SOCIAL NETWORK ANALYSIS

GROUP MEMBERS

RUTURAJ, NITESH, MUHAMMAD USMAN, MANJUNAGARAJ, ALI

GAME OF THRONES – SOCIAL NETWORK ANALYSIS

Executive Summary:

This report presents social network analysis of most popular TV show and book series Game of thrones. We start by introduction to dataset, followed by calculating centrality measures and interpreting results. Next, we run community detection algorithms to assess what kind of information we can gather. Lastly, we explore the possibility of a linear relationship between centrality and community information. Since we are aware of different characters in game of thrones it becomes feasible for us to cross check network measure results and confirm their relevance.

Dataset:

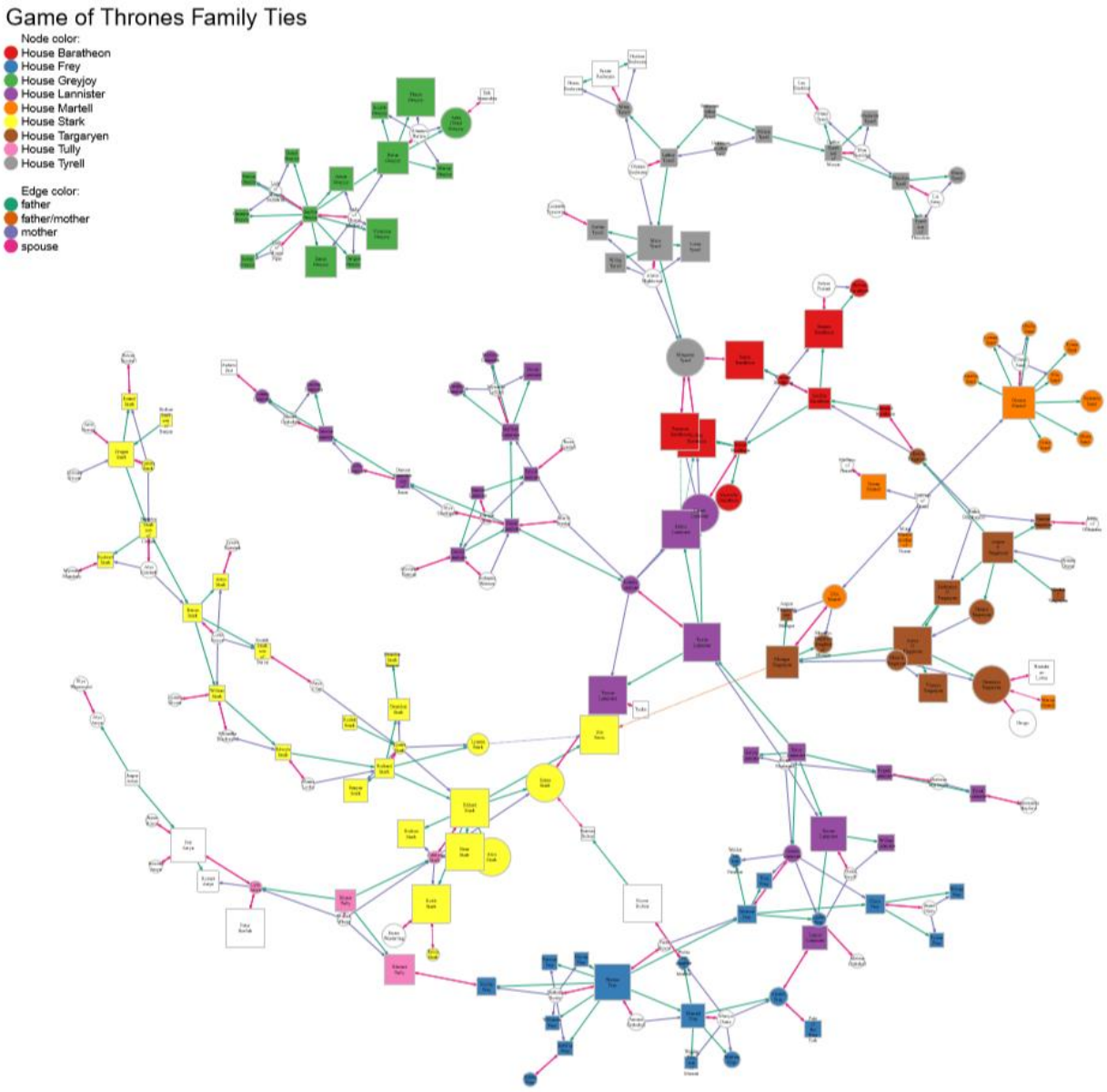
We used data from Kaggle Game of thrones datasets, the link is provided in references section. We have a big network comprising of 208 nodes and 404 edges. A network graph is constructed from the edge and node-table. An edge-table contains source and target nodes in the first two columns and optionally additional columns with edge attributes. We have type of interaction (mother, father or spouse), the color and line-type we want to assign to each edge.

```
> head(union_edges)
                source            target   type   color   lty
1         Lysa Arryn      Robert Arryn mother #7570B3 solid
2       Jasper Arryn        Alys Arryn father #1B9E77 solid
3       Jasper Arryn         Jon Arryn father #1B9E77 solid
4          Jon Arryn      Robert Arryn father #1B9E77 solid
110 Cersei Lannister  Tommen Baratheon mother #7570B3 solid
210 Cersei Lannister Joffrey Baratheon mother #7570B3 solid
```

The node table has one row for each character that is either a source or a target in the edge table. We chose following attributes from original Kaggle dataset: gender/male (male = 1, female = 0), house (as the house each character was born into) and popularity. House2 was meant to assign a color to only the major houses. Shape represents the gender.

```
> head(union_characters)
          name male culture          house popularity      house2   color  shape
1    Alys Arryn    0    <NA>    House Arryn 0.08026756        <NA>    <NA> circle
2 Elys Waynwood    0    <NA> House Waynwood 0.07023411        <NA>    <NA> circle
3  Jasper Arryn    1    <NA>    House Arryn 0.04347826        <NA>    <NA> square
4   Jeyne Royce    0    <NA>    House Royce 0.00000000        <NA>    <NA> circle
5     Jon Arryn    1 Valemen    House Arryn 0.83612040        <NA>    <NA> square
6    Lysa Arryn    0    <NA>    House Tully 0.00000000 House Tully #F781BF circle
```

Graphical Representation:



Node color shows the major houses, node size shows the character's popularity and node shape represents their gender (square for male, circle for female). Edge color shows interaction type.

What we can see right away is that there are only limited connections between houses and that the Greyjoys (green color network above) is the only house that has no ties to any of the others.

