Study of the communities inside the European Parliament before and after the Brexit vote

Denis Donadel

January 20, 2019

Contents

0.1 Introduction

In this essay we study with an engineering approach a network based on the connection between MEPs of European Parliament.

We study in the same way a network before and a network after the day when in United Kingdom people have been called to vote "remain" or "leave" the European Union. The day was 23 June 2016 when 51.9 per cent of those who voted supported withdrawal. Withdrawal has been advocated by Eurosceptics, both left-wing and right-wing, while Pro-Europeanists (or European Unionists), who also span the political spectrum, have advocated continued membership.

In this research we hope to find about two/three different big groups, one of the MPs pro-EU, one with MEPs against-EU and maybe a third "in between".

The key point is to understand how a big event like Brexit changes the political system in the European Union.

1 Build of the networks

In order to create the two networks we have collected votes from all the MPs in about 40 resolutions. To build the network before the vote we have considered resolution from April to May, for the other one from July to September.

Votes are available only from verbals of plenary sessions of the European Parliament that are accessible on the European Parliament website [1].

Verbals are available in .pdf format so we have made a Python code in order to count the vote from MEPs and build the adjacency matrix.

We add also an extra column in order to have an immediate index of the European groups, quite useful for plotting graph or other stuff and make them immediately understandable. Finally, we create a table with the correspondence between number of row and surname of the MEPs, always in order to a better quickly understanding of the plots.

First we create a rectangular matrix X where row represent MEPs and columns represent votes. So $x_{ij} = 1$ if MEP i voted *favour* at resolution j, $x_{ij} = -1$ if he voted *against*, $x_{ij} = 0$ if he *abstained* from the vote and finally $x_{ij} = -A$ if he was *absent*.

The adjacency matrix A measure how two different MEPs are linked by count the number of times where they agreed. In other words, it count how many times two MEPs vote in the same way. It's pretty simple starting from the matrix X.

The main diagonal of A have then the biggest values in the matrix because represent the number of votes of every MEP. Note that generally the value is lower than 40 because of the absences at some sessions.

The setting of the adjacency matrix (with groups row) is shown in figure 1.

2 First part

In this first part we choose to focus on the second network (from July to September) that anyway is really similar to the first one, at least for how we are going to analyze it.

There are different way to work on the matrix: for some things we choose to leave the matrix weighted as it is, for other stuff we choose instead to create a

	0	1	2	3	4	5	6	7	8	9	...	845	846	847	848	849	850	851	852	853	854
0	16	0	9	9	9	2	12	9	1	7	...	13	1	13	0	0	10	14	14	13	1
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
2	9	0	40	39	39	8	23	39	7	21	...	26	7	24	0	0	39	25	26	24	2
3	9	0	39	39	39	8	22	39	7	20	...	25	7	23	0	0	38	24	25	23	7
4	9	0	39	39	39	8	22	39	7	20	...	25	7	23	0	0	38	24	25	23	2
5	2	0	8	8	8	40	0	8	35	15	...	1	26	1	0	0	7	0	0	0	2
6	12	0	23	22	22	0	39	22	0	17	...	36	3	37	0	0	24	37	36	34	3
7	9	0	39	39	39	8	22	39	7	20	...	25	7	23	0	0	38	24	25	23	2
8	1	0	7	7	7	35	0	7	36	13	...	1	27	1	0	0	6	0	0	0	2
9	7	0	21	20	20	15	17	20	13	40	...	16	14	17	0	0	21	17	17	16	6
10	10	0	27	26	26	2	29	26	2	21	...	31	5	30	0	0	28	31	32	30	7

Figure 1: First rows of the adjacency matrix with the extra column for the group

threshold in order to obtain a 0s-1s matrix or at least to remove the *backgroud noise*.

2.1 Graph and general information

In the net there are 854 MEPs and so nodes, but some of them are always absent or in some cases some MEPs have lapsed so we can delete them and obtain 748 nodes without any threshold. In this last case the number of edges is 272285.

From now on we considered a $threshold = 20$ and we erase all the isolated nodes that are created in this way. So we get $N = 706$ nodes and "only" $L = 138433$ edges. The graph corresponding to this case is represent in figure 2.

We can compare this data with a random net with $p = 0.55$ that have an expected numbers of links equal to $L_{rnd} = \frac{pN(N-1)}{2} = 136876$, not so far from our case.

2.2 Degree distribution

Estimate a degree distribution in this network is really hard. First, we evaluate the strength of every links based on the weighted matrix and the degree based on the 0-1 adjacency matrix.

Representing the comparison between this two set of values (figure 3) we can see that the behavior is really similar and so we can consider only the 0-1 matrix without lack of generalization. It's really good because simplify the study.

In the net there are many hubs and some node with low degree. In between there are really few nodes. This is visible in figure 4. It is a bimodal behavior, difficult to estimate and out of the program of this course.

We also compute the average degree that is $\langle k \rangle = 392.1615$ and compare as before with random case where $\langle k \rangle_{rnd} = p(N-1) = 387.75$. We verify that $\langle k \rangle > \ln(N) = 6.56$ and so we are in a connected regime as expected.

We compute the moments of the degree distribution:

