Université de Bordeaux

Network analysis : project

Arno BARRABÈS
Jean DESSAGNE-ARENDO

April, 20th of 2022

Contents

1 Introduction **1**

2 Comparison of the two networks **2**
2.1 Presentation 2
2.2 Degree 2
2.3 Clustering 3
2.4 Centrality 3
2.5 Communities 4

3 Evolution of the indicators regarding to the variables **5**
3.1 Gender 5
3.2 Core 6
3.3 Coord 7

4 Conclusion **7**

1 Introduction

The objective of the present project is to work on a network in order to visualise it in the form of graphs and to analyse its structure.

The data in our possession are a cluster network : hundreds of scientists were observed before and after a policy intervention. The variables that we look at are "coord" (if the scientist is the coordinator of the cluster), "core" (if the scientist is a core member of the cluster) and "gender" (female/male).

After creating the two networks (before and after the policy intervention), we will depict and analyze their structure and their evolution between the two periods, showing some relevant network statistics and graphs. Then we will discuss network connections with respect to gender, coord and core attributes.

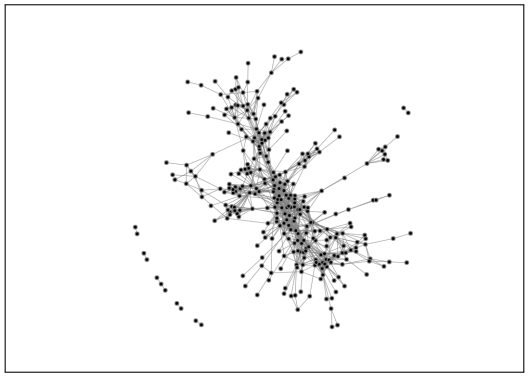
We must keep in mind that this is not a rigorous scientific experiment due to the absence of a control cluster that would not be impacted by the policy. Therefore, we cannot affirm that the evolution of the cluster is one hundred percent due to this policy intervention. However, in the following document we won't remind it each time ; it would take too much words.

One shall find the code we have created and used to get our results in the appendices inside the attached document.

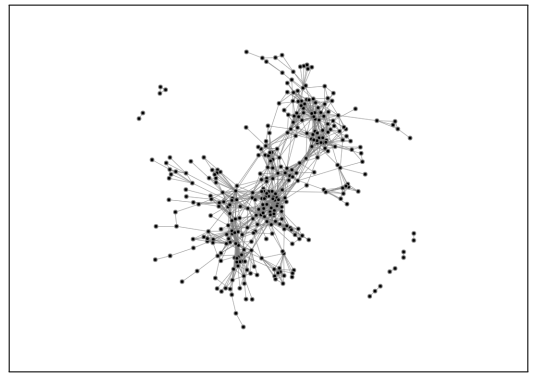
2 Comparison of the two networks

2.1 Presentation

During this whole project, we will study two different networks created from the same dataframe : one for each period, before and after the policy intervention. Let's start with a graphic representation of the two networks.



(a) Before the policy.



(b) After the policy.

Figure 1: Graphic representation of the two networks.

Both networks are composed of seven components. Our first reflex is to count the edges and the vertices of the two networks and to compute their ratio "number of vertices / number of edges". We can see below that the policy seems to increase both the number of vertices and edges while maintaining a quasi constant ratio.

	Before	After	Evolution
Number of vertices	318	347	+9.12%
Number of edges	2502	2880	+15.11%
Ratio	0.13	0.12	

An interesting thing to look at is how many links between scientists the policy destroyed, created or preserved : 1 316 links disappeared, 1 694 new links appeared and 1 186 stayed untouched. We can conclude that the policy intervention greatly restructured the cluster.