Centrality Measures – Network Analysis:

Network science helps us identify the important vertices. A person can play a central role in multiple ways. For example, he / she could be well connected, be centrally located, or be uniquely positioned to help disperse information or influence others in the storyline.

We consider a character as “important” if he has connections to many other characters. There are a few network properties, that tell us more about this. For this, we are considering the network as undirected to account for parent/child relationships as being mutual.

```
> #Entire Network
> centr_degree(union_graph_undir, mode = "total")$centralization
[1] 0.04282795
> centr_clo(union_graph_undir, mode = "total")$centralization
[1] 0.01414082
Warning message:
In centr_clo(union_graph_undir, mode = "total") :
  At centrality.c:2784 :closeness centrality is not well-defined for disconnected graphs
> centr_eigen(union_graph_undir, directed = FALSE)$centralization
[1] 0.8787532
```

Node Degree Centrality:

Node degree or degree centrality describes how central a node is in the network (i.e. how many in- and outgoing edges it has or to how many other nodes it is directly connected via one edge). We can calculate the number of out- or in-going edges of each node, here we consider the sum of both.

Results:

	rowname	degree	degree_std
1	Quellon Greyjoy	12	0.05797101
2	Walder Frey	10	0.04830918
3	Oberyn Martell	10	0.04830918
4	Eddard Stark	9	0.04347826
5	Catelyn Stark	8	0.03864734
6	Emmon Frey	7	0.03381643
7	Genna Lannister	7	0.03381643
8	Merrett Frey	7	0.03381643
9	Balon Greyjoy	7	0.03381643
10	Jason Lannister	7	0.03381643

In this case, the node degree seems to reflect how many offspring and spouses a character had. With many wives and several children, Quellom Greyjoy, the grandfather of Theon and Asha/Yara leads the list.

Closeness Centrality:

In a connected graph, closeness centrality of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the *closer* it is to all other nodes. Its normalized form represents the average length of the shortest paths instead of their sum. It is given by

$$C(x) = \frac{N}{\sum_y d(y,x)}.$$

Where d(y, x) is the distance between vertices x and y. N is the number of nodes in the graph.

Results:

	rowname	closeness	closeness_std
1	Sansa Stark	0.0002013288	9.726028e-07
2	Tyrion Lannister	0.0002012882	9.724070e-07
3	Tywin Lannister	0.0002011668	9.718201e-07
4	Joanna Lannister	0.0002005616	9.688965e-07
5	Eddard Stark	0.0002002804	9.675381e-07
6	Catelyn Stark	0.0001986492	9.596579e-07
7	Cersei Lannister	0.0001984915	9.588960e-07
8	Jaime Lannister	0.0001975894	9.545382e-07
9	Jeyne Marbrand	0.0001966568	9.500330e-07
10	Tytos Lannister	0.0001966568	9.500330e-07

The characters with highest closeness are all central characters that connect various storylines and houses in Game of Thrones. Here Sana and Tyrion have a slight edge over others, and this is true in the TV show as well.

Betweenness centrality:

The betweenness centrality of a vertex measures how frequently that vertex lies on short paths between other pairs of vertices. Mathematically, the betweenness zi of vertex i is

$$\sum_{j,k \in V} \frac{\sigma_{jk}(i)}{\sigma_{jk}},$$

Where σ_{jk} is the number of (j, k)- shortest paths and $\sigma_{jk}(i)$ is the number of these (j,k)-shortest paths that go through vertex i. A vertex that appears on many short paths is a broker of information in the network: Efficient communication between different parts of the network will frequently pass through such a vertex. They are key connections or bridges between different groups of nodes.

Node betweenness Results:

	rowname	betweenness	betweenness_std
1	Eddard Stark	6926.864	0.3248846
2	Sansa Stark	6165.667	0.2891828
3	Tyrion Lannister	5617.482	0.2634718
4	Tywin Lannister	5070.395	0.2378123
5	Joanna Lannister	4737.524	0.2221999
6	Rhaegar Targaryen	4301.583	0.2017533
7	Margaery Tyrell	4016.417	0.1883784
8	Jon Snow	3558.884	0.1669192
9	Mace Tyrell	3392.500	0.1591154
10	Jason Lannister	3068.500	0.1439191

Edge Betweenness:

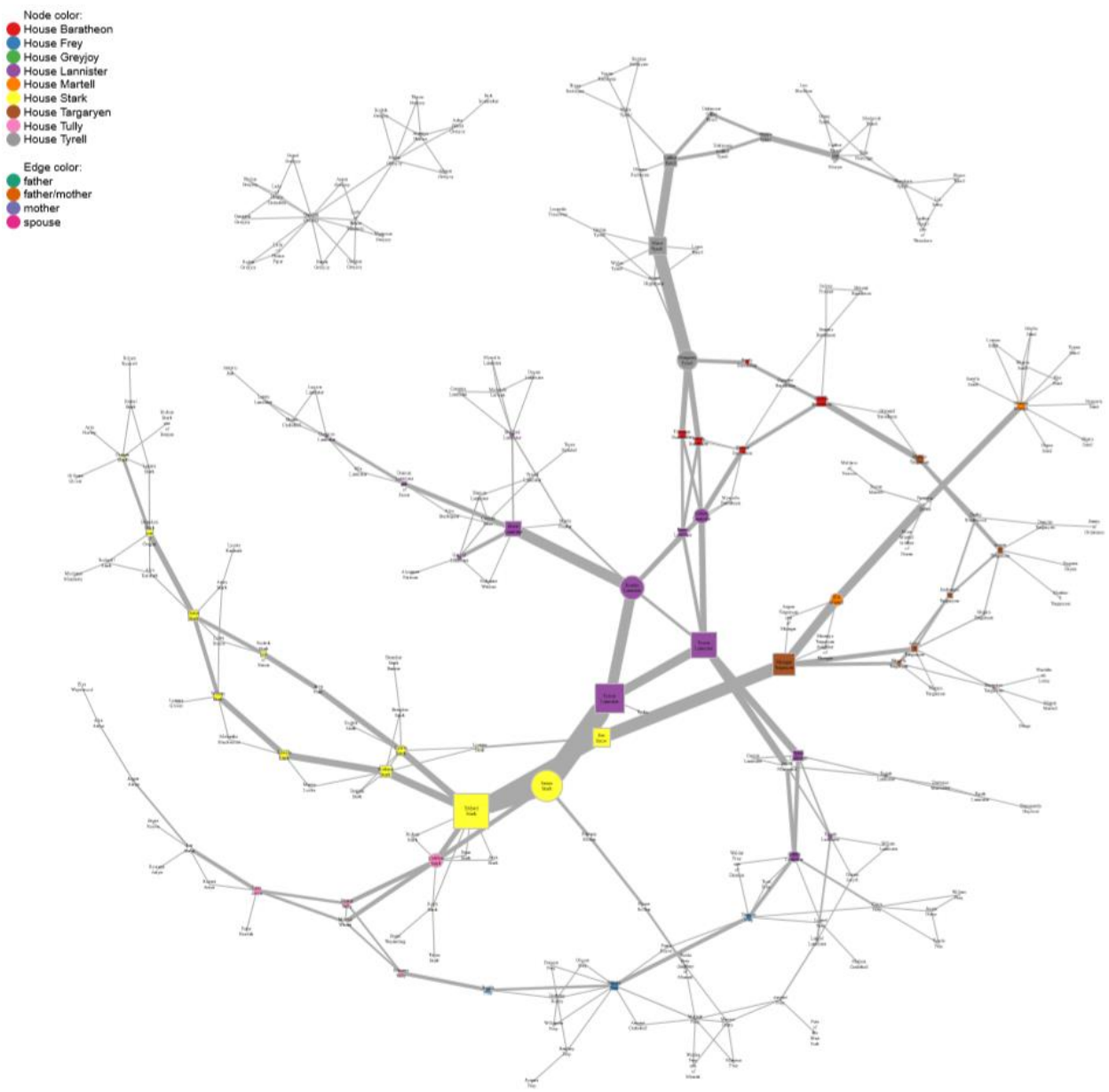
The edge betweenness centrality is defined as the number of the shortest paths that go through an edge in a graph or network. Each edge in the network can be associated with an edge betweenness centrality value. An edge with a high edge betweenness centrality score represents a bridge-like connector between two parts of a network, and the removal of which may affect the communication between many pairs of nodes through the shortest paths between them.

