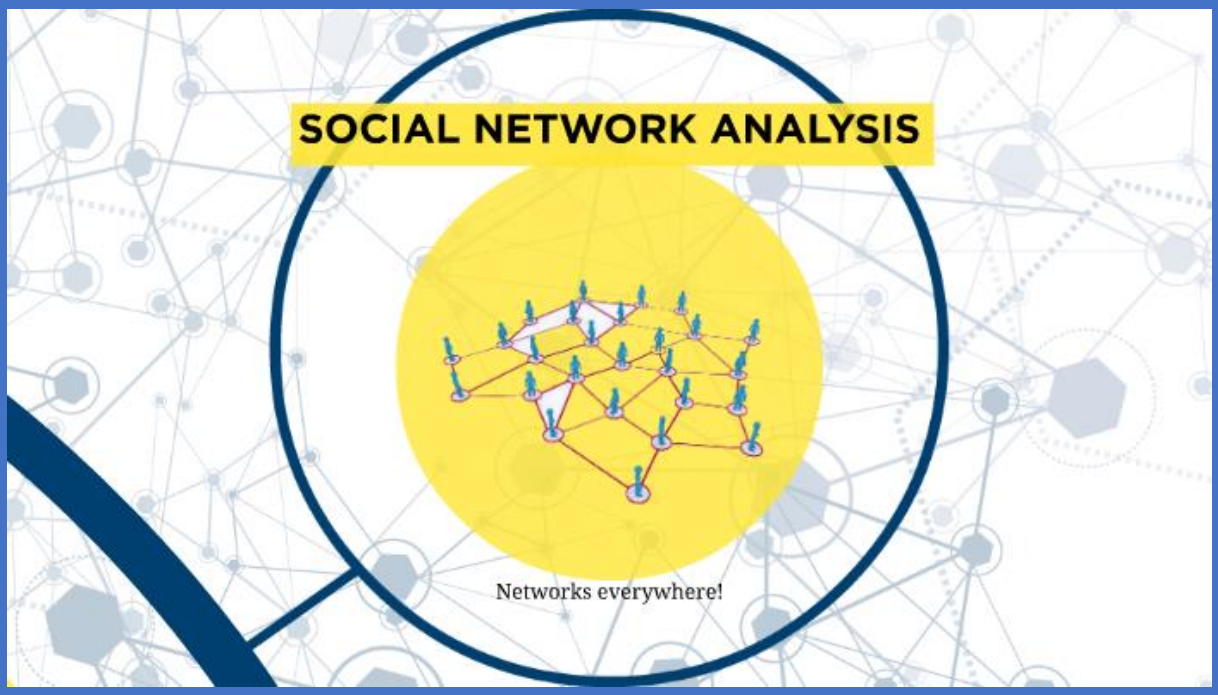
# SOCIAL NETWORK ANALYSIS INDIVIDUAL PROJECT ON CENTRALITY MEASURES




**Submitted By** –

BHURJI Prineet Kaur

## **Topic Background** –

*Centrality Measures*

In graph theory and network analysis, indicators of centrality identify the most important vertices within a graph. Introduced in 1979 in social network analysis, centrality measures are used to identify the most embedded vertices of a network.

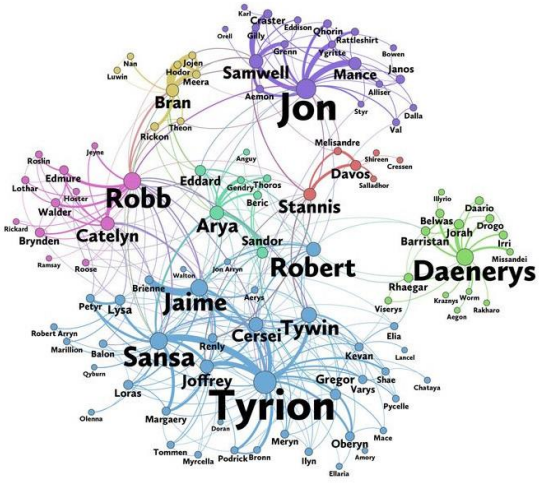
Centrality measures address the question: who is the most important or central person in this network? Because there is no consensus on exactly how centrality is defined and the proper procedure for its measurement, we will used different methods of centrality measure for our analysis including - the degree, closeness, betweenness, pagerank and eigenvector measures of centrality.