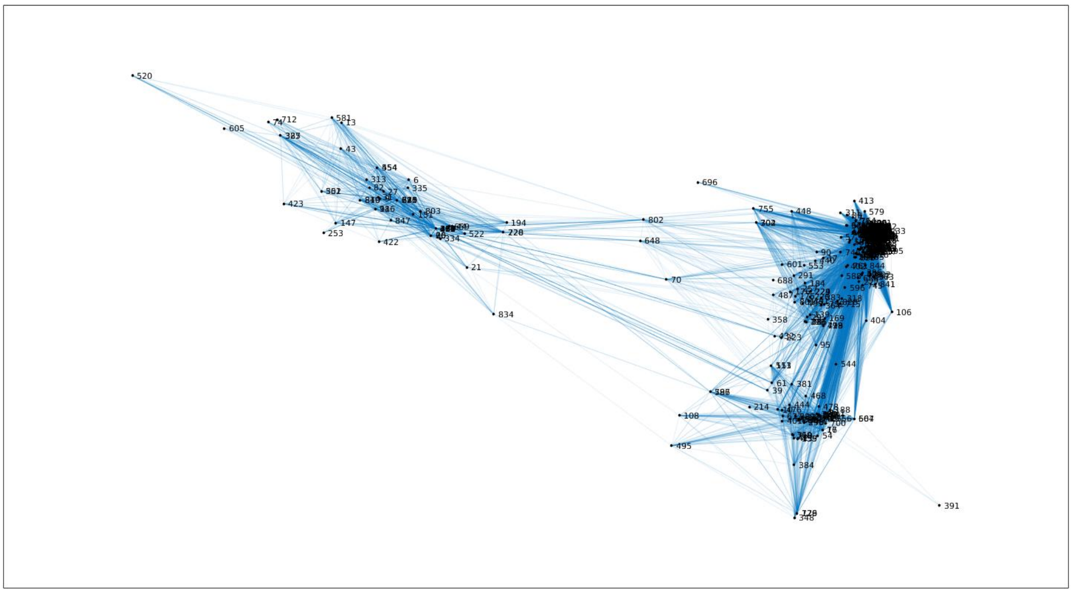


Figure 2: Graph of the network with $threshold = 20$

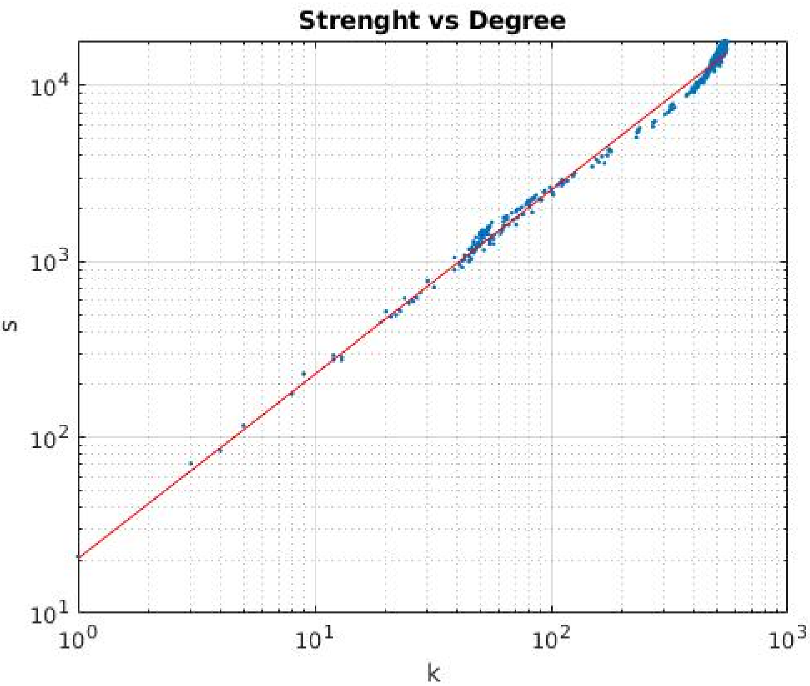


Figure 3: Degree vs strength

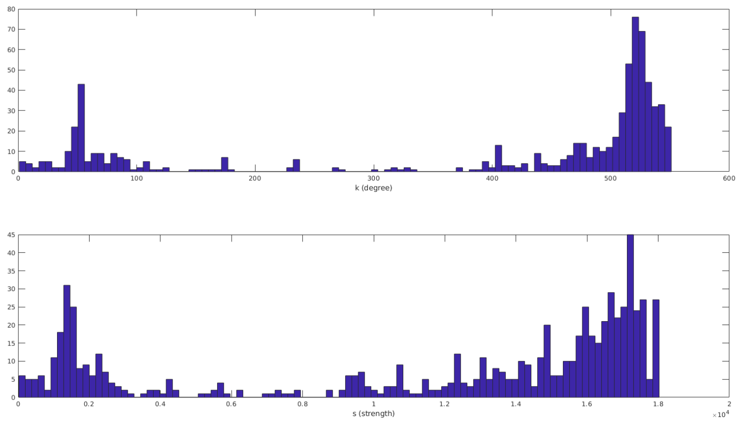


Figure 4: Degree and strength distribution

- $\langle k \rangle = 392.1615$
- $\langle k^2 \rangle = 4.6656 \cdot 10^4$
- $\langle k^3 \rangle = 1.0077 \cdot 10^7$
- $\langle k^4 \rangle = 2.1768 \cdot 10^9$

So we observe that the first ones moments are finite as we expect from a random network.

We measure the diameter $diam = 6$ and the average distance $\langle dist \rangle = 1.6751$. Similar values come from random network: $diam_{rnd} = \ln(N) = 6.56$ and $\langle dist \rangle_{rnd} = \frac{\ln(N_{rnd})}{\ln(\langle k \rangle_{rnd})} = 1.1005$.

Thanks to all of this observation we can say that we are in random regime.

In figure 5 there are some other plots of degree distribution.

Anyway we try to estimate a power law as illustrated in figure 6 with the maximum likelihood fitting. You can see that using all the data the estimation is bad, with a $kmin = 600$ the approximation improve a little bit and give a gamma coefficient $\gamma = 3.2057$. Having such value of $\gamma > 3$ is a further property of random networks.

2.3 Cluster coefficient

In figure 7 is shown the cluster coefficient of the nodes with respect to the degree. In general there are high values of clustering coefficients as we expected due to the huge number of links and hubs in the network. The average cluster coefficient is $\langle C \rangle = 0.89892$. In order to compare with some similar network, in a random net the expected coefficient is about $\langle C \rangle_{rnd} = \langle k \rangle_{rnd}/N = p = 0.55$.

2.4 Assortativity

We estimate a structural cutoff of about $k_s = 526$ and a natural cutoff $kmax$ lower than the structural one.