Edge Betweenness Results:

	rowname	edge	betweenness
1	160	Sansa Stark|Tyrion Lannister	5604.149
2	207	Sansa Stark|Eddard Stark	4709.852
3	212	Rhaegar Targaryen|Jon Snow	3560.083
4	296	Margaery Tyrell|Mace Tyrell	3465.000
5	213	Eddard Stark|Jon Snow	3163.048
6	131	Jason Lannister|Joanna Lannister	3089.500
7	159	Joanna Lannister|Tyrion Lannister	2983.591
8	171	Tyrion Lannister|Tywin Lannister	2647.224
9	192	Elia Martell|Rhaegar Targaryen	2580.000
10	300	Luthor Tyrell|Mace Tyrell	2565.000

Graphical Representation:

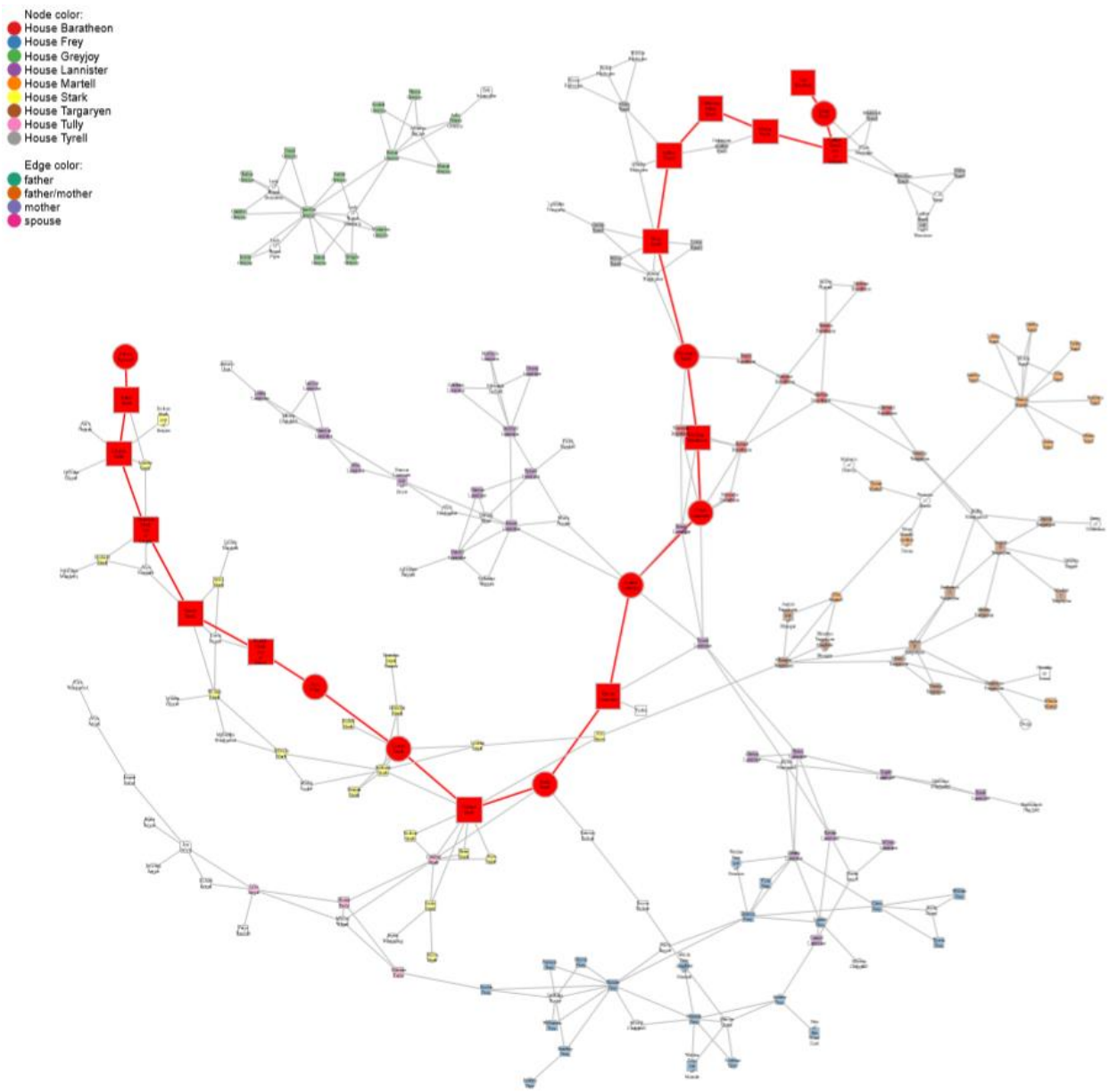
We can now plot by feeding the node betweenness as vertex.size and edge betweenness as edge.width to plot function.



