Athens University of Economics and Business Dept. of Informatics
M.Sc. Program in Business Analytics
Course: Social Network Analysis Spring 2019
Instructor: Katia Papakonstantinopoulou

Network Analysis and Visualization with R and igraph

A Song of Ice and Fire

Assignment done by: Koh Miao Ling

*P.S: I have never watched any episode of Games of Thrones

Martin's A Game of Thrones 5-Book Boxed Set (Song of Ice and Fire Series) has many characters mentioned in passing and major characters meet each other only occasionally. With this social network analysis assignment, I have researched and I find it interesting to see how various characters are connected and who the important people in the series are. As it turns out that network analysis can be used to explore relationships in social networks, who is connected with the important people and what the key players that connect clusters of people.

I was given the dataset that has an undirected weighted information between any 2 source and target combinations. These weighted edges can be used to refer to the strength of relationship between each pair of characters. I found out that the Eddard-Stark and Robert-Baratheon is the strongest weighted edge (334) among the many other characters. These 2 characters are close as brothers and have created bonds of friendship which helps to build their houses later in their lives.

Using the dataset given, I use the graph_from_data_frame to create igraph graph for the network of characters of “A Song of Ice and Fire”. The entire network have 796 vertices (nodes) and 2823 edges among them.

```
> g <- graph_from_data_frame(d=edges, directed=FALS
E)
> print(g, e=TRUE, v=TRUE)

IGRAPH 27c0b3a UNW- 796 2823 --
+ attr: name (v/c), weight (e/n)
+ edges from 27c0b3a (vertex names):
 [1] Addam-Marbrand--Brynden-Tully     Addam-Marbra
nd--Cersei-Lannister
 [3] Addam-Marbrand--Gyles-Rosby       Addam-Marbra
nd--Jaime-Lannister
 [5] Addam-Marbrand--Jalabhar-Xho      Addam-Marbra
nd--Joffrey-Baratheon
 [7] Addam-Marbrand--Kevan-Lannister   Addam-Marbra
nd--Lyle-Crakehall
 [9] Addam-Marbrand--Oberyn-Martell    Addam-Marbra
nd--Tyrion-Lannister
[11] Addam-Marbrand--Tywin-Lannister   Addam-Marbra
nd--Varys
+ ... omitted several edges
```

2. Network Properties

Considering we have an undirected weight graph above, where vertices represents the characters and the edges represent the (bilateral) connections among them. I will look into the network properties of the data to understand further into their connections.

We counted the number of vertices, edges and diameter of the graph. Also we measure the diameter and the degrees (weighted and non-weighted) of the characters. As mentioned earlier, the graph has 796 vertices (nodes) and 2823 edges among them.

And here a "triangle" is a set of three vertices that are mutually adjacent in the graph[1]. Number of triangles are can be used to calculate the measure of degree to which nodes in a graph tend to cluster together. It means that friends of my friend tend to be my friend. We found a total of 16965 triangles in this undirected network.

As for the diameter of a network, it is the distance between the two most distant nodes in the network. In undirected network as mentioned above, it's the connected components linked to each other by following links. The diameter of this network is 53 ignoring their directions which mean that they weakly connected components.

The notion of centrality can be measured differently. It can be simply be found by relying on their degree which mean the in-degree and out-degree. Since we have a undirected network, we will calculate using "total".

As for weighted degree, I used a function "strength" in R to calculate the "total" degree. Results can be found below.

The top-10 characters of the network as far as their degree is concerned are as follows:

```
> top_10_degree
 Tyrion-Lannister          Jon-Snow   Jaime-Lannister  Cersei-Lannister
              122               114               101                97
Stannis-Baratheon        Arya-Stark     Catelyn-Stark       Sansa-Stark
               89                84                75                75
     Eddard-Stark        Robb-Stark
               74                74
```

The top-10 characters of the network as far as their weighted degree is concerned are as follows:

```
> top_10_weighted_degree
  Tyrion-Lannister           Jon-Snow   Cersei-Lannister
              2873               2757               2232
 Joffrey-Baratheon       Eddard-Stark Daenerys-Targaryen
              1762               1649               1608
   Jaime-Lannister        Sansa-Stark         Bran-Stark
              1569               1547               1508
  Robert-Baratheon
              1488
```

