# Homework 1 Network Science
# A Storm of Nodes
network analysis of the third book of George R.R. Martin's series

Daniele Mari 1231806

December 2019

## 1 Introduction

"A Song of Ice and Fire" is a famous fantasy book series written by George R. R. Martin with almost two thousands characters named along the books. Thanks to the complexity of the relationships between the characters, it is interesting to build and analyze their social network, since it has enough nodes for quantitative evaluation. The adjacency matrix was built from the third book, "A storm of swords", and the mathematical analysis is based on [1].

## 2 Building the Dataset

The characters are the nodes of the network and they are connected either if they are mentioned within 20 words, or if one of the two nodes is the protagonist of the chapter where the other one is mentioned.

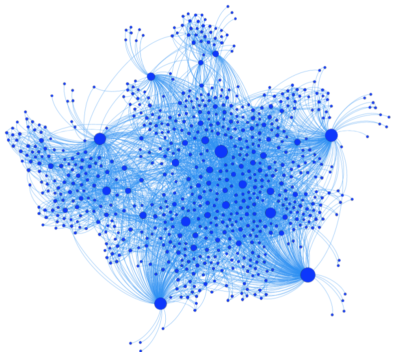
To do so it is necessary to know which words are names and, if so, who they are referring to.

First of all the file was parsed: removed indices, appendix, changed punctuation and so on. After that, in order to find proper names, only words that are not present in the British dictionary and whose first letter was capital were appointed as possible character names; if two of these were found consecutively they were considered as if they belonged to the same character.

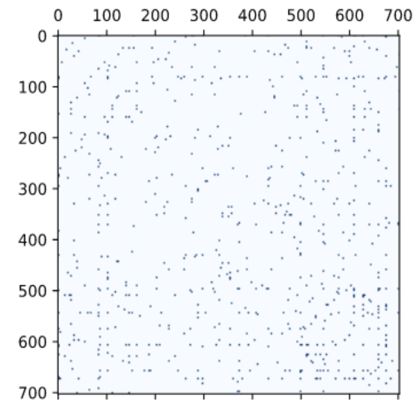
Names like Jon, Snow, Rivers were actually present in the dictionary so they were manually added as exceptions in an extra file.

At this point words that were not characters names (i.e. city names, archaisms, ...) were removed from the list and nicknames were all assigned to their owner with a key value relation. If a name was not univocal, for example surnames, it was not used to generate the network in order to avoid fake links.

At this point since it is possible to find out when and where a certain character is mentioned it is very easy to build the actual network.



(a) Network representation generated with Gephi




(b) Sparse matrix representation of the network

| **Power law** | **Momentums** | **Custering** | **Distance** |
|---|---|---|---|
| $\gamma = 2.45$<br>$C = 73.75$ | $\langle k \rangle = 12.3$<br>$\langle k^2 \rangle = 394.7$<br>$\langle k^3 \rangle = 41757$ | $\langle C \rangle = 0.6646$<br>Expected $\langle C \rangle = 0.061$ | Diameter $= 4$<br>$\langle d \rangle = 2.69$<br>Expected $\langle d \rangle = 1.88$ |
| **Fitness** | **Assortativity** | **Robustness** | |
| $\alpha = 4.42$<br>$\beta = 0.0436$ | $\mu = -0.225$<br>$\mu_{rewire} = 0.0059$ | $f_{rand} = 0.947$<br>$f_{attack} = 0.231$ | |

Table 1: Table containing the main parameters of the network

# 3 Network Analysis

## 3.1 Network overview

One of the easiest ways to represent a graph mathematically is through the adjacency matrix (figure 1b) while a more human readable representation is given in figure 1a.

## 3.2 Network Degree

The degree distribution is the first thing that should be analyzed in order to understand the main properties of the network.