Eddad Stark is the character with highest betweenness. This makes sense, as he and his children (specifically Sansa and her arranged marriage to Tyrion) connect to other houses and are the central points from which the story unfolds. In edge betweenness Sansa and Tyrion's connection is the most important. This measure accurately represents the storyline.

Diameter:

In contrast to the shortest path between two nodes, we can also calculate the longest path, or diameter. The longest path connects 21 nodes in our network.



Eigenvector centrality

In graph theory, eigenvector centrality is a measure of the influence of a node in a network. Relative scores are assigned to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. A high eigenvector score means that a node is connected to many nodes who themselves have high scores.

For a given graph G := (V,E) with |V| vertices, let A = ($a_{v,t}$) be the adjacency matrix. The relative centrality, x , score of vertex, v , can be defined as:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

Where M(v) is a set of neighbors of v and lamda is a constant. this can be rewritten in vector notation as the eigenvector equation:

$$\mathbf{Ax} = \lambda\mathbf{x}$$

We calculate the eigenvalues and eigenvectors of the adjacency matrix. The eigenvector with the highest eigenvalue scores those vertices highly, that have many edges or that are connected to vertices with many edges.

	name	eigenvector
1	Quellon Greyjoy	-0.6625628
2	Balon Greyjoy	-0.3864950
3	Lady of House Sunderly	-0.3312814
4	Alannys Harlaw	-0.2760678
5	Lady of House Stonetree	-0.2208543
6	Asha (Yara) Greyjoy	-0.1656407
7	Robin Greyjoy	-0.1104271
8	Euron Greyjoy	-0.1104271
9	Urrigon Greyjoy	-0.1104271
10	Victarion Greyjoy	-0.1104271

Because of their highly connected family ties (i.e. there are only a handful of connections, but they are almost all triangles), the Greyjoys have been scored with the highest eigenvalues.

Eigenvector Centrality Results:

	name	eigen_centrality
1	Tywin Lannister	1.0000000
2	Cersei Lannister	0.9168980
3	Joanna Lannister	0.8358122
4	Jeyne Marbrand	0.8190076
5	Tytos Lannister	0.8190076
6	Genna Lannister	0.7788376
7	Jaime Lannister	0.7642870
8	Robert Baratheon	0.7087042
9	Emmon Frey	0.6538709
10	Walder Frey	0.6516021

Here we see members of Lannister family score highly and this has to do with them as current rulers of the westeros world and their connections throughout the seven kingdoms.

PageRank Centrality:

PageRank was the founding idea behind Brin and Page's Google search engine. Each vertex has an inherent importance $\beta \geq 0$, along with an importance acquired from its neighbors. Unlike eigenvector centrality, a vertex does not get full credit for the total importance of its neighbors. Instead, that neighbor's importance is divided equally among its direct connections. In other words, a vertex of very high degree passes along only a small fraction of its importance to each neighbor.

The PageRank y_i of vertex i is given by :

$$y_i = \alpha \sum_{j \in V} \frac{w_{ji}}{k_j} y_j + \beta,$$

where $\alpha + \beta = 1$ and $\alpha, \beta \geq 0$.

PageRank Centrality Results:

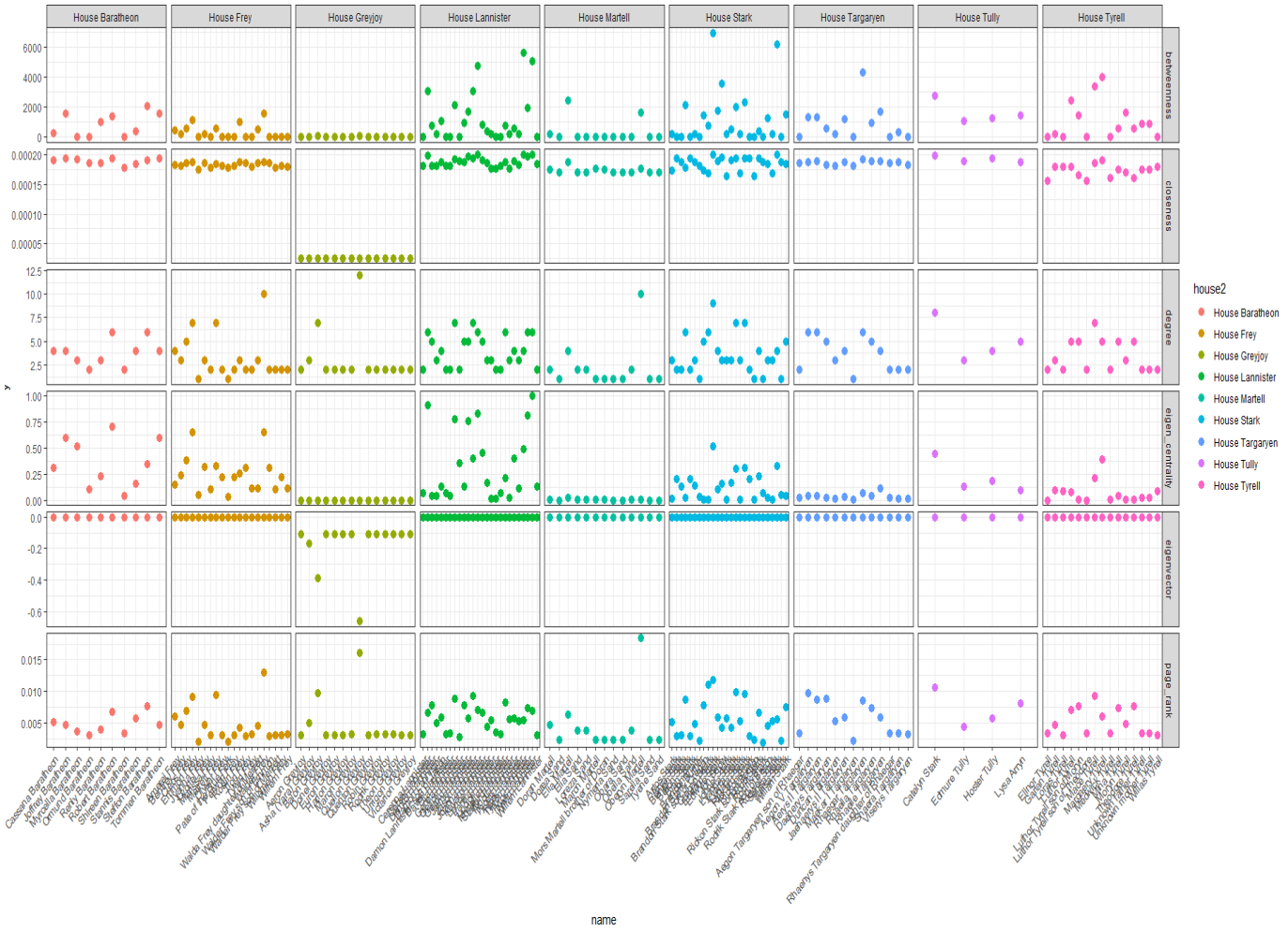
	name	page_rank
1	Oberyn Martell	0.018402407
2	Quellon Greyjoy	0.016128129
3	Walder Frey	0.012956029
4	Eddard Stark	0.011725019
5	Cregan Stark	0.010983561
6	Catelyn Stark	0.010555473
7	Lyarra Stark	0.009876629
8	Aegon V Targaryen	0.009688458
9	Balon Greyjoy	0.009647049
10	Jon Arryn	0.009623742

Oberyn Martell, Quellon Greyjoy and Walder Frey all have the highest number of spouses, children and grandchildren. Therefore, they scored highest for PageRank.