## **Project Definition** –

*Game of Thrones*

Game of Thrones is a fantasy series that has aired on HBO since 2011. It is based on the novels A Song of Ice and Fire written by the writer George R. R. Martin. It takes place in a fictional universe and narrates the power struggles of noble families for control of the Iron Throne.

Through this project we will analyse the co-occurrence network of the characters in the Game of Thrones books. Here, two characters are considered to co-occur if their names appear in the vicinity of 15 words from one another in the books.




This report presents social network analysis of the given GOT dataset by calculating centrality measures and exploring various methods to measure the centralities.

## About the Dataset –

The Data used for the analysis has been taken from Kaggle Game of thrones datasets. The dataset constitutes a network and is given as a text file describing the edges between characters, with some attributes attached to each edge. The complete network graph consists of 208 nodes and 404 edges.

*NODES Table* →

```
> # Exploring the Node Table
> str(union_characters)
'data.frame':   208 obs. of  8 variables:
 $ name      : chr  "Alys Arryn" "Elys Waynwood" "Jasper Arryn" "Jeyne Royce" ...
 $ male      : num  0 0 1 0 1 0 1 0 0 0 ...
 $ culture   : Factor w/ 65 levels "","Andal","Andals",..: NA NA NA NA 57 NA 57 NA NA 59 ...
 $ house     : chr  "House Arryn" "House Waynwood" "House Arryn" "House Royce" ...
 $ popularity: num  0.0803 0.0702 0.0435 0 0.8361 ...
 $ house2    : chr  NA NA NA NA ...
 $ color     : 'AsIs' chr  NA NA NA NA ...
 $ shape     : chr  "circle" "circle" "square" "circle" ...
> 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
```

The table consists of one row for each character that is either a source or a target in the edge table. Attributes included – gender (whether male or not), house (each character was born in) and popularity. House2 to assign a colour to only the major houses and Shape depicts the gender.