From the plots in figure 2 it is easy to see that the distribution follows the power law $p(k_i) = Ck_i^{-\gamma}$. $\gamma$ can be calculated via linear interpolation (in this case $k_{min} = 20$ was used), since it's value is $\gamma = 2.45$ it means that the network is scale free. This property can also be seen through the higher order momentums, reported in table 1, because they tend to explode as their degree increases. At this point it is easy to find $C = (\gamma - 1)k_{min}^{\gamma-1} = 73.75$

The biggest hub is Tyrion Lannister who is the adviser of the king and he

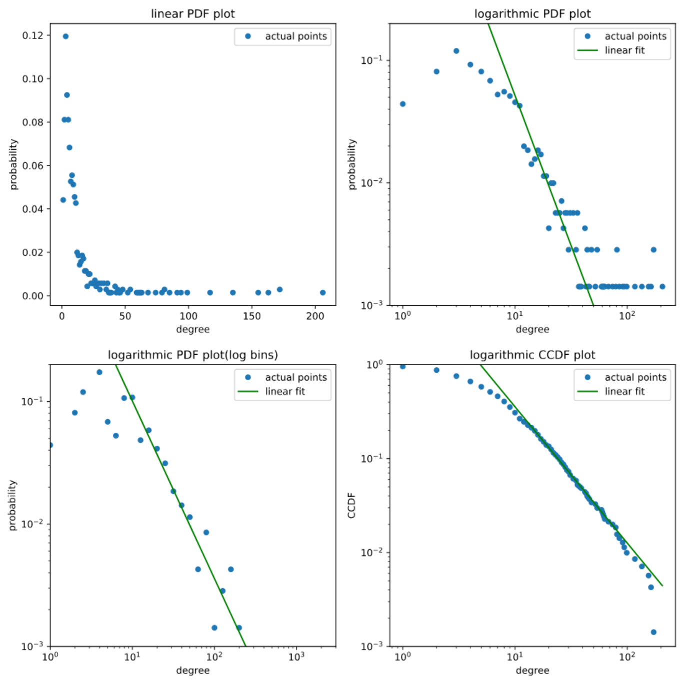



Figure 2: degree distribution of the network

handles the politics of the reign; he also comes from a very important family and suffers from dwarfism. Because of these characteristics many people in the reign know him, this is why he has so many connections.

### 3.3 Small World Property

Since the network is scale free it should have the ultra small world property. By calculating the distance between each node, using the breadth first search algorithm, it is found that the diameter of the network is 4 and the average distance between nodes is 2.695.

This value is a little bit higher than the expected average distance $\langle d \rangle = ln(ln(N)) = 1.88$.

The measure though is not too strange since it is still lower than the expected average distance for networks with $\gamma = 3$, that in this case would be $\langle d \rangle = \frac{ln(N)}{ln(ln(N))} = 3.49$

The distance distribution can be seen in figure 3

### 3.4 Clustering coefficients

In the network a lot of nodes have $C_i = 1$ as it can be seen in figure 4 meaning that all their neighbours are connected.

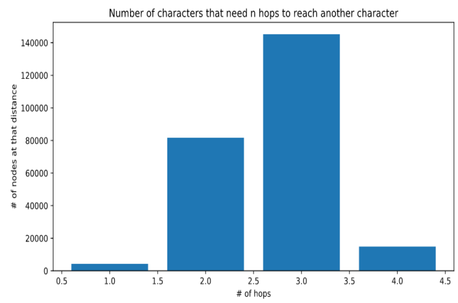



Figure 3: Distance distribution

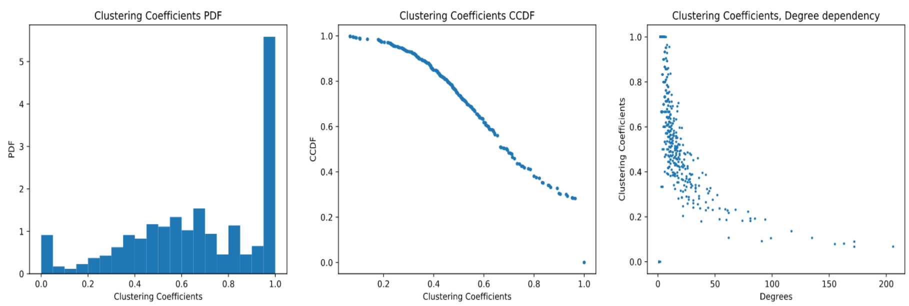



Figure 4: Clustering Distribution

The average clustering coefficient is $\langle C \rangle = 0.6646$ that is ten times higher than the expected clustering coefficient $\langle C_{exp} \rangle = \frac{ln(N)^2}{N} = 0.06113$ meaning that the nodes of the network are much more connected than expected.

By plotting the clustering coefficients against degrees, the relation that appears is that low degree nodes tend to have an higher clustering coefficient than high degree nodes. This is true in general but it might be enforced by how the network was built. As a matter of fact, since characters are connected every time they are mentioned in proximity of one another, it is very likely that there are groups of low degree nodes that are fully connected.