Summary of Centrality Results:

Each measure has confirmed relevance of different nodes and they all make sense from the story point of view. We can now compare all the node-level information to decide which characters are the most important in Game of Thrones.

Let's look at all characters from the major houses:



When comparing across different centrality measures, we can say that House Stark (specifically Eddard and Sansa) and House Lannister (especially Tyrion) are the most important family connections in Game of Thrones.

Community Detection:

At the start we saw game of thrones world divided into 9 major houses. But other than family ties we have communities within the storyline that are related together by participating in same events. Next we look at different community detection methods to see how different characters are divided into communities. Below we share results from different methods before finally choosing louvain method that gives us 13 communities that make the most sense.

Network communities:

Networks often have different clusters or communities of nodes that are more densely connected to each other than to the rest of the network. Let's cover some of the different existing methods to identify these communities.

The most straightforward way to partition a network is into **connected components**. Each component is a group of nodes that are connected to each other, but *not* to the rest of the nodes. We look at giant connected component that includes most nodes.

```
> giant
[[1]]
IGRAPH 06a42bd DN-- 189 368 --
+ attr: name (v/c), male (v/n), culture (v/c), house (v/c), popularity (v/n), house2 (v/c), color (v/c), shape (v/c), type (e/c), color (e/c),
+ edges from 06a42bd (vertex names):
 [1] Lysa Arryn       ->Robert Arryn                Jasper Arryn     ->Alys Arryn                  Jasper Arryn     ->Jon Arryn
 [4] Jon Arryn        ->Robert Arryn                Cersei Lannister ->Tommen Baratheon            Cersei Lannister ->Joffrey Baratheon
 [7] Cassana Baratheon->Stannis Baratheon           Cersei Lannister ->Myrcella Baratheon          Selyse Florent   ->Shireen Baratheon
[10] Cassana Baratheon->Renly Baratheon             Rhaelle Targaryen->Steffon Baratheon           Cassana Baratheon->Robert Baratheon
[13] Robert Baratheon ->Tommen Baratheon            Robert Baratheon ->Joffrey Baratheon           Steffon Baratheon->Stannis Baratheon
[16] Robert Baratheon ->Myrcella Baratheon          Stannis Baratheon->Shireen Baratheon           Steffon Baratheon->Renly Baratheon
[19] Ormund Baratheon ->Steffon Baratheon           Steffon Baratheon->Robert Baratheon            Jaime Lannister  ->Tommen Baratheon
[22] Jaime Lannister  ->Joffrey Baratheon           Jaime Lannister  ->Myrcella Baratheon           Mariya Darry     ->Walda Frey daughte
+ ... omitted several edges

[[2]]
IGRAPH 06a42bd DN-- 19 36 --
+ attr: name (v/c), male (v/n), culture (v/c), house (v/c), popularity (v/n), house2 (v/c), color (v/c), shape (v/c), type (e/c), color (e/c),
+ edges from 06a42bd (vertex names):
 [1] Alannys Harlaw         ->Asha (Yara) Greyjoy Alannys Harlaw         ->Balon Greyjoy       Alannys Harlaw         ->Maron Greyjoy
 [4] Alannys Harlaw         ->Rodrik Greyjoy      Alannys Harlaw         ->Theon Greyjoy       Asha (Yara) Greyjoy    ->Erik Ironmaker
 [7] Balon Greyjoy          ->Alannys Harlaw      Balon Greyjoy          ->Asha (Yara) Greyjoy Balon Greyjoy          ->Maron Greyjoy
[10] Balon Greyjoy          ->Rodrik Greyjoy      Balon Greyjoy          ->Theon Greyjoy       Erik Ironmaker         ->Asha (Yara) Greyjoy
[13] Lady of House Piper    ->Quellon Greyjoy     Lady of House Piper    ->Robin Greyjoy       Lady of House Stonetree->Donel Greyjoy
[16] Lady of House Stonetree->Harlon Greyjoy      Lady of House Stonetree->Quellon Greyjoy     Lady of House Stonetree->Quenton Greyjoy
[19] Lady of House Sunderly ->Aeron Greyjoy       Lady of House Sunderly ->Balon Greyjoy       Lady of House Sunderly ->Euron Greyjoy
[22] Lady of House Sunderly ->Quellon Greyjoy     Lady of House Sunderly ->Urrigon Greyjoy     Lady of House Sunderly ->Victarion Greyjoy
+ ... omitted several edges
```

As expected, we can see two giant components, recall that house Greyjoy only had connections within their family and hence we see it as a separate component of nodes.

Even within a giant component, there can be different subsets of the network that are more connected to each other than to the rest of the network. The goal of **community detection algorithms** is to identify these subsets.

There are a few different algorithms, each following a different logic.