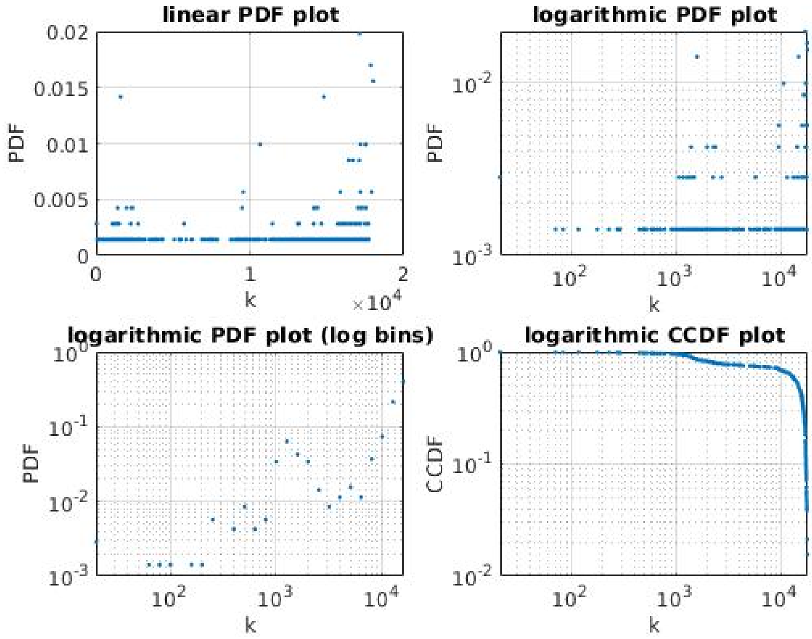


Figure 5: Some other plots of degree distribution

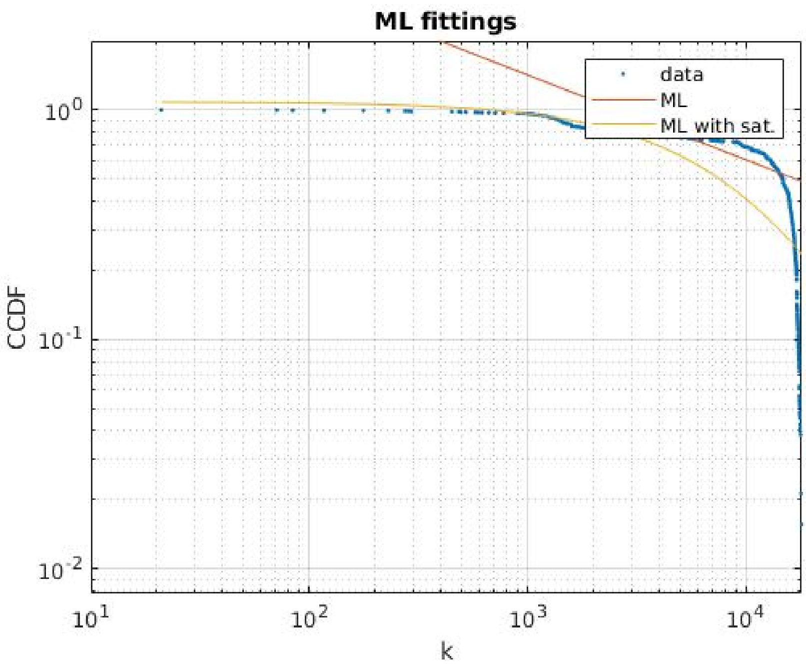


Figure 6: ML fitting

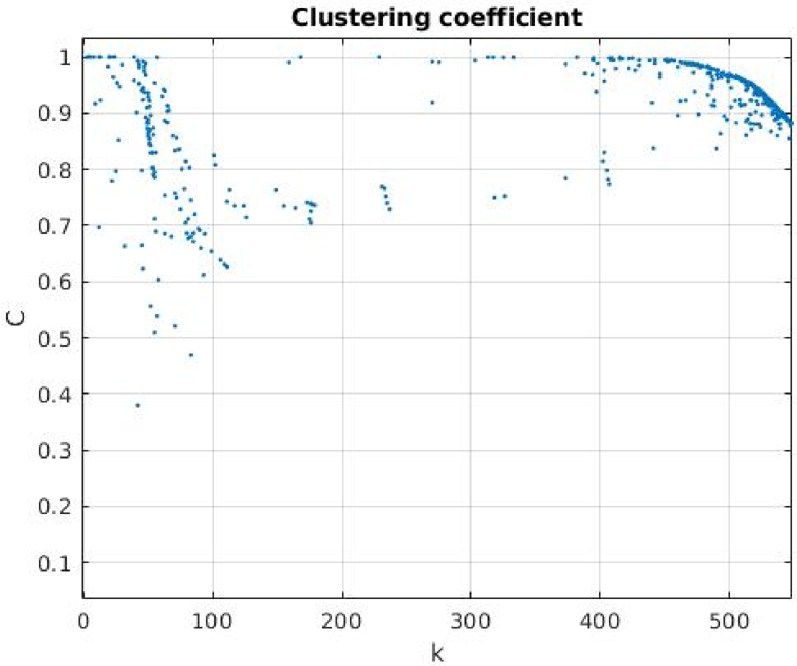


Figure 7: Clustering coefficient

In order to estimate assortativity we plot a graph shown in figure 8a with a fitting that provide an assortativity factor $\gamma = 0.47937$. In figure 8b we can see an enlarged version of the previous one plot with a fit only in the last ones values bigger than k_s. You can see a structural disassortativity behavior with a $\gamma = -0.52621$ but, in general, we can say that the network is an assortativity one.

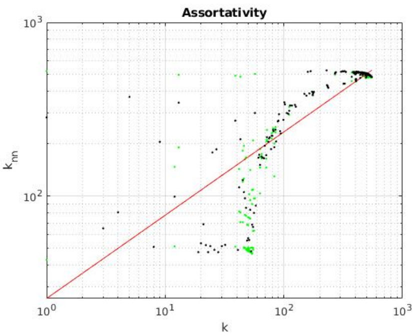
2.5 Robustness

First we estimate the inhomogeneity ratio $\kappa = \langle k^2 \rangle / \langle k \rangle = 216$ and so we compute the breaking point $f_c = 1 - 1/(\kappa - 1) = 0.9953$. In random network it should be $f_{c(rnd)} = 1 - 1/\langle k \rangle = 0.9954$ that means robustness not came from the power law but by the huge average degree. The plot that you can see in figure 9a shown two a random removing of nodes compared to an attack to the node starting with those with highest degree. Instead, figure 9b shows how the network react to a random removal of links.

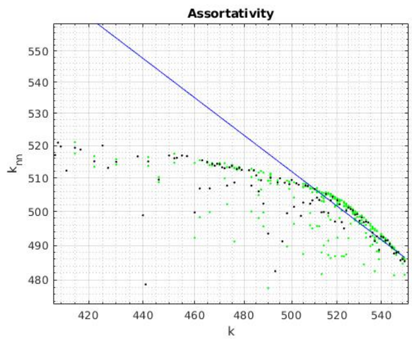
Both the two plot in figure 9 confirms the facts that, due to the fact that the majority of the nodes are hubs, the net is extremely resistant both to random removal and to attacks.

3 Second part

In this second part of this essay we discuss ranking and clustering methods in order to find the better community detection in our case.