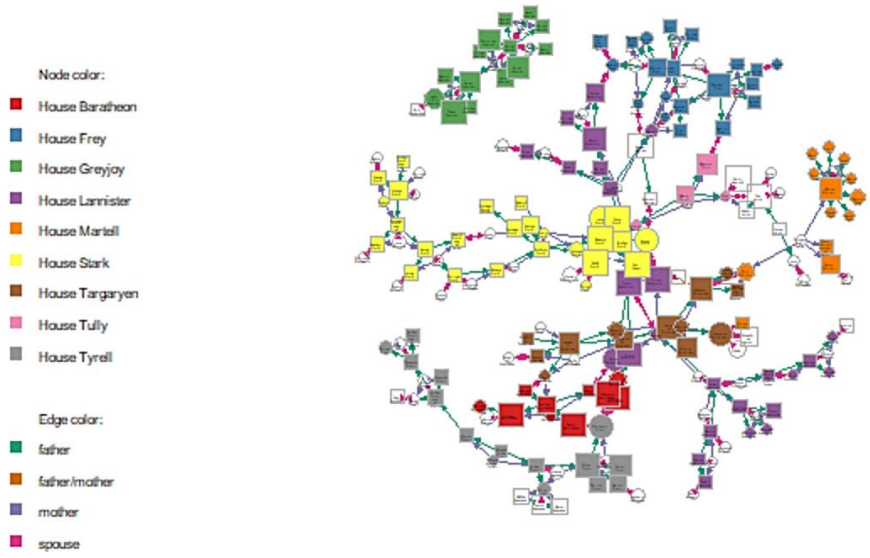
*EDGES Table* →

```
> # Exploring the Edge Table
> str(union_edges)
'data.frame':   404 obs. of  5 variables:
 $ source: chr  "Lysa Arryn" "Jasper Arryn" "Jasper Arryn" "Jon Arryn" ...
 $ target: chr  "Robert Arryn" "Alys Arryn" "Jon Arryn" "Robert Arryn" ...
 $ type  : chr  "mother" "father" "father" "father" ...
 $ color : 'AsIs' chr  "#7570B3" "#1B9E77" "#1B9E77" "#1B9E77" ...
 $ lty   : chr  "solid" "solid" "solid" "solid" ...
> 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 table consists of source nodes, target nodes and edge attributes including type of interaction (whether mother, father or spouse), the color and line-type assigned to each edge.

## Network Analysis using Centrality Measures –

*Graphical representation of the GOT Family Ties Network*




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

## Exploring various Centrality Measures –

Centrality describes the number of edges that are in- or outgoing to/from nodes. High centrality networks have few nodes with many connections, low centrality networks have many nodes with similar numbers of edges. Given below is Centrality for complete network followed by various measures of Centrality.

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

### *Node Degree Centrality*

The degree centrality of a node is simply its degree—the number of edges it has. The higher the degree, the more central the node is. This can be an effective measure since many nodes with high degrees also have high centrality by other measures.

Degree centrality is a good measure of the total connections a node has but will not necessarily indicate the importance of a node in connecting others or how central it is to the main group.

| | 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 |

For our dataset of GOT, the node degree seems to depict how many family members a character has (this may include spouses or children). Quellom Greyjoy, has the highest degree as he was a grandfather and had many wives and children.

### *Closeness Centrality*

Closeness centrality measures everyone's position in the network via a different perspective from the other network metrics, capturing the average distance between each vertex and every other vertex in the network. The nodes that are more central are also closer to all the other nodes. This centrality measure tests the ability of a node to affect the other nodes in the network.

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