The **walktrap** algorithm finds communities through a series of short random walks. The idea is that these random walks tend to stay within the same community. The length of these random walks is 4 edges by default, but you may want to experiment with different values (longer random walks will lead to fewer communities). The goal of this algorithm is to identify the partition that maximizes a modularity score. Our highest Modularity score is 0.85 with 15 groups.

```
> cluster_walktrap(g)
IGRAPH clustering walktrap, groups: 19, mod: 0.83
+ groups:
  $`1`
   [1] "Aegon Targaryen son of Rhaegar"
   [2] "Elia Martell"
   [3] "Rhaegar Targaryen"
   [4] "Rhaenys Targaryen daughter of Rhaegar"
   [5] "Aegon V Targaryen"
   [6] "Aerys II Targaryen"
   [7] "Betha Blackwood"
   [8] "Daenerys Targaryen"
   [9] "Drogo"
  + ... omitted several groups/vertices


> cluster_walktrap(g, steps=10)
IGRAPH clustering walktrap, groups: 16, mod: 0.84
+ groups:
  $`1`
   [1] "Doran Martell"                "Dorea Sand"
   [3] "Elia Sand"                    "Ellaria Sand"
   [5] "Loreza Sand"                  "Mellario of Norvos"
   [7] "Mors Martell brother of Doran" "Nymeria Sand"
   [9] "Obara Sand"                   "Obella Sand"
  [11] "Oberyn Martell"               "Princess of Dorne"
  [13] "Sarella Sand"                 "Tyene Sand"

  $`2`
  + ... omitted several groups/vertices
```

```
> cluster_walktrap(g, steps=20)
IGRAPH clustering walktrap, groups: 16, mod: 0.84
+ groups:
  $`1`
   [1] "Doran Martell"                "Dorea Sand"
   [3] "Elia Sand"                    "Ellaria Sand"
   [5] "Loreza Sand"                  "Mellario of Norvos"
   [7] "Mors Martell brother of Doran" "Nymeria Sand"
   [9] "Obara Sand"                   "Obella Sand"
  [11] "Oberyn Martell"               "Princess of Dorne"
  [13] "Sarella Sand"                 "Tyene Sand"

  $`2`
  + ... omitted several groups/vertices
```

```
> cluster_walktrap(g, steps=30)
IGRAPH clustering walktrap, groups: 14, mod: 0.85
+ groups:
  $`1`
   [1] "Alys Stackspear"    "Alysanne Farman"    "Cerenna Lannister"
   [4] "Cerissa Brax"       "Damon Lannister"    "Daven Lannister"
   [7] "Gerold Lannister"   "Jason Lannister"    "Marla Prester"
  [10] "Myranda Lefford"    "Myrielle Lannister" "Rohanne Webber"
  [13] "Stafford Lannister" "Teora Kyndall"      "Tybolt Lannister"

  $`2`
   [1] "Cleos Frey"              "Dorna Swyft"
   [3] "Emmon Frey"              "Genna Lannister"
  + ... omitted several groups/vertices
```

The **infomap** method attempts to map the flow of information in a network, and the different clusters in which information may get remain for longer periods. Like walk trap, but not necessarily maximizing modularity, but rather the so-called “map equation”.

```
> cluster_infomap(union_graph_undir)
IGRAPH clustering infomap, groups: 27, mod: 0.8
+ groups:
  $`1`
   [1] "Aeron Greyjoy"          "Donel Greyjoy"          "Euron Greyjoy"          "Harlon Greyjoy"         "Lady of House Piper"    "Lady of House Stonetree"
   [7] "Lady of House Sunderly" "Quellon Greyjoy"        "Quenton Greyjoy"        "Robin Greyjoy"          "Urrigon Greyjoy"        "Victarion Greyjoy"

  $`2`
   [1] "Aegon V Targaryen"      "Aerys II Targaryen"     "Betha Blackwood"        "Duncan Targaryen"       "Dyanna Dayne"           "Jaehaerys II Targaryen"
   [7] "Jenny of Oldstones"     "Maekar I Targaryen"     "Rhaella Targaryen"      "Shaera Targaryen"       "Viserys Targaryen"

  $`3`
   [1] "Cersei Lannister"   "Jaime Lannister"    "Joffrey Baratheon"  "Myrcella Baratheon" "Robert Baratheon"   "Tommen Baratheon"   "Joanna Lannister"   "Tyrion
Lannister"
  + ... omitted several groups/vertices
```

The **edge-betweenness** method iteratively removes edges with high betweenness, with the idea that they are likely to connect different parts of the network. Here betweenness (gatekeeping potential) applies to edges, but the intuition is the same.

```
> cluster_edge_betweenness(union_graph_undir)
IGRAPH clustering edge betweenness, groups: 12, mod: 0.84
+ groups:
  $`1`
   [1] "Alys Arryn"       "Elys Waynwood"    "Jasper Arryn"     "Jeyne Royce"      "Jon Arryn"        "Lysa Arryn"       "Robert Arryn"     "Rowena Arryn"
   [9] "Edmure Tully"     "Sansa Stark"      "Arya Stark"       "Bran Stark"       "Catelyn Stark"    "Eddard Stark"     "Jeyne Westerling" "Rickon Stark"
  [17] "Robb Stark"       "Talisa Stark"     "Hoster Tully"     "Minisa Whent"     "Petyr Baelish"

  $`2`
   [1] "Cassana Baratheon" "Cersei Lannister"  "Jaime Lannister"   "Joffrey Baratheon" "Margaery Tyrell"   "Myrcella Baratheon" "Renly Baratheon"
   [8] "Robert Baratheon"  "Selyse Florent"    "Shireen Baratheon" "Stannis Baratheon" "Steffon Baratheon" "Tommen Baratheon"   "Joanna Lannister"
  [15] "Tyrion Lannister"  "Tysha"

  + ... omitted several groups/vertices
```

The **label propagation** method labels each node with unique labels, and then updates these labels by choosing the label assigned to the majority of their neighbors and repeat this iteratively until each node has the most common labels among its neighbors.

```
IGRAPH clustering label propagation, groups: 27, mod: 0.79
+ groups:
  $`1`
  [1] "Alys Arryn"    "Elys Waynwood" "Jasper Arryn"  "Jeyne Royce"   "Jon Arryn"     "Lysa Arryn"    "Robert Arryn"  "Rowena Arryn"  "Petyr Baelish"

  $`2`
  [1] "Cassana Baratheon" "Ormund Baratheon"  "Renly Baratheon"   "Rhaelle Targaryen" "Selyse Florent"    "Shireen Baratheon" "Stannis Baratheon"
  [8] "Steffon Baratheon"

  $`3`
   [1] "Cersei Lannister"   "Jaime Lannister"    "Joffrey Baratheon"  "Myrcella Baratheon" "Robert Baratheon"   "Tommen Baratheon"   "Joanna Lannister"
   [8] "Tyrion Lannister"   "Tysha"              "Tywin Lannister"
  + ... omitted several groups/vertices
```

The **Louvain algorithm** initially assigns each node to its own community; nodes are then sequentially assigned to the community that increases modularity (if any) so that communities are merged; this merging process continues until modularity cannot increase or only one community remains.

Final Choice: Louvain Algorithm Results:

After analyzing results from different community detection methods, we choose output of 13 communities within the Game of thrones world. We get a modularity score of 0.84. These different communities make the most sense based on family ties and the storyline. These communities participate in events that drive different wars in the story. We now present the final 13 communities with its group members.