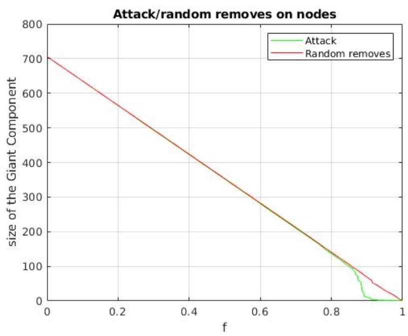


(a) Assortativity plot

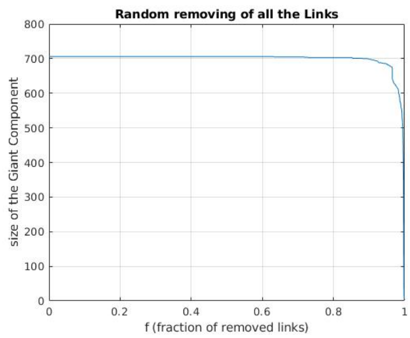


(b) Enlarged around highest values

Figure 8: Assortativity



(a) Removing and attack on nodes



(b) Removing of links

Figure 9: Robustness

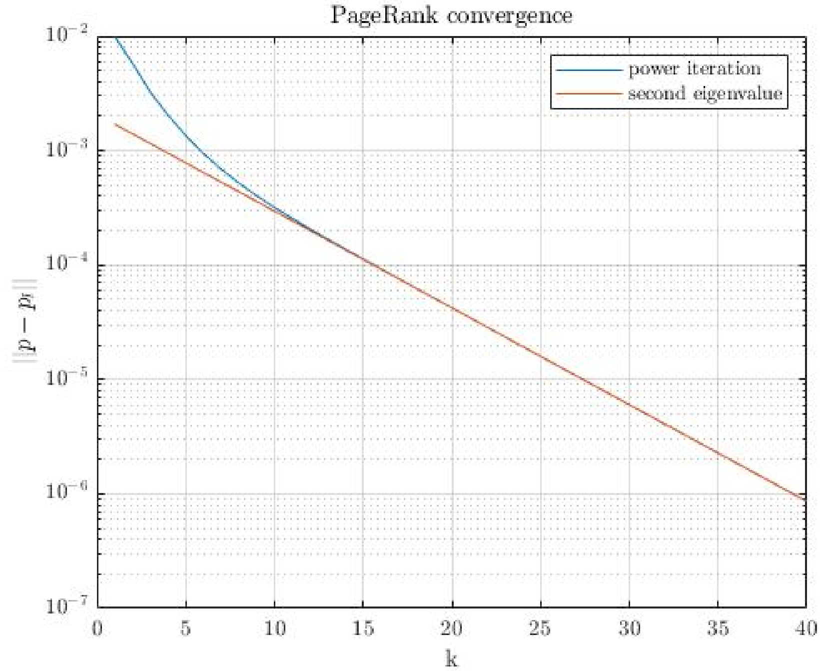


Figure 10: Convergence of power iterations

3.1 PageRank

Firstly we use the PageRank algorithm. We try to solve it both with linear system equation and with power iterations.

The linear system algorithms solved the following equation:

$$\mathbf{r} = [(\mathbf{I} - c\mathbf{M})/(1-c)]^{-1}\mathbf{q} \tag{1}$$

where $\mathbf{M} = \mathbf{A}diag^{-1}(\mathbf{d})$ and $\mathbf{d} = \mathbf{A}^T\mathbf{1}$ is the degree vector. It takes about 0.063404 seconds.

Solving with power iteration take instead about 0.087211 seconds. We try with 40 iterations of the following equations:

$$\mathbf{p}_{t+1} = c\mathbf{M}\mathbf{p}_t + (1-c)\mathbf{q} \tag{2}$$

but, as is shown in figure 10 already after about 13 iterations the results converge to the second eigenvalue.

In figure 11 there is a plot of the eigenvalues of Page Rank.

In table 1 there are the first ten MEPs with highest Page Rank values. Unfortunately, in this network this algorithm wasn't able to catch leaders of the parties as we would have liked.

3.2 HITS

After PageRank we try also another ranking algorithm: HITS. We search for the principal eigenvector of

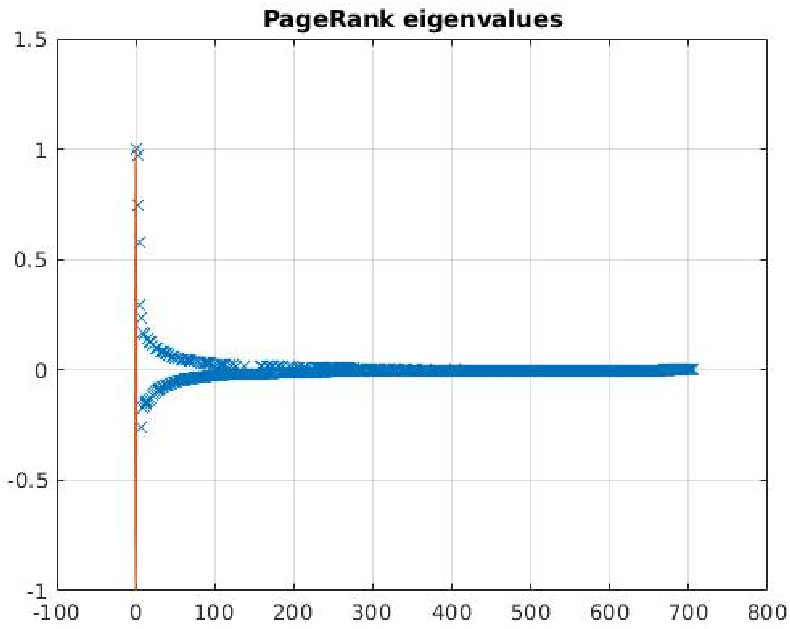


Figure 11: Eigenvalues of Page Rank

MEP	Party	PageRank value
McCLARKIN	EPP	0.0018169
PARKER	EFDD	0.0018014
KOHLÍČEK	GUE-NGL	0.0018004
ROCHEFORT	ALDE	0.0017919
SALAFRANCA SÁNCHEZ-NEYRA	EPP	0.0017748
ŁUKACIJEWSKA	EPP	0.0017745

Table 1: First ones MEPs according to PageRank

MEP	Party	HITS value
HAZEKAMP	GUE-NGL	0.0460868
KYLLÖNEN	GUE-NGL	0.0460868
SELLSTRÖM	EPP	0.0460868
ASHWORTH	EPP	0.0460868
FLACK	ECR	0.0460868
LECHEVALIER	ENF	0.0460525
FERRANDINO	S&D	0.0460469
LÓPEZ AGUILAR	S&D	0.0460408

Table 2: First ones MEPs according to HITS

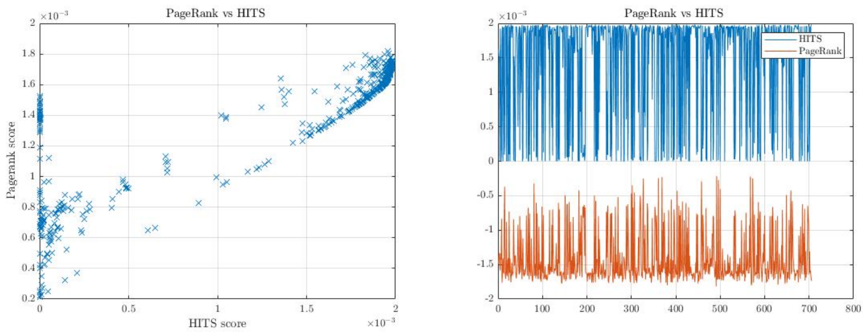


Figure 12: HITS vs PageRank

$$\mathbf{M} = \mathbf{A}^T \cdot \mathbf{A} \tag{3}$$

The computation takes 0.035570 seconds. Results are shown in table 2 but, unfortunately, also with this algorithm we aren't able to catch the leaders of the parties.