## *Betweenness Centrality*

Betweenness centrality measures the extent to which a vertex plays this bridging role in a network. Specifically, betweenness centrality measures the extent that the user falls on the shortest path between other pairs of users in the network. If a node has a high betweenness centrality, it means that it is on the path between many other nodes and is likely an important connector in the network. Vertices with high betweenness are key bridges between different parts of a network.

Node Betweenness –

| | 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 |

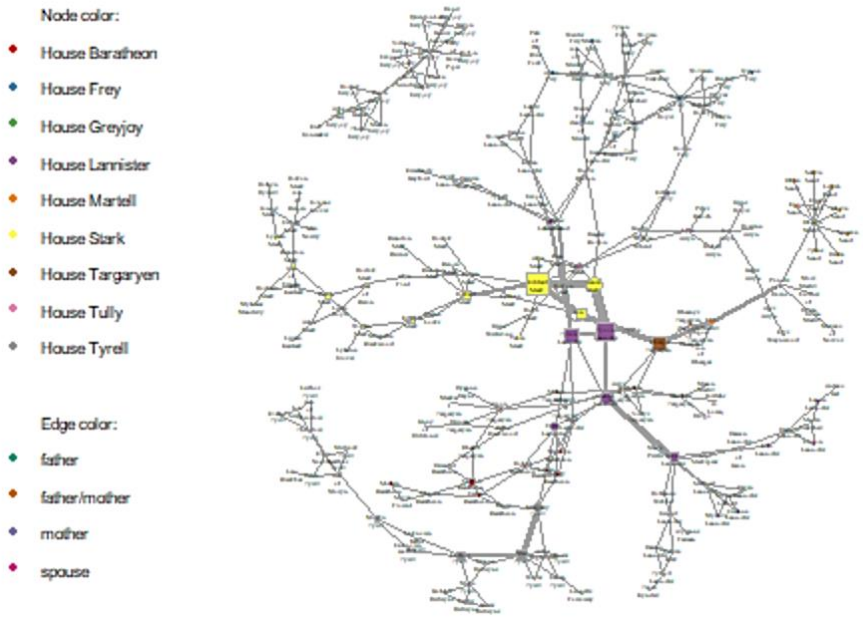
As we see above, Eddard Stark tops the chart of betweenness followed by Sansa Stark and Tyrion Lannister.

Edge Betweenness –

| | 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 |

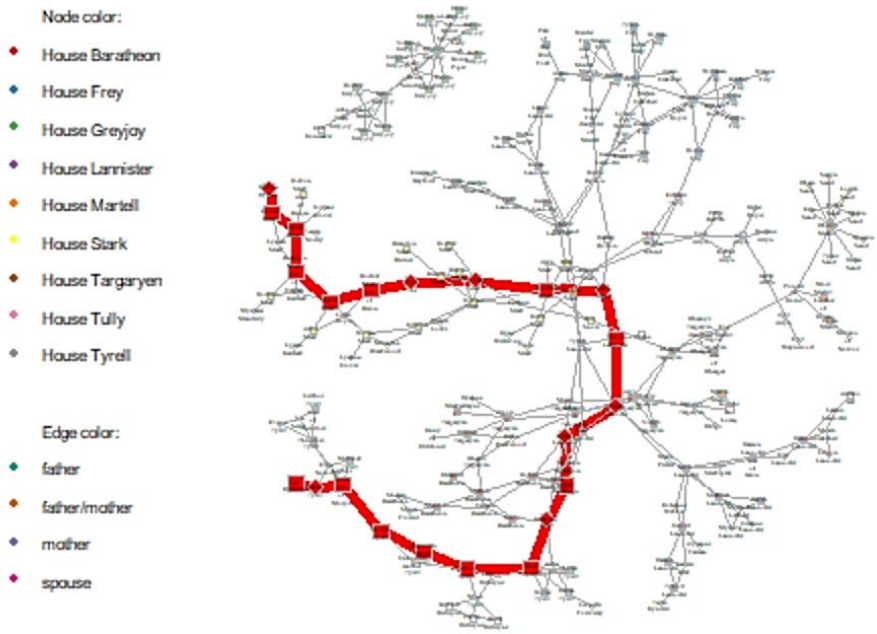
*Graphical representation of the Node betweenness vs Edge betweenness*

For Plotting purpose took node betweenness as “vertex.size” and edge betweenness as “edge.width”.



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.




### *PageRank Centrality*

Page Rank is an algorithm used by Google Search to rank web pages in their search engine result. It is considered as a variant of Eigen Centrality but designed for web ranking and accounting for link direction. It uses hyperlinks between pages as a measure of importance to rank web contents. Each node of the network is assigned a score based on its number of incoming links. These links are also weighted depending on the relative score of its originating node. As a result, nodes with many incoming links are influenced by nodes to which they are connected.

| | name | page_rank |
|---|---|---|
| Oberyn Martell | Oberyn Martell | 0.018402407 |
| Quellon Greyjoy | Quellon Greyjoy | 0.016128129 |
| Walder Frey | Walder Frey | 0.012956029 |
| Eddard Stark | Eddard Stark | 0.011725019 |
| Cregan Stark | Cregan Stark | 0.010983561 |
| Catelyn Stark | Catelyn Stark | 0.010555473 |
| Lyarra Stark | Lyarra Stark | 0.009876629 |
| Aegon V Targaryen | Aegon V Targaryen | 0.009688458 |
| Balon Greyjoy | Balon Greyjoy | 0.009647049 |
| Jon Arryn | Jon Arryn | 0.009623742 |

Oberyn Martell, Quellon Greyjoy and Walder Frey all have maximum number of spouses, children, and grandchildren and thus they score highest in terms of PageRank.

### *Eigen Vector Centrality*

Eigenvector centrality is a measure of how well connected a vertex is. The eigenvector approach to centrality calculates a relative score based on the number of links a node has to other nodes in the network. This shows the importance of a node given the importance of its neighbours. For this reason, this centrality measure is often referred to as the prestige centrality. It basically starts by measuring each nodes' degree score.

We calculate the eigenvalues and eigenvectors of the adjacency matrix. The eigenvector 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 |