## 3.5 Temporal evolution

In order to study the temporal evolution of the network the chapters were used as a time unit. So for each character a signal $d_i(t)$ was computed and it describes the degree increment with respect to time.

For some of these signals the degree growth suddenly stops like in figure 5. By inspecting the chapters where this happened it is easy to find out that it is when the characters died or shortly after if they had funerals or there was a trial for the murderers. In some other less likely cases the character might have just disappeared from the story.

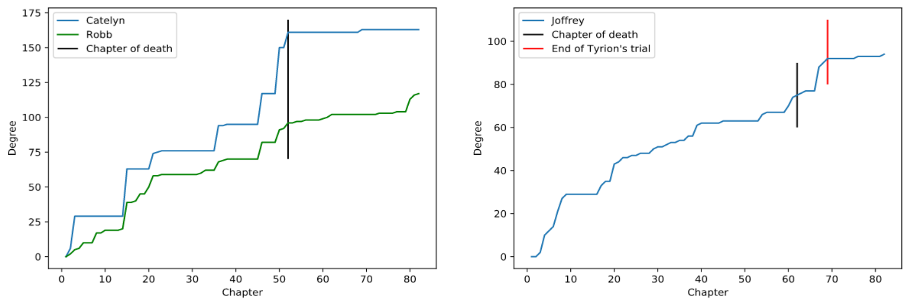



Figure 5: Degree evolution over time of Catelyn and Robb Stark that died in the same chapter and degree evolution of Joffrey Baratheon

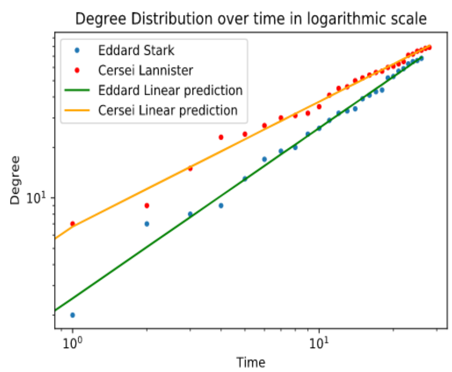



Figure 6: Temporal evolution of the degrees of Eddard Stark and Cersei Lannister in logarithmic scale

The values of the signals though are very often stationary because the character is not mentioned for some chapters. In order to remove these idle times $d_i(t)$ was transformed in $d_{i,cleaned}(t)$ by removing all the entrances where $d_i(t) = d_i(t-1)$.

By inspecting figure 6 it is clear that the degree of the node grows as a power law with time, but the exponents are different for each character so fitness based preferential attachment can be assumed to be a meaningful model for the time evolution of the network.

Because of that the degree of character $i$ at time $t$ can be written as $d_{i,cleaned}(t) = at^{\frac{\eta_i}{C}}$, where $a$ is a constant, $\eta_i$ is the fitness value of node $i$ and $C$ is the normalization constant previously found by analyzing the degree distribution.

The histogram of the fitness values in figure 7 looks like a Gamma Distribution $\Gamma(\alpha, \beta)$. Under this assumption the two parameters can be calculated with a curve fitting function yeilding $\alpha = 4.426$ and $\beta = 0.0436$.

It is interesting to see that the hubs are not the characters with the highest fitness, but they have more connections just because they had more time to

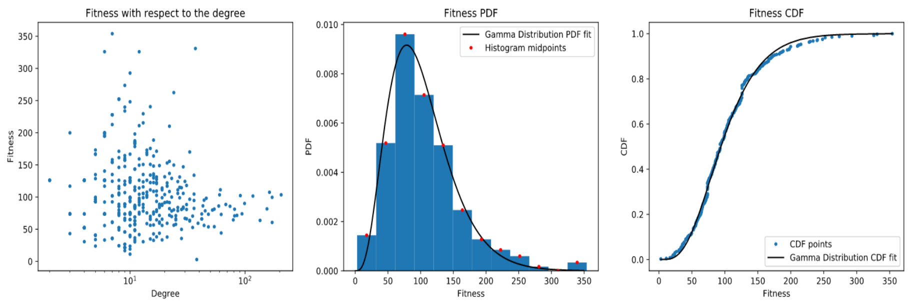
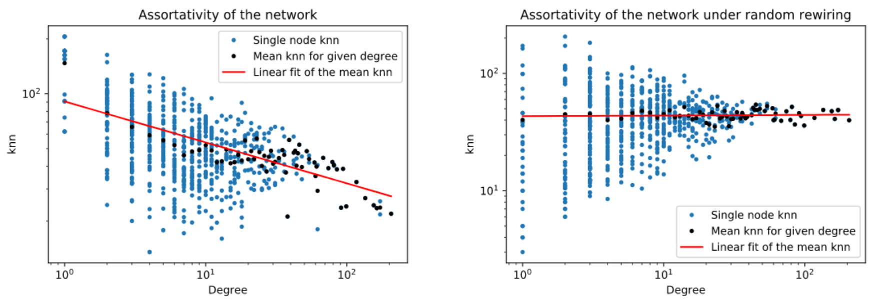