A comparison between PageRank and HITS is shown in figure 12.

3.3 Spectral clustering

In order to divide in two communities the net we try to use a Spectral approach. First of all we compute the eigenvalues of the normalized Laplacian:

$$\mathbf{L}_1 = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} \tag{4}$$

and plot them in figure 13.

As we can immediately see the first two eigenvalues in the plot are not so far one from the other:

$$\lambda_0 = 0.0000; \quad \lambda_1 = 0.0316; \quad \lambda_2 = 0.2567$$

so we cannot expect a good division. Instead, the fact that the first eigenvalues is equal to zero is right as we expected.

However first we try the simplest approach using the sign of the *Fiedler's* vector to divide in two communities the net. The results is shown in figure 14.

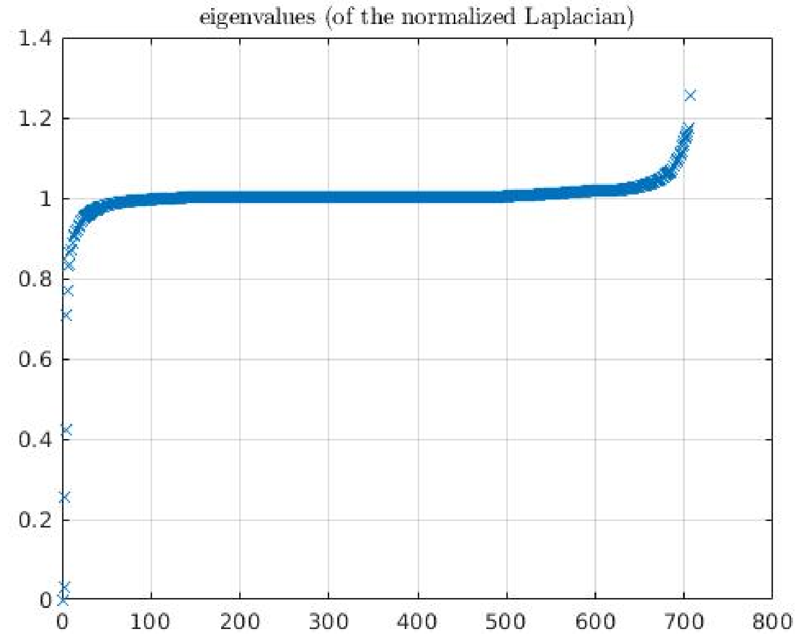


Figure 13: Eigenvalues of Page Rank (smallestabs)

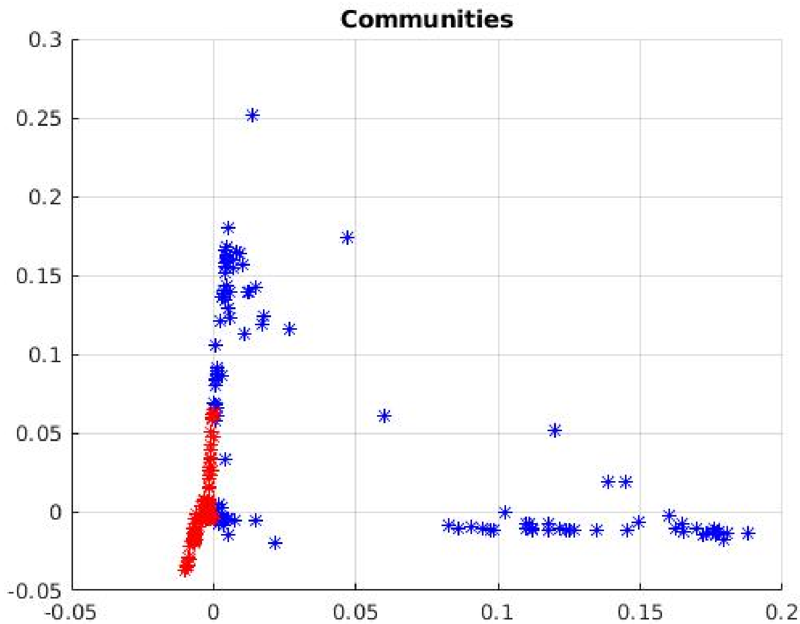
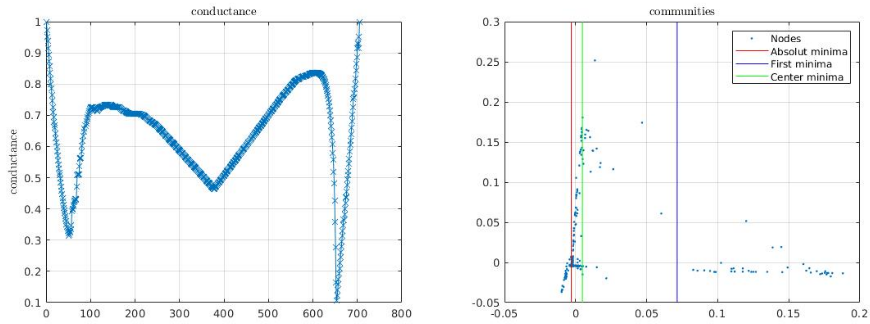


Figure 14: Communities via *Fiedler* vector



(a) Conductance (b) Community with different threshold

Figure 15: Spectral Approch

The result is not so satisfactory and then we try with the Spectral clustering approach from which we get the conductance shown in figure 15a. We can see three different local minima that are three different good point to use as threshold for generate the two communities. If we select the absolute minimum value we get:

- Minimum conductance: 0.10647
- *Cheeger's* upper bound: 0.25157
- Cut value: 255
- Assoc value: 2395
- Community size #1: 654
- Community size #2: 52

With different threshold we get similar values that for brevity we do not report. In figure 15b there are a representation of the three threshold and the relatives communities.

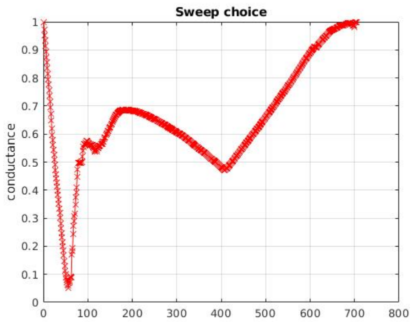
3.4 PageRank Nibble

After the Spectral clustering we try another algorithm: PageRank Nibble. It's purpose is again to divide the network in two different clusters. This algorithm needs a starting point: we choose the best threshold of the Spectral clustering algorithm (see subsection 3.3). Another idea to initialize the teleport vector is to try different MEPs that is important in the political scenario: we tried to do that but the results were really similar to the ones that we present following (using as teleport vector the threshold of Spectral clustering).