In Community 1 :

```
> union_characters$name[union_characters$cluster==1]
 [1] "Joanna Lannister"               "Amarei Crakehall"
 [3] "Lanna Lannister"                "Gerion Lannister"
 [5] "Walda Frey daughter of Merrett" "Rodrik Greyjoy"
 [7] "Willem Frey"                    "Lyanna Glover"
 [9] "Darlessa Marbrand"              "Aeron Greyjoy"
[11] "Luthor Tyrell son of Moryn"     "Tywin Lannister"
[13] "Viserys Targaryen"              "Jasper Arryn"
[15] "Marla Prester"                  "Elia Sand"
[17] "Cleos Frey"                     "Arya Stark"
```

For Community 2 :

```
> union_characters$name[union_characters$cluster==2]
 [1] "Leo Blackbar"
 [2] "Willas Tyrell"
 [3] "Emmon Frey"
 [4] "Kevan Lannister"
 [5] "Walder Frey son of Emmon"
 [6] "Robyn Ryswell"
 [7] "Stafford Lannister"
 [8] "Rickon Stark son of Benjen"
 [9] "Quenton Greyjoy"
[10] "Aerys II Targaryen"
[11] "Tywin Frey"
[12] "Tysha"
[13] "Rhaenys Targaryen daughter of Rhaegar"
[14] "Hoster Tully"
[15] "Luthor Tyrell"
[16] "Hizdahr zo Loraq"
```

For Community 3 :

```
> union_characters$name[union_characters$cluster==3]
 [1] "Genna Lannister"        "Minisa Whent"           "Doran Martell"
 [4] "Lady of House Sunderly" "Benfrey Frey"           "Petyr Baelish"
 [7] "Myrielle Lannister"     "Cerenna Lannister"      "Drogo"
[10] "Amerei Frey"            "Brandon Stark Burner"
```

For Community 4:

```
> union_characters$name[union_characters$cluster==4]
 [1] "Elyn Norridge"                "Melesa Crakehall"
 [3] "Lorra Royce"                  "Lyarra Stark"
 [5] "Shaera Targaryen"             "Rhaelle Targaryen"
 [7] "Betha Blackwood"              "Quellon Greyjoy"
 [9] "Tommen Baratheon"             "Benjen Stark"
[11] "Bethany Rosby"                "Daven Lannister"
[13] "Luthor Tyrell son of Theodore" "Rodwell Stark"
[15] "Oberyn Martell"               "Alysanne Farman"
[17] "Jeyne Westerling"             "Jeyne Royce"
[19] "Myranda Lefford"
```

For Community 5:

```
> union_characters$name[union_characters$cluster==5]
 [1] "Dorna Swyft"            "Perra Royce"            "Mellario of Norvos"
 [4] "Princess of Dorne"      "Robert Baratheon"       "Ellaria Sand"
 [7] "Willam Stark"           "Jaehaerys II Targaryen" "Edwyle Stark"
[10] "Obara Sand"             "Ramsay Bolton"          "Bran Stark"
[13] "Arra Norrey"            "Gerold Lannister"       "Marissa Frey"
[16] "Jenny of Oldstones"     "Renly Baratheon"        "Tyrion Lannister"
[19] "Robin Greyjoy"          "Daenerys Targaryen"     "Myrcella Baratheon"
[22] "Rickon Stark"
```

For Community 6:

```
> union_characters$name[union_characters$cluster==6]
 [1] "Brandon Stark son of Cregan" "Lancel Lannister"
 [3] "Jyanna Frey"                 "Catelyn Stark"
 [5] "Lady of House Piper"         "Euron Greyjoy"
 [7] "Tytos Lannister"             "Pate of the Blue Fork"
 [9] "Dorea Sand"                  "Rodrik Stark"
[11] "Alys Stackspear"
```

For Community 7:

```
> union_characters$name[union_characters$cluster==7]
 [1] "Arya Flint"                    "Shireen Baratheon"
 [3] "Shiera Crakehall"              "Ermesande Hayford"
 [5] "Lysara Karstark"               "Lucion Lannister"
 [7] "Lady of House Stonetree"       "Willem Lannister"
 [9] "Jaime Lannister"               "Jon Snow"
[11] "Mors Martell brother of Doran" "Marna Locke"
```

```
[13] "Loras Tyrell"                  "Sarella Sand"
```

For Community 8:

```
> union_characters$name[union_characters$cluster==8]
 [1] "Talisa Stark"        "Mina Tyrell"         "Obella Sand"
 [4] "Lia Serry"           "Alys Karstark"       "Tion Frey"
 [7] "Beron Stark"         "Tybolt Lannister"    "Erik Ironmaker"
[10] "Roose Bolton"        "Asha (Yara) Greyjoy"
```

For Community 9:

```
> union_characters$name[union_characters$cluster==9]
 [1] "Elinor Tyrell"              "Walder Frey son of Merrett"
 [3] "Olyvar Frey"                "Eddard Stark"
 [5] "Cersei Lannister"           "Robb Stark"
 [7] "Maron Martell"              "Jeyne Marbrand"
 [9] "Donel Greyjoy"              "Tyrek Lannister"
[11] "Melantha Blackwood"         "Duncan Targaryen"
[13] "Teora Kyndall"              "Victarion Greyjoy"
[15] "Steffon Baratheon"          "Leonette Fossoway"
[17] "Lyanna Stark"               "Antario Jast"
```

For Community 10:

```
> union_characters$name[union_characters$cluster==10]
 [1] "Unknown father Tyrell" "Edmure Tully"          "Mariya Darry"
 [4] "Jeyne Darry"           "Alerie Hightower"      "Mace Tyrell"
 [7] "Alannys Harlaw"        "Rowena Arryn"          "Lynara Stark"
[10] "Moryn Tyrell"          "Stannis Baratheon"     "Cregan Stark"
[13] "Cerissa Brax"          "Tyene Sand"            "Perwyn Frey"
[16] "Elia Martell"          "Ormund Baratheon"      "Elys Waynwood"
[19] "Olenna Redwyne"        "Maekar I Targaryen"
```

For Community 11:

```
> union_characters$name[union_characters$cluster==11]
 [1] "Rhaegar Targaryen"              "Selyse Florent"
 [3] "Rohanne Webber"                 "Theodore Tyrell"
 [5] "Urrigon Greyjoy"                "Nymeria Sand"
 [7] "Aegon Targaryen son of Rhaegar" "Sansa Stark"
 [9] "Horas Redwyne"                  "Jonnel Stark"
[11] "Jon Arryn"                      "Unknown mother Tyrell"
[13] "Rickard Stark"                  "Lyonel Frey"
[15] "Willamen Frey"                  "Garlan Tyrell"
[17] "Rhaella Targaryen"              "Aegon V Targaryen"
[19] "Tygett Lannister"               "Cassana Baratheon"
[21] "Lysa Arryn"                     "Olene Tyrell"
[23] "Rodrik Stark son of Beron"
```