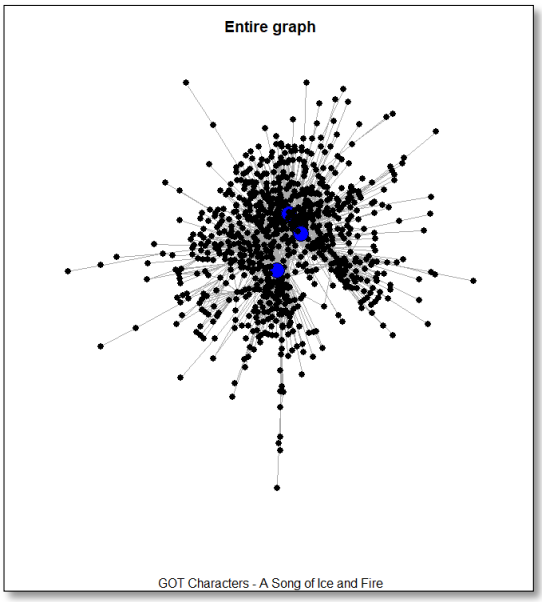
[1] http://theory.stanford.edu/~tim/s14/l/l1.pdf

3. Subgraph

3 blue dots shown in the network graph below are namely Tyrion-Lannister, Jon Snow, and Jaime Lannister.

I have plotted the entire network graph by setting parameters of the graphs first. The colours for vertices are set to black and those nodes with more than 100 connections are set to blue. I have also tuned the size of those nodes with more connections to make it more visible and stand out among the others.

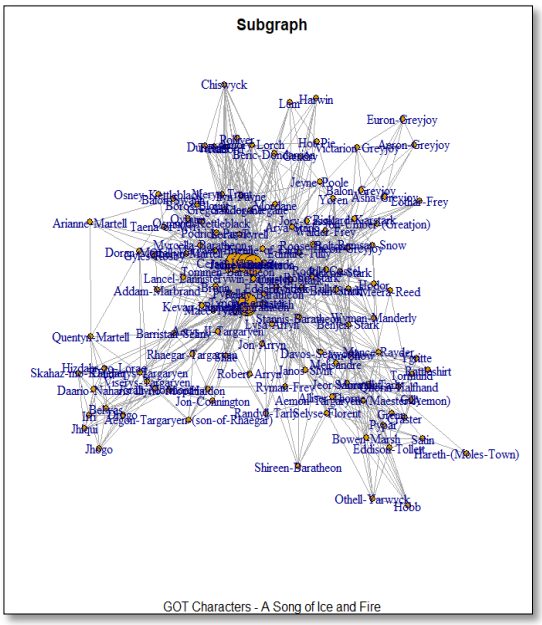
Next, I measure the density of a network and it gives us a ready index of the degree of dyadic connection in a population. When the data is undirected, density is calculated relative to the number of unique pairs ((n*n-1)/2)2. The entire graph density is 0.0089 and subgraph density is 0.12. The subgraph is more compact than entire network which means that they have more average strength of ties across the possible ties.



However, the density is more compact for the subgraph because I have removed those with connections less than 10.

One character in subgraph is more likely to connect with more people in the system as compared rest across the entire network.

```
> # Entire graph
> edge_density(g)
[1] 0.008921968
```




```
> # Subnetwork
> edge_density(subgraph, loops = FALSE)
[1] 0.1258612
```

[2] http://faculty.ucr.edu/~hanneman/nettext/C8_Embedding.html

4. Centrality

Closeness centrality in the network is calculated as the reciprocal of the sum of the length (distance) of the shortest path between the nodes which means that who can have more useful contacts to reach out to even more people. In real life, this person is very useful as it will be helpful to reach others who acts as intermediator.

I calculated this by using the "closeness" function in R and the top 15 useful characters as follows.