2.2 Degree

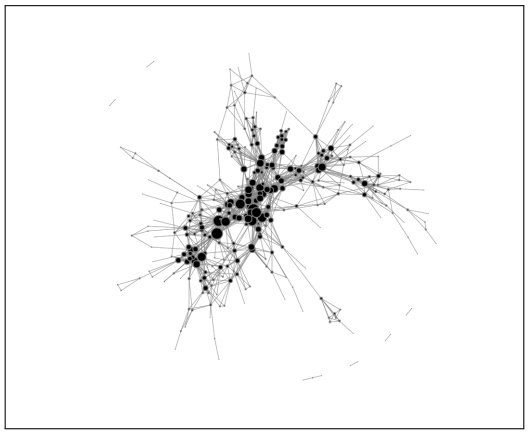
We are now going to take an interest in the notion of degree (the number of links connected to a given vertex). We can present the results the same way as before :

	Mean			Standard deviation		
	Before	After	Evolution	Before	After	Evolution
Degree	15.74	16.6	+5.46%	13.22	12.32	-6.81%

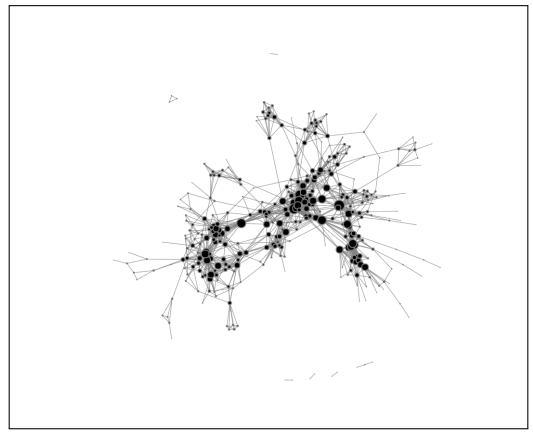
The mean degree went from 15.74 to 16.6, meaning that after the policy, a scientist is connected to 16.6 other scientists on average, almost one more than before. The standard deviation decreased a little so there are less extreme values (scientists with small/big degrees), so we can conclude by

saying that the policy seems to reinforce connections between scientists.

Let's modify the two previous graphs by adding the notion of degree : the higher the degree of a vertex is, the bigger it will appear.



(a) Before the policy.



(b) After the policy.

Figure 2: Graphic representation of the two networks regarding to the degree.

We note that the big vertices are inside or near the center of the network while the vertices at the periphery are almost invisible ; it is not a surprise because the plot function we used is based on aesthetics criteria that put high-connected vertices at the center.

2.3 Clustering

Now, we shall compute the clustering coefficient of our two networks. Those two numbers correspond to the probability of the existence of a link between two vertices knowing that they share a common neighbour, so it measures in a way the density of the network. Before the policy, the coefficient is 43.5% and after it is 46.1% : another proof that the policy tends to get the scientists closer to each other.

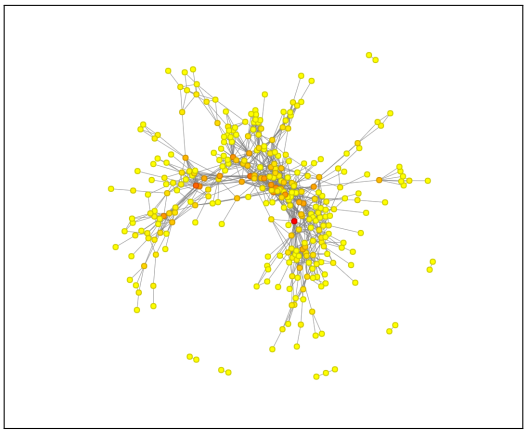
2.4 Centrality

A very important notion in network analysis is centrality. There exist various types of centrality ; we will use the "betweenness centrality". The betweenness centrality (denoted "BC" from now) of a vertex is an indicator of how many shortest paths contain this vertex : it measures how important the node is to the flow of information through a network.