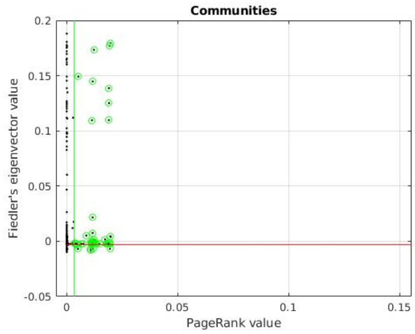
In figure 16a there are a plot of the conductance where you can see the at the beginning a good minimum. We use this to create the two communities plotted in figure 16b.

The results and the parameters of the algorithm are listed below:

- Complexity/D: 21.8193



(a) Sweep



(b) Community

Figure 16: PageRank Nibble Algorithm

- ϵ: 0.001
- Precision: 0.00048876
- Minimum conductance: 0.051089
- Number of links: 138433
- Cut value: 136
- Assoc value: 2662
- Community size #1: 57
- Community size #2: 649

3.5 K-Means

After the algorithms mentioned before we thought that only two communities are not able to catch all the aspects of the network. For instance, we hypothesized that besides one group "pro-EU" and another "against-EU" there might be another group with MEPs that don't have a clear idea (i.e. MEPs without a political group).

In order to do that we decided to use K-Means clustering algorithm with three clusters. Results are plot in figure 17.

Instead, in figure 18 we highlight side by side in the same graph two different thinks: in 18a we highlight nodes based on the cluster they belong. Instead, in 18b, we highlight nodes according on the fact that MEPs are in one of the three big different political groups that we consider: "pro-EU", "against-EU" or "not-attached/not clear idea on this topic". See subsection 4.1 for more information on those. We can immediately see that the results are good and there are a lot of similarities between the two plots.

The dimensions of the three clusters are:

- Cluster #1: 598
- Cluster #2: 51
- Cluster #3: 57

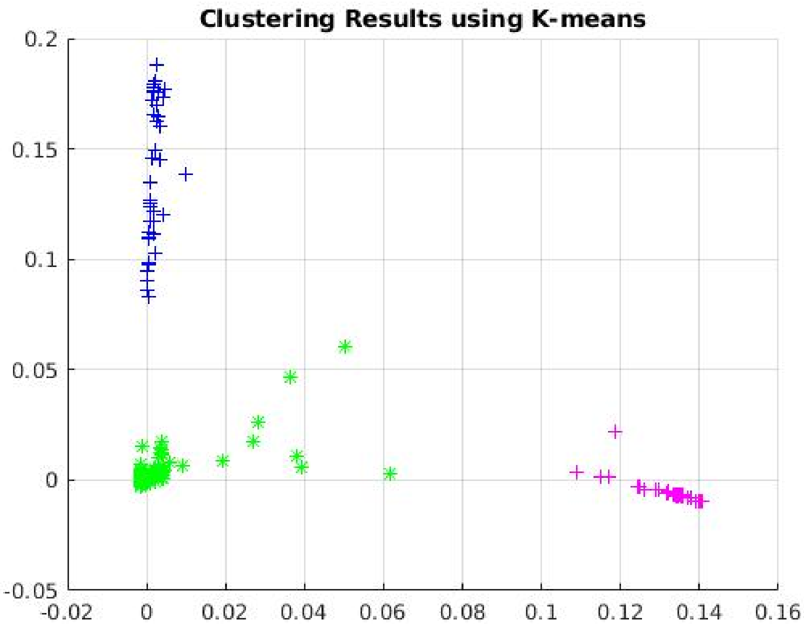
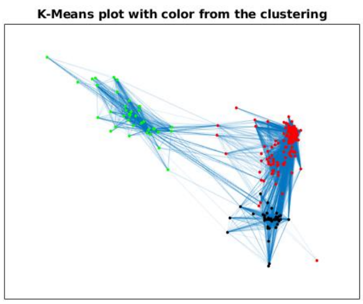
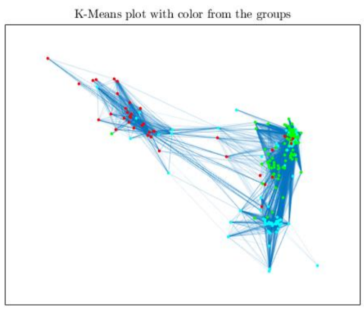


Figure 17: Clustering result of K-Means with three clusters



(a) Graph with color from the clusters



(b) Graph with color from the groups

Figure 18: K-Means in graph

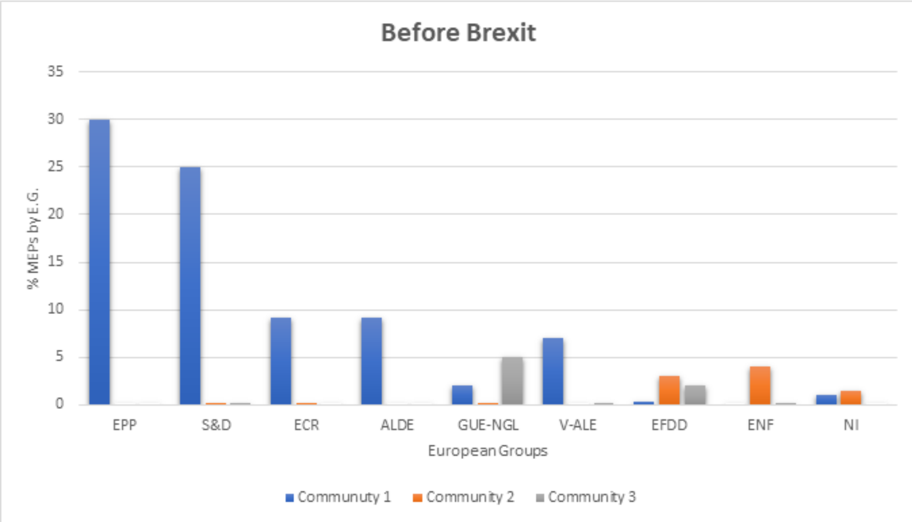


Figure 19: Composition of the groups before Brexit

4 Comparison of the two networks

4.1 The three big groups

In order to compare the network created with votes before Brexit and the one with votes after Brexit we try to create three different big groups where about all the MEPs can fit. It is an hard job, especially in some cases where some parliamentary groups do not have a definite position on the topic.