For Community 12:

```
> union_characters$name[union_characters$cluster==12]
[1] "Dyanna Dayne"     "Margaery Tyrell"  "Harlon Greyjoy"   "Alys Arryn"
[5] "Brandon Stark"    "Walder Frey"      "Maron Greyjoy"    "Damion Lannister"
[9] "Robert Arryn"
```

For Community 13:

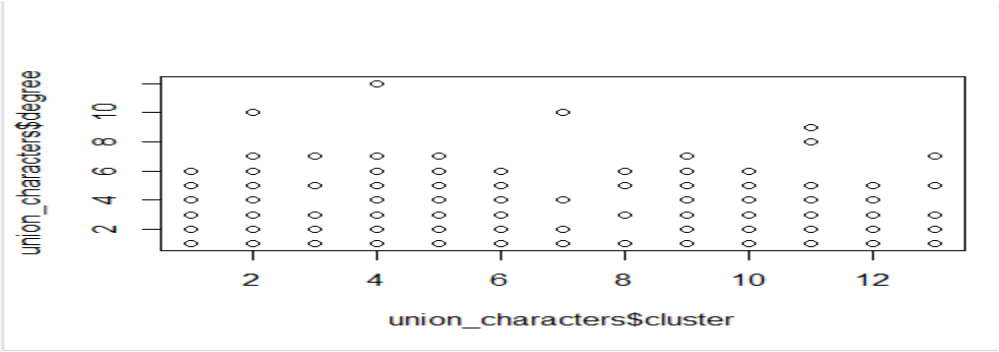
```
> union_characters$name[union_characters$cluster==13]
 [1] "Paxter Redwyne"               "Medwick Tyrell"
 [3] "Gilliane Glover"              "Merrett Frey"
 [5] "Artos Stark"                  "Balon Greyjoy"
 [7] "Damon Lannister son of Jason" "Damon Lannister"
 [9] "Jason Lannister"              "Theon Greyjoy"
[11] "Joffrey Baratheon"            "Roslin Frey"
[13] "Myriame Manderly"             "Loreza Sand"
[15] "Hobber Redwyne"               "Ella Lannister"
```

Exploring Relationship between centrality measures and community:

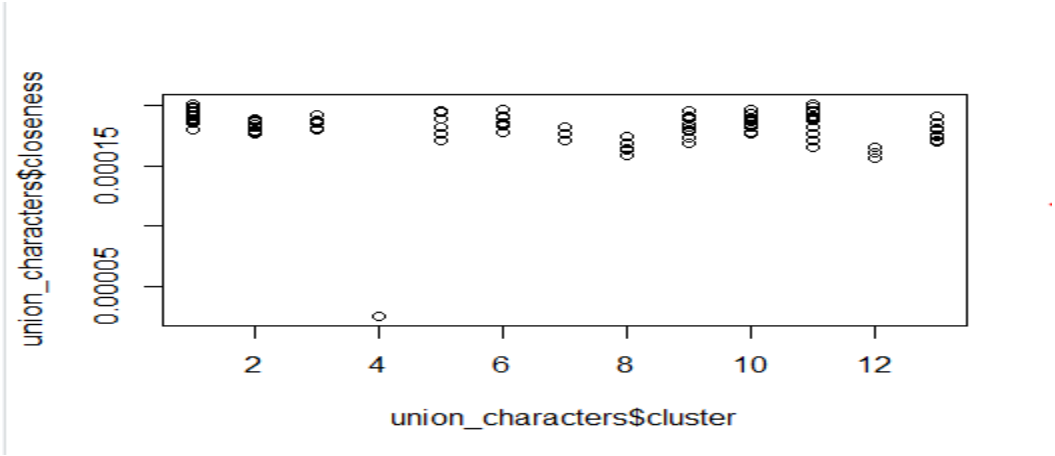
We consider if there's a linear relationship between centrality measures and communities that can potentially help us predict centrality measures based on community information. If there was, we could model centrality measure as dependent variable and community as predictor.

So, to investigate we plotted predictor (Community on x axis) and degree on y axis but as we can see this in below plot there was no linear relationship.

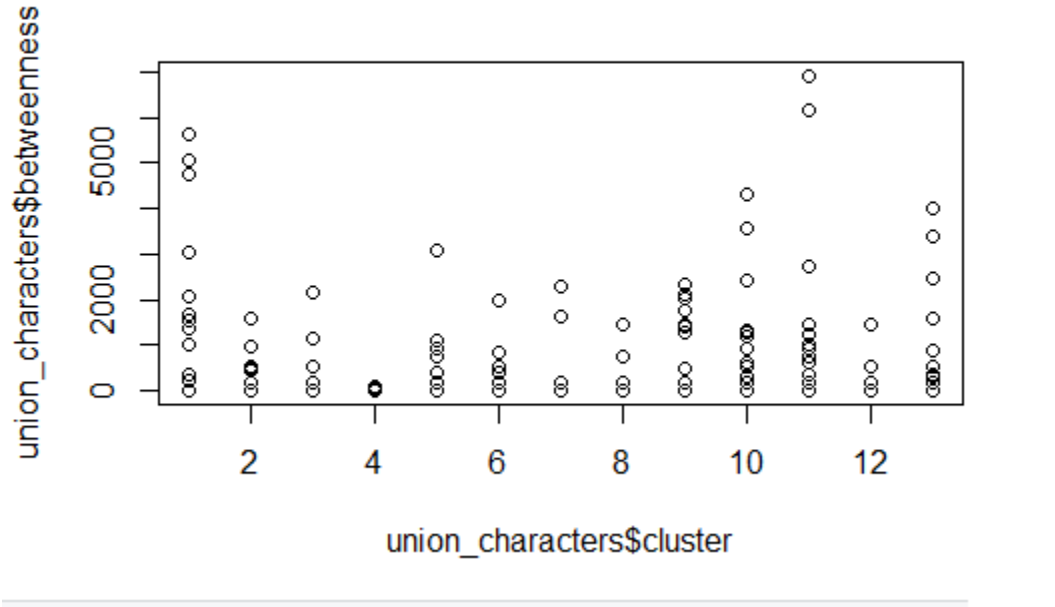
plot(union_characters$cluster,union_characters$degree)



Similarly we checked between closeness and Community and we couldn't find any.



We checked between Betweenness and Community:



Based on our understanding, we could not detect a linear relationship between centrality measures and community information.

References:

Kaggle GOT https://www.kaggle.com/mylesoneill/game-of-thrones

Wikipedia https://www.wikipedia.org/