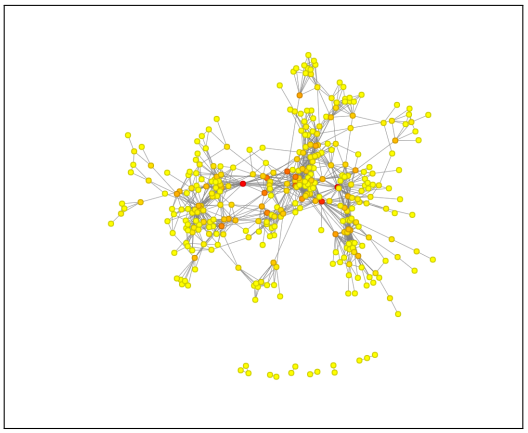
	Mean			Standard deviation	
	Before	After	Evolution	Before	After
Betweenness centrality	490.4	544.4	+11%	866.3	989.9

The increasing of the BC on average confirms the conlusion we made with the clustering coefficient : the policy seems to give more importance to every scientists by making the cluster denser. We note that the standard deviation is quite high before and after because we can see that there are various 2-tuple and 3-tuple disconnected of the giant component, so their BC are very low. Let's modify the two previous graphs by adding the notion of BC : the higher the BC of a vertex is, the

more red it will appear. Unsurprisingly, there are more red and orange vertices after than before the policy intervention.



(a) Before the policy.



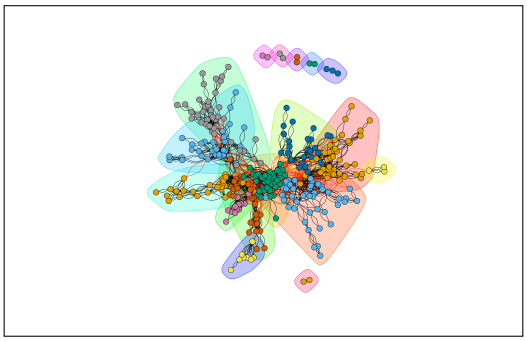
(b) After the policy.

Figure 3: Graphic representation of the two networks regarding to the BC.

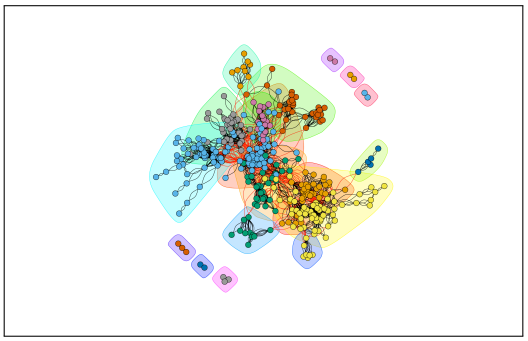
2.5 Communities

Our last statistic is the notion of community : a network is said to have community structure if the nodes of the network can be easily grouped into sets of nodes such that each set of nodes is densely connected internally. Pairs of nodes are more likely to be connected if they are both members of the same community, and less likely to be connected if they do not share communities.

We use Louvain's method and we get 17 (respectively 18) communities before (resp. after) the policy intervention with a modularity of 67% (resp. 71%) : it is greater than 30% so the segmentation made is pertinent. The policy does not seem to have a significant impact on communities. Let's modify the two previous graphs by adding the notion of community.



(a) Before the policy.



(b) After the policy.

Figure 4: Graphic representation of the two networks regarding to the communities.