The *pro-EU* group contains Group of the European People's Party (Christian Democrats) (**EPP**), Group of the Progressive Alliance of Socialists and Democrats in the European Parliament (**S&D**), Group of the Alliance of Liberals and Democrats for Europe (**ALDE**) and European Conservatives and Reformists Group (**ECR**).

The *against-EU* group contains Europe of Freedom and Direct Democracy Group (**EFDD**) and Europe of Nations and Freedom Group (**ENF**).

The *not-attached* group contains the Non-attached Members (**NA**).

Remains two groups, the Confederal Group of the European United Left - Nordic Green Left (**GUE-NGL**), that have not a clear idea on this topic (they are reformist anti-EU) and so we can consider it in the *not-attached* group or in the *against-EU*. The second one is the Group of the Greens/European Free Alliance (**V-ALE**): they are reformist pro-EU and so we can consider it in the *not-attached* group or in the *pro-EU*.

4.2 Differences between the two networks

In the first two histograms (figure 19 and figure 20) you can see the percentage distribution of MEPs in the European Parliament from April to May (before Brexit) and from July to September (after Brexit). Every bar represents one of the three communities. It's immediately visible that community one is much more bigger that the other, as we have discuss in subsection 3.5. This biggest

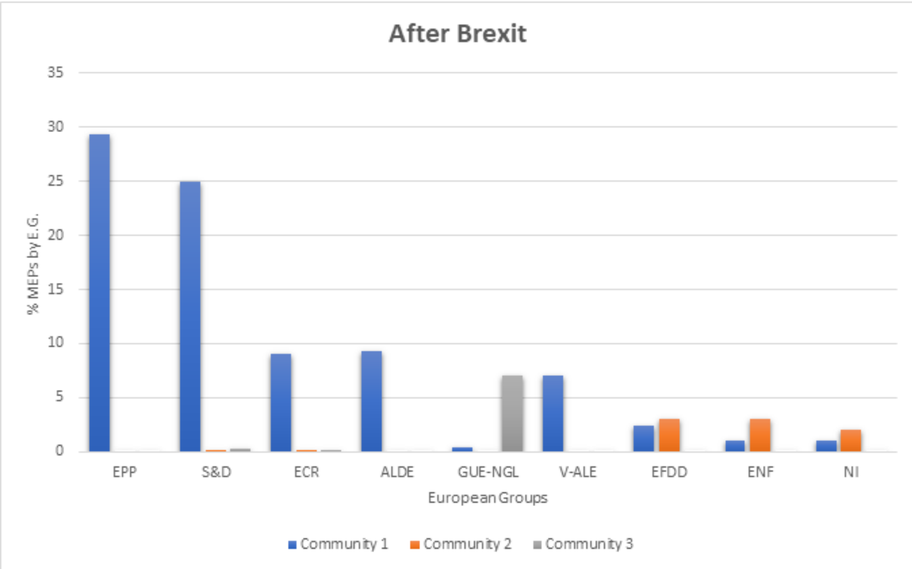


Figure 20: Composition of the groups after Brexit

community is composed predominantly from old formations characterized by positions in general "pro-EU". The two major groups are EPP and S&D (that are also the biggest groups in the European Parliament).

We can see essentially a contraposition between Pro-Europeanists (first community) and Eurosceptics (second and third communities). The components of GUE-NGL, EFDD and ENF are rather ridiculous but are the roots of the second community, right-wing Eurosceptics composed predominantly by EFDD and ENF and the third, left-wing Eurosceptics composed predominantly by GUE-NGL.

Not-attached are distributed equivalently in the Pro-Europeanists community and in the right-wing Eurosceptics community.

By the way, the difference between the first and the second period are more visible in figure 21 and figure 22 where are shown the percentage distribution of the parliamentary groups in the communities.

PPE and ALDE are the only two formations that are only in one of the three communities, the first one. S&D, V-ALDE and ECR instead have a paltry percentage of MEPs from other communities.

The other groups instead have a really different behavior.

GUE-NGL changes from 64% in the third community and 34% in the first to 94% in the third and only 6% in the first. This behavior is probably due to the desire to strengthen its internal cohesion both in Eurosceptic and anti-right.

EFDD is the group that change most from the first to the second period. In the first presents 51% of MEPs in the second community (right-wing Eurosceptic), 40% in the third community (left-wing Eurosceptic) and a poor 9% in the first community. In this period there are a strength political line based on an hard opposition between Pro-Europeanists and Eurosceptic. The second phase will instead see the contrast of a 44% in the first community against a 56% in the second one (and no one in the third). In the first faction there is

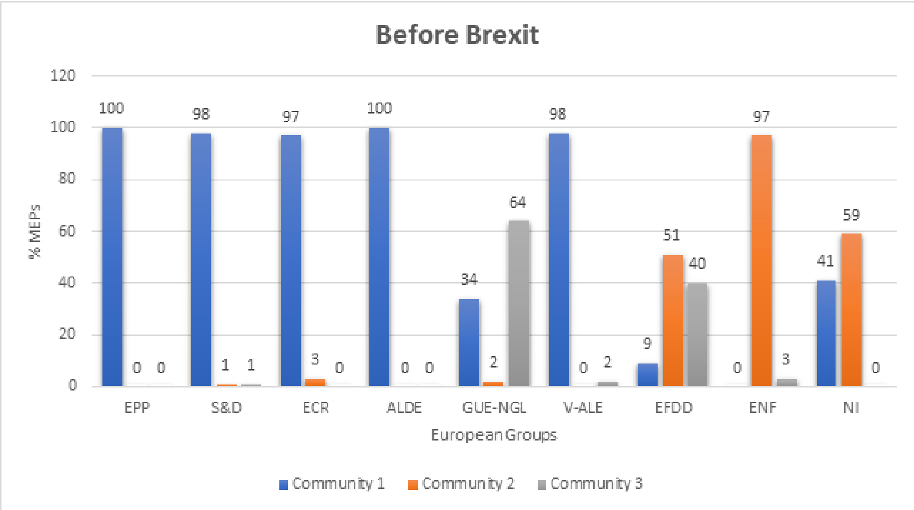


Figure 21: Composition of parliamentary groups before Brexit in percentage

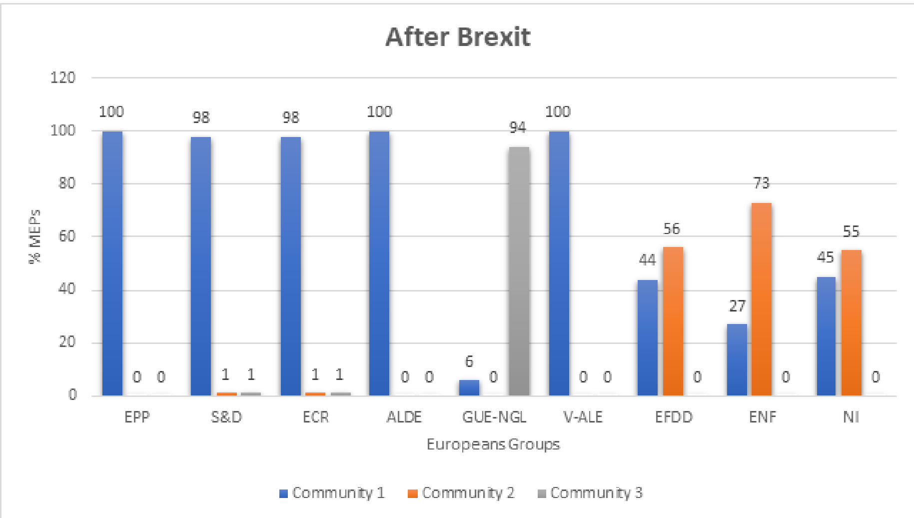


Figure 22: Composition of parliamentary groups after Brexit in percentage

in majority "Movimento 5 Stelle" (M5S) from Italy, while in the second it's predominant "United Kingdom Independence Party" (UKIP). From these two parties, initially allies, there was a break between the two periods.

ENF shows a curios behavior too: if in the first period it was compact on the second community (97%), after Brexit we can see 27% in the first community and 73% faithful to the second one. In this 73% it's big the French component (68%), while the 27% is predominantly composed by Italian "Lega Nord" and, in partly minor, the German "FPO".

Finally, the non-attached group doesn't change significantly in these two periods.

4.3 Clustering of the biggest community

In order to understand a little bit more about the biggest community found as specify in subsection 3.5 (the pro-European one) we decided to use K-Means algorithm also on it. We try different number of clusters but with four we found a division that made sense. Method to do this study is the same as in 3.5.

The dimensions of the four clusters are:

- Community #1: 582 (pre-Brexit) // 552 (post-Brexit)
- Community #2: 31 (pre-Brexit) // 40 (post-Brexit)
- Community #3: 31 (pre-Brexit) // 4 (post-Brexit)
- Community #4: 2 (pre-Brexit) // 2 (post-Brexit)

We found, as is show in figure 23 and in figure 24, a big community (#1) quite stable in the two periods which include MEPs from EPP, S&D, ALDE and V-ALE. We suppose that this big community it's kept compact from a pro-European and moderate political action.

The other three communities are really little and MEPs in them change significantly across the two different periods. From April to May the #2 and the #4 communities are mainly build by ECR MEPs, probably due to a Eurosceptic and nationalist trend of this formation. The #3 is instead mainly compose by members of V-ALE and GUE-NGL: it's the progressive community.

In the second period, from July to September, there is a community that continue to be build almost totally by ECR members, while it appears a third community of only members of EFN: again, this two communities highlight the fraction between Europeist and Eurosceptic.

Note that, on every run of the code, the results can be a little bit different due to the K-Means function, but the general behavior doesn't change too much.

5 Conclusions

In conclusion we can say that the expected results came out from this study: we have been able to divide MEPs in meaningful community based only in our votes. We can say for instance that the majority of the MEPs in general vote following the party instruction and thanks to this we have been able to identify parties from the data. We were also able to see change between before/after Brexit and to explain it in a meaningful way.

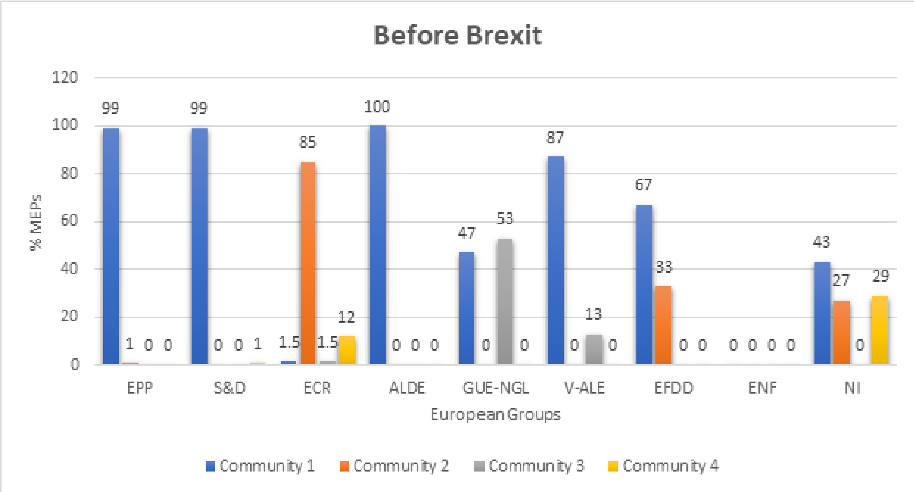


Figure 23: Composition of parliamentary groups before Brexit in percentage inside the biggest community

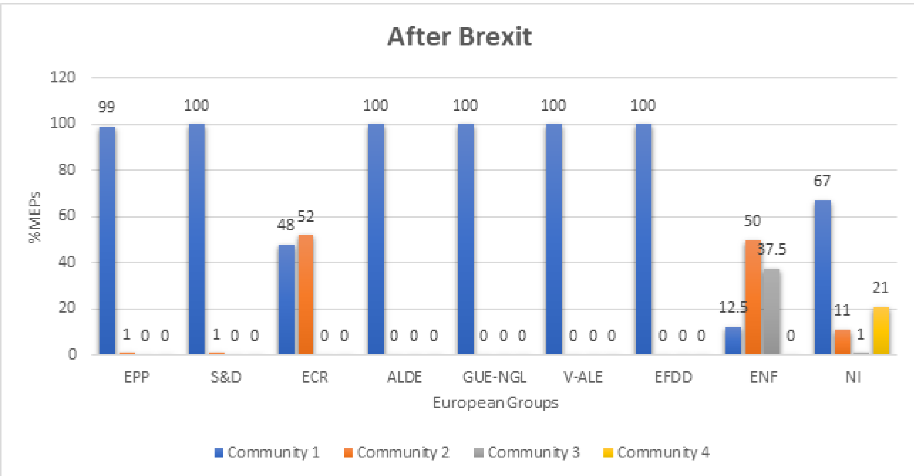


Figure 24: Composition of parliamentary groups after Brexit in percentage inside the biggest community

References

[1] European Parliament website, http://www.europarl.europa.eu, December 2018

[2] Brexit on Wikipedia, https://en.wikipedia.org/wiki/Brexit, December 2018

[3] Network Science by Albert-László Barabási, http://networksciencebook.com/, January 2019