Jon Snow is ranked 10th place. Top intermediator is Jaime Lannister and his role is a king of the kingdom with many important connections. The rest of the characters with highest closeness are all surrounded by central characters that connect various storylines and houses in Game of Thrones.

```
> closeness_top_15
  Jaime-Lannister  Robert-Baratheon     Theon-Greyjoy       Jory-Cassel
     0.0001193460      0.0001137527      0.0001135203      0.0001131734
Stannis-Baratheon   Tywin-Lannister  Cersei-Lannister  Tyrion-Lannister
     0.0001131606      0.0001128286      0.0001116695      0.0001114454
 Brienne-of-Tarth           Jon-Snow Joffrey-Baratheon     Rodrik-Cassel
     0.0001112718      0.0001106072      0.0001093733      0.0001083658
    Doran-Martell      Eddard-Stark       Harys-Swyft
     0.0001079098      0.0001073192      0.0001072961
```

Betweeness centrality describes the number of shortest paths between nodes. Nodes with high betweenness centrality are on the path between many other nodes, i.e. they are people who are key connections or bridges between different groups of nodes. In a social network, these nodes would be very important because they are likely to pass on information to a wide reach of people.

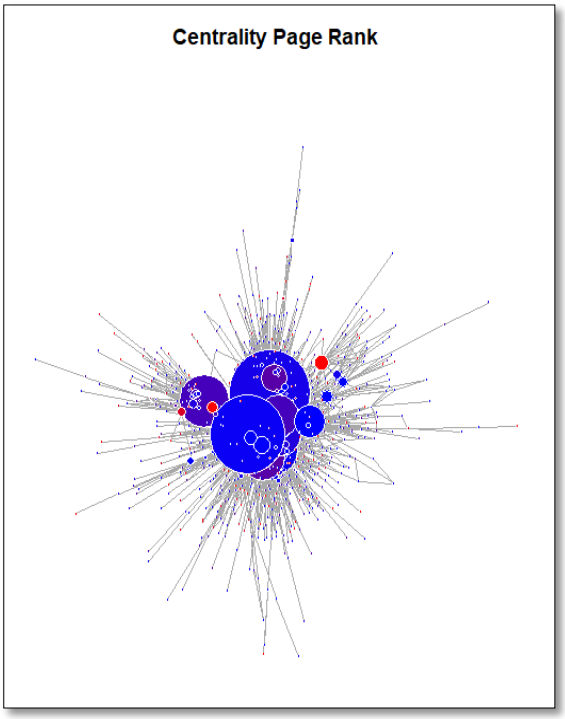
Jon Snow is at the top for betweeness centrality.

```
> betweeness_top_15
          Jon-Snow      Theon-Greyjoy    Jaime-Lannister
          41698.94          38904.51          36856.35
Daenerys-Targaryen  Stannis-Baratheon   Robert-Baratheon
          29728.50          29325.18          29201.60
  Tyrion-Lannister   Cersei-Lannister    Tywin-Lannister
          28917.83          24409.67          20067.94
        Robb-Stark         Arya-Stark    Barristan-Selmy
          19870.45          19354.54          17769.29
      Eddard-Stark       Sansa-Stark   Brienne-of-Tarth
          17555.36          15913.44          15614.41
```

5. Ranking and Visualization

PageRank is the most popular measure of importance because it considers nodes as more important if they have many incoming edges (or links). Similarly, eigenvector centrality scores nodes in a network according to the number of connections to high-degree nodes they have as well. So this shows that John Snow has many connections to multiple high degree nodes.

Unfortunately I have not watched the show and google gave me another new list of top 40 important characters. But I've heard Jon Snow is quite a popular guy but I don't know the reason. And the mother of dragon is in the 4th place.




```
> cent_top_10
                  name  page
_rank
1            Jon-Snow 0.035
70539
2    Tyrion-Lannister 0.032
91094
3    Cersei-Lannister 0.023
66461
4  Daenerys-Targaryen 0.022
28040
5     Jaime-Lannister 0.019
79001
6        Eddard-Stark 0.018
96426
7          Arya-Stark 0.018
57171
8   Stannis-Baratheon 0.018
05099
9   Joffrey-Baratheon 0.017
46037
10         Robb-Stark 0.017
36071
```