3 Evolution of the indicators regarding to the variables

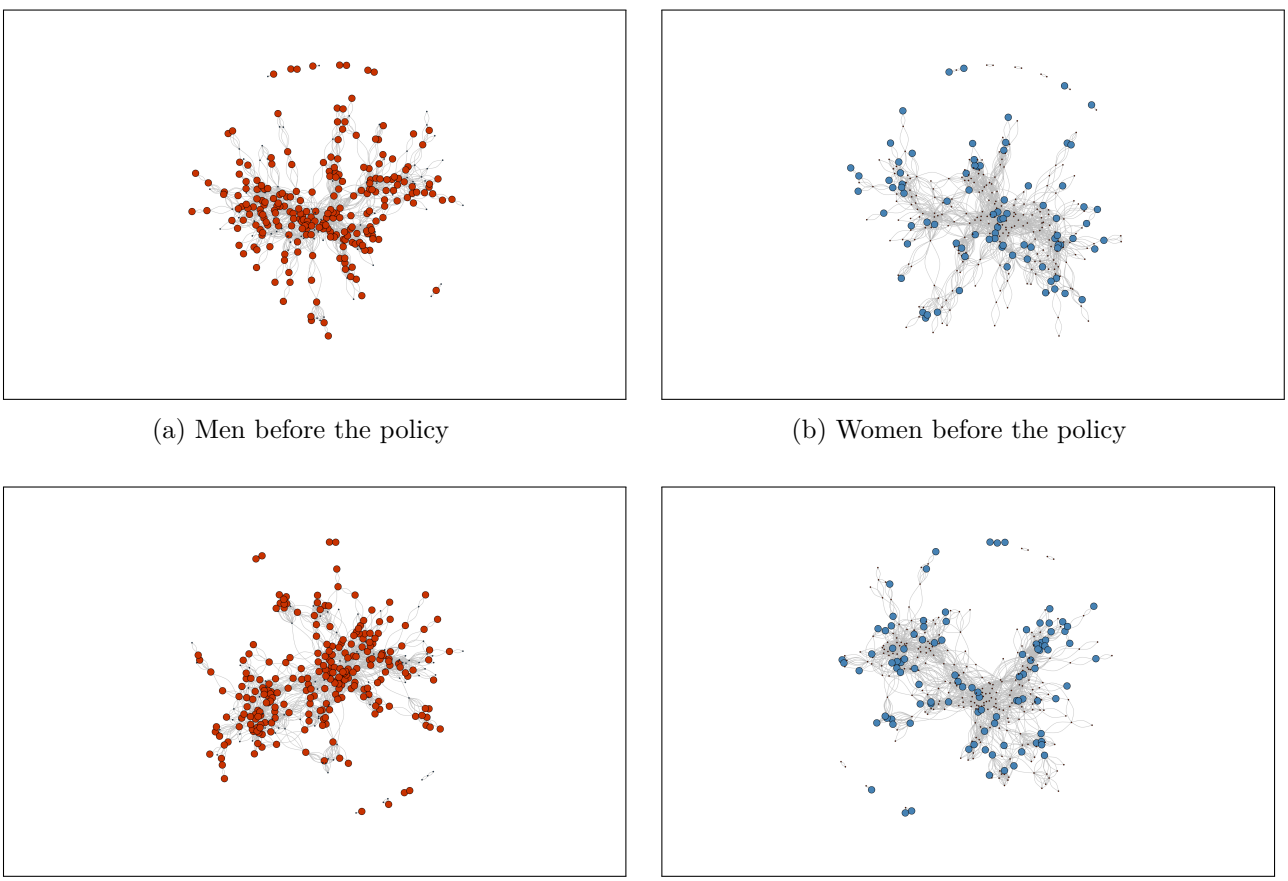
3.1 Gender

Since the gender of every scientist is known, we can analyze the networks regarding to the gender : does the policy intervention favoured the women in order to improve the density of the women network or the centrality of the women scientist ?

The first observation is that there are much more men than women in the cluster ; nevertheless, the policy added additional women (and some men, to a lesser extent if we look at the evolution) scientists, so the ratio is a little bit more egalitarian.

	Before	After	Evolution
Number of women	84	100	+19.1%
Number of men	234	247	+5.6%
Proportion of women	26.4%	28.8%	

Let's now create the graphs regarding to the period and the gender : we colour the vertices of women in blue and the vertices of men in red.



(a) Men before the policy

(b) Women before the policy

(c) Men after the policy

(d) Women after the policy

Figure 5: Graphic representation of the two networks regarding to the two genders.

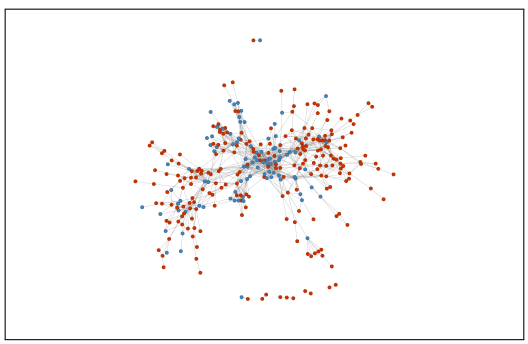
We will now show how the degree and the betweenness centrality are altered among men and women in the cluster.