Figure 7: Distribution of the fitness values found




(a) (b)

Figure 8: knn distributions without (a) and with (b) random rewiring

form new connections.

### 3.6 Assortativity

By inspecting the nearest neighbours degree it is possible to see that the network is Disassortative, this is particularly clear in figure 8a.

As a matter of fact by supposing that $k_{nn}(k_i) = k_i^{\mu}$, for this network it holds that $\mu = -0.2248$ which is negative.

In order to check if the this is a structural or natural property, random rewiring was applied obtaining the knn distribution in figure 8b. Since $\mu_{rewire} = -0.030$ it means that the network has natural Disassortativity.

As expected of a scale free network the natural cutoff $k_{max} = k_{min} N^{\frac{1}{\gamma-1}} = 1375$ is greater than the structural cutoff $k_s = (2L)^{\frac{1}{2}} = 93$. The structural cutoff is located just before the distribution stops following the power law as shown in figure 9.

The natural disassortativity of this network is probably due to the fact that most of the hubs live in different places. Because of that they tend to be very connected with low degree nodes in their area and less connected to other hubs that they usually never meet.

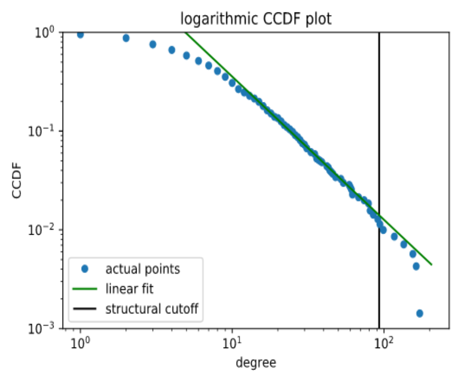



Figure 9: Structural cutoff of the network

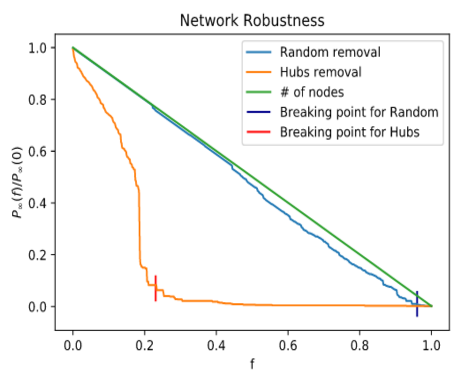



Figure 10: Robustness of the network under random and attack node removal

### 3.7 Robustness

In order to analyze the robustness of the network two approaches were followed:

- Random node removal: Nodes are removed at random, the network is really robust to this kind of node failures since the breaking point is $f_{rand} = 0.925$ which is very high.
- Hubs removal: Hubs are removed first, the network breaks really early, in this case $f_{attack} = 0.23$.

As expected from a scale free network it is very robust to random node removal but it breaks very easily if hubs are removed, this is very clear in figure 10.

In order to calculate the breaking points the inhomogeneity ratio $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}$ was used, in particular when $\kappa < 2$ it means that a giant component can't be identified anymore.

Since in the book there are adversarial factions it might be more interesting to study the robustness of the communities in the network in order to see which one is weaker.

## 4 Conclusions

The network is scale free, this was expected since lords and kings are obviously hubs as many people know them, meaning that they can get high degrees.

The clustering coefficient analysis shows that many small nodes are almost fully connected, probably forming small communities, because if they live in the same city it is very likely that they know each other.

Finally from the robustness of the network it is seen how killing lords tends to isolate their servants from the network, this quickly leads to the giant component disappearing. Common people are usually just connected with their countrymen because most times only the lord has the resources to travel and meet other people.

## References

[1] Albert-László Barabási et al. *Network science.* Cambridge university press, 2016.