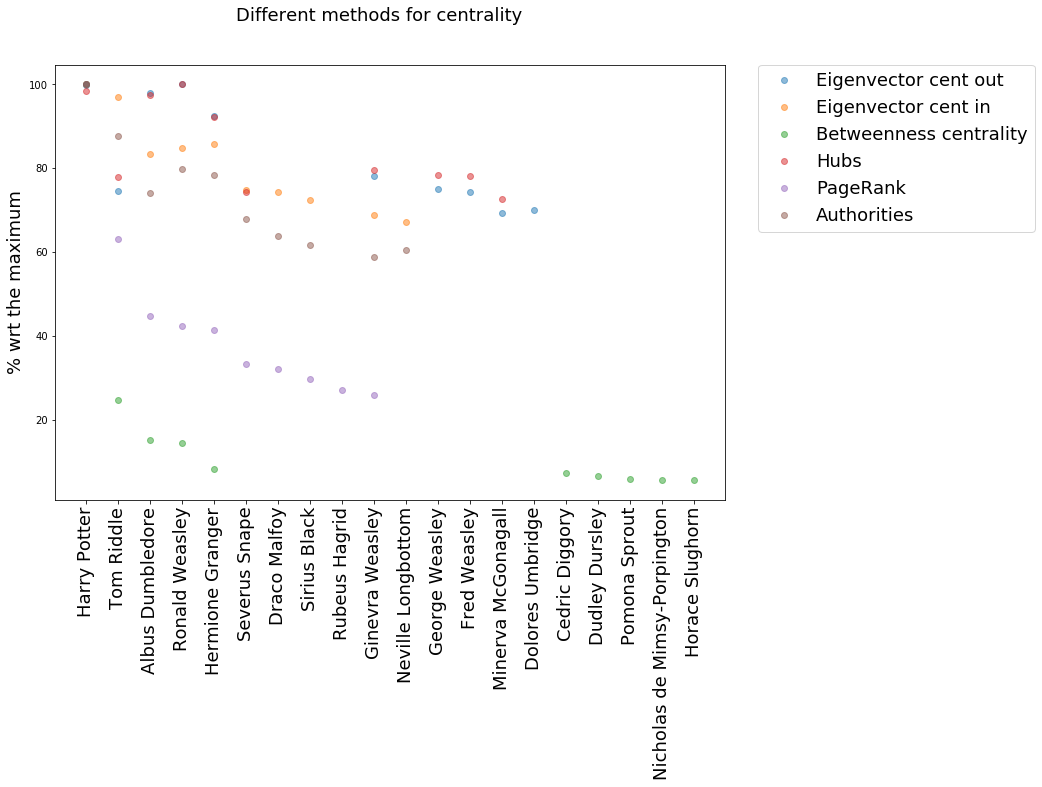
**Statistics of the network**

In this section we are tr ying to find out which characters are the most central, more specifically the most important characters. For this we used different tools such as betweenness centrality, PageRank, eigenvector centrality (for in and out links) and the HITS algorithm (finds Hubs and Authorities). We thus use these different measures to be able to compare them and to see if indeed they capture the most important characters in our network. The tools are rather technical to explain, thus we have explained them in our notebook.



In the above graph one can see the results of the different measures.

Harry Potter nearly always appears on top. He is just distanciated by Ron on the Hubs and eigenvector centrality out. This is because Harry Potter has plenty of links linking to him, thus plenty of people know him. While proportionally Ron has more links linking to other characters (out links). Ron thus knows plenty of other characters, but he is not really well known by the others. This can be due to the fact that he is really close to Harry, this means he will know a lot of people that Harry knows, but then those people will have much more connections to Harry than to him. This happens when your bestfriend has 7 books and movies about him.

Both of the measures (Hubs and eigenvector centrality out) focus on the out links.

We notice that the top 5 characters (Harry Potter, Tom Riddle (Voldemort), Dumbledore, Ron and Hermione) in all the measures are the most important characters in the Harry Potter universe (or at least what we would have said).

The next 9 characters are all still important but more side characters in the Harry Potter universe. Except from Sirius Black they appear in all the books.

Dolores Umbridge is more special. She is the headmaster of Hogwarts in book 5: “Harry Potter and the Order of the Phoenix”. Thus she is very important in book 5 but otherwise she only appears in book 7 for a minor role.

The last 5 characters of the plot represent less famous characters that only appear in one or two Harry Potter book. Some of the names were even not that familiar to us. We can thus take into consideration that betweenness centrality is a poor measure to find central nodes in this graph. In our view the best measure for this graph is PageRank that captures the first 10 characters in the plot.

Finally, we notice a big correlation between the eigenvector out centrality and the Hubs. The same yields for eigenvector in centrality and the Authorities.

**Community analysis**

We will use the Louvain algorithm to analyse the underlying communities in the harry potter network.

We have thus applied this algorithm to 2 different sets of communities. On the one hand we were interested in the **4 different houses of Hogwarts ( Gryffindor, Slytherin, Ravenclaw and Hufflepuff)**, but on the other hand we were also interested in the interaction between the **good and bad guys in Harry Potter.** There is in Harry Potter one community of bad guys: the Death Eaters. Yes the name sounds horrible! But also several communities of good guys such as the Order of the Phoenix and the Army of Dumbledore. It seemed thus interesting to asses these two separated problems.

We do not know what exactly to expect. We could have 2 different scenarios. One scenario could be that characters interact a lot in there own House. Another scenario would be that there is a lot of interaction between the most famous characters regardless of their house or conviction. The solution will probably be a blurry mix of these two scenarios.

Plot of the network houses

Here is a plot of network with only the House members. Red nodes are Gryffindor nodes. Green nodes are Slytherin nodes. Yellow nodes are Hufflepuff nodes. Blue nodes are Ravenclaw nodes.

We notice in the center a higher density of Gryffindor nodes. This makes sense since in Harry Potter, the main house (the house of Harry Potter) is Gryffindor.

We had a problem since not all of our nodes had a house or a good or bad label. We thus only kept the nodes that had such a label and the edges between these nodes. We have plotted our results in a confusion matrix in order to make our results more visual. The confusion matrix is a table to compare two different sets of non-overlapping communities.

A number in the square (with row i and column j), is the number of characters that are in both community corresponding to row i and community corresponding to row j. We will use the confusion matrix to compare the communities that are naturally present in Harry Potter (houses, Good and Bad Guys) versus the communities found by the Louvain algorithm (these have no name and thus are just numbered).

The confusion Matrix of the houses

This matrix tells us that the categories of houses are not so good to distinguish the different communities. The only community that is somewhat better captured is the slytherin house. On the one hand, we see that the community 3 we have a vast majority of Slytherins. This is also the case in community 6. On the other hand we distinguish community 2 which has members of several different houses but only one from Slytherin.

Confusion Matrix of the Good and bad guys

Death eaters versus the Good guys

A first observation here is that there are way more good than bad guys. This is because we took several different lists for the Good Guys :'First Order of the Phoenix, Dumbledore's Army and the Order of the Phoenix allies. While we only had one interesting list for the bad Guys: the Death Eaters.

We see though that the distinction is pretty good! In the first community all but one are members of Dumbledore's Army. The 3rd community captures the vast majority of the Death Eaters. We notice that people who were in the frontline to fight against the death eaters such as Severus Snape, Neville’s parents or Sirius Black end up in this community also. The two remaining community contain other good guys but it is harder to see a deep structure.