	Mean degree			Mean BC		
	Before	After	Evolution	Before	After	Evolution
Women	13.48	14.84	+10.1%	277	383	+38.3%
Men	16.55	17.31	+4.6%	567	610	+7.6%

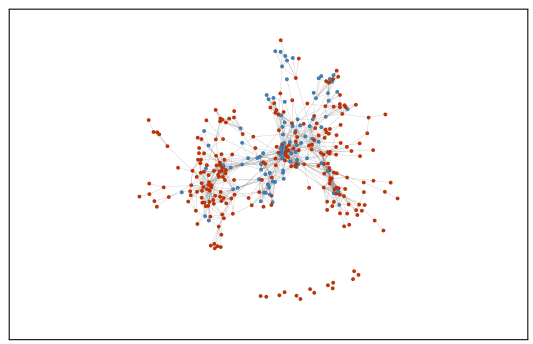
We can observe that even though the women mean degree and the women mean BC still remain lower than for the men, the policy seems to greatly increase the BC and the degree of the women scientists, meaning that the women have more importance in the cluster than before. In average, woman earned a more central place inside the cluster of scientists, they expanded more their network relative to men.

3.2 Core

First, we can create the graphs regarding to the period and the gender : we colour the vertices of core scientists in blue and the vertices of the non core ones in red.



(a) Before the policy.



(b) After the policy.

Figure 6: Graphic representation of the two networks regarding to the core.

Since the core attribute of every scientist is known, the same way we did for the gender, we can analyze the networks regarding to the core. More precisely, we will determine if the fact of being a core scientist has an impact on the degree and the betweenness centrality.

	Core	
	Before	After
Women	21	20
Men	73	75
Total	94	95
Proportion of women	28.8%	26.7%

We can see that the policy did not affect significatively the number of core men and of core women, neither the proportion of core women. It is not grave because what interests us is the following question : does belonging to the core group modify the evolutions of mean degree and BC we saw earlier among women scientists ?

	Mean degree			Mean BC		
	Before	After	Evolution	Before	After	Evolution
Core women	15.9	16.8	+5.7%	383	578	+50.9%
Non core women	12.7	14.4	+13.4%	242	335	+38.4%

We can conclude from those results that the restructuration of the network after the policy intervention did not impact the core women the same way as the non core women. Actually, it seems to improve more the degree of the non core women than the core ones while increasing a little bit more the centrality of core women than non core ones.

3.3 Coord

In this cluster, the scientist whose label is 7440 appears to be the coordinator of the cluster of scientists. He is a man and we can compare his degree and his BC before and after the policy with the mean degree and mean BC of the networks.

	Mean degree		Mean BC	
	Before	After	Before	After
Coordinator	19	18	1074	3881
Whole network	15.7	16.6	490	544

We can oberve that his number of neighbours decreased of one but his betweenness centrality exploded with the policy : he was the 47th vertex (over 318) with the highest BC before the policy intervention, he is the 7th after. Much more shortest paths pass through the coordinator ; he is more central, which seems to be a good quality for a coordinator scientist.

4 Conclusion

Although this study is not designed for rigorously showing the true impact of the policy intervention on the cluster of scientists at our disposal, we can moderate our comments by saying that this policy *seems* that it increased almost every indicator we have studied during the "Network analysis" course.

As a matter of fact, in one hand, the number of vertices and the number of edges grew up ; the mean degree, the mean betweenness centrality and the clustering coefficient raised too ; only the number of communities increased weakly (from 17 to 18).

In another hand, we saw that those rises are even more significant for the women, in front of men, and also how the effects differs between core women and non core women.

Finally, we created various graphs in order to clearly visualise the objects we manipulate and can support our analysis.

Therefore, to conclude with this project, we will say that the policy intervention *tends* to improve the density, the clustering, the centrality and the expansion of the network of scientists by generating more collaborations between more scientists, and more particularly for the women ; it is understandable when we think about the too-small place holded by women in the scientific community : maybe one of the goals of the policy is to correct this aspect ...