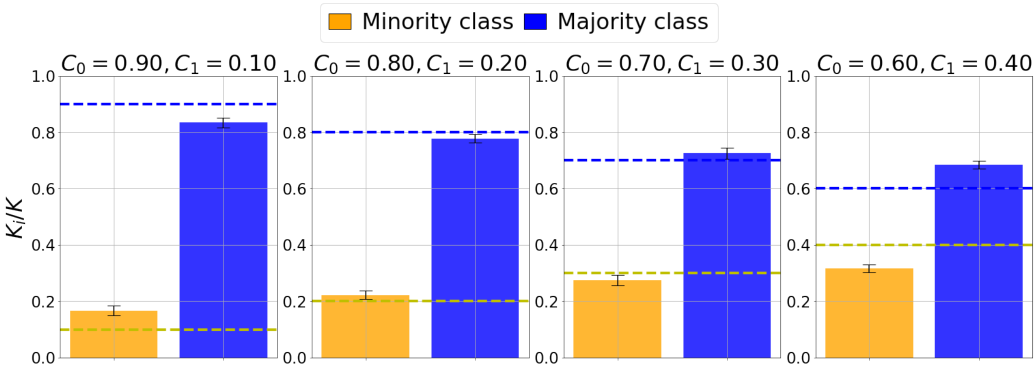
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0$, $P_{1,i} = 1$, $P_0 = 0.5$, $P_1 = 0.5$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.26: *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 0.8$



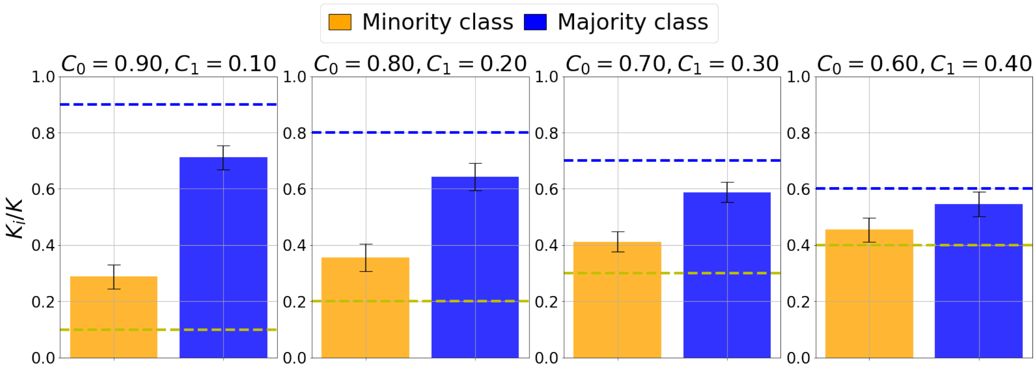
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0$, $P_{1,i} = 0.8$, $P_0 = 0.5$, $P_1 = 0.5$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.27: *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0$ and $P_{1,i} = 0.6$



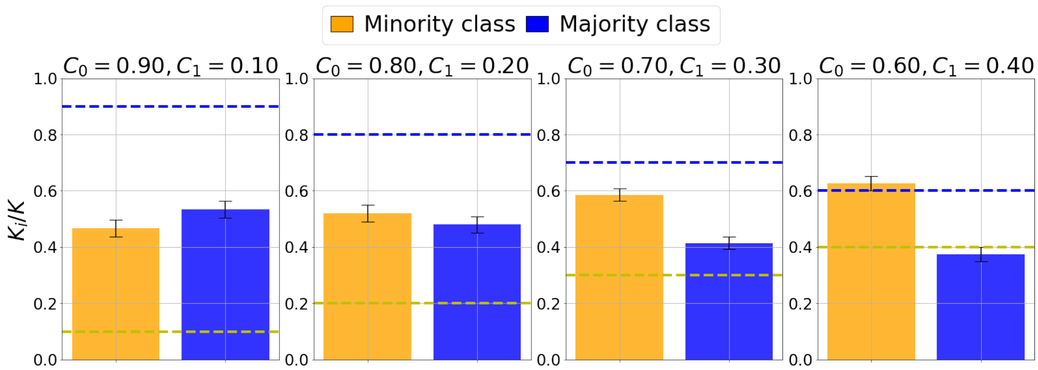
Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0$, $P_{1,i} = 0.6$, $P_0 = 0.5$, $P_1 = 0.5$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.28: *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.5$, $P_{0,i} = 0.5$ and $P_{1,i} = 0.5$



Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0.5$, $P_{1,i} = 0.5$, $P_0 = 0.5$, $P_1 = 0.5$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Figure A.29: *Few-get-richer* effect experiment for $m = 5$, $P_0 = 0.2$, $P_{0,i} = 0.5$ and $P_{1,i} = 0.5$



Degree of each class for several values of C_0 and C_1. In all cases, $P_{0,i} = 0.5$, $P_{1,i} = 0.5$, $P_0 = 0.2$, $P_1 = 0.8$ and $m = 5$. The orange and blue dashed lines indicate which would be the degree of each class if their representation were proportional to their size. The vertical lines correspond to the standard deviation of the class degree centred in the mean value.

Bibliography

[Avin et al., 2015] Avin, C., Keller, B., Lotker, Z., Mathieu, C., Peleg, D., and Pignolet, Y. A. (2015). Homophily and the glass ceiling effect in social networks. pages 41–50. Association for Computing Machinery, Inc.

[Barabási and Pósfai, 2016] Barabási, A.-L. and Pósfai, M. (2016). *Network science*. Cambridge University Press, Cambridge. http://barabasi.com/networksciencebook/, last accessed on 28/03/2022.

[Chen et al., 2018] Chen, L., Ma, R., Hannák, A., and Wilson, C. (2018). Investigating the impact of gender on rank in resume search engines. volume 2018-April. Association for Computing Machinery.

[Espín-Noboa et al., 2022] Espín-Noboa, L., Wagner, C., Strohmaier, M., and Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific Reports*, 12.

[Fabbri et al., 2021] Fabbri, F., Croci, M. L., Bonchi, F., and Castillo, C. (2021). Exposure inequality in people recommender systems: The long-term effects.

[Germano et al., 2019] Germano, F., Gómez, V., and Le Mens, G. (2019). The few-get-richer: A surprising consequence of popularity-based rankings. In *The World Wide Web Conference*, pages 2764–2770.

[Gómez et al., 2021] Gómez, E., Boratto, L., and Salamó, M. (2021). Disparate impact in item recommendation: A case of geographic imbalance. *Lecture Notes in Computer Science*, page 190–206.

[Karimi et al., 2018] Karimi, F., Génois, M., Wagner, C., Singer, P., and Strohmaier, M. (2018). Homophily influences ranking of minorities in social networks. *Scientific reports*, 8(1):1–12.

[Mcpherson et al., 2001] Mcpherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks.

[Singh and Joachims, 2018] Singh, A. and Joachims, T. (2018). Fairness of exposure in rankings. pages 2219–2228. Association for Computing Machinery.

[Zhang and Luo, 2017] Zhang, J. and Luo, Y. (2017). Degree centrality, betweenness centrality, and closeness centrality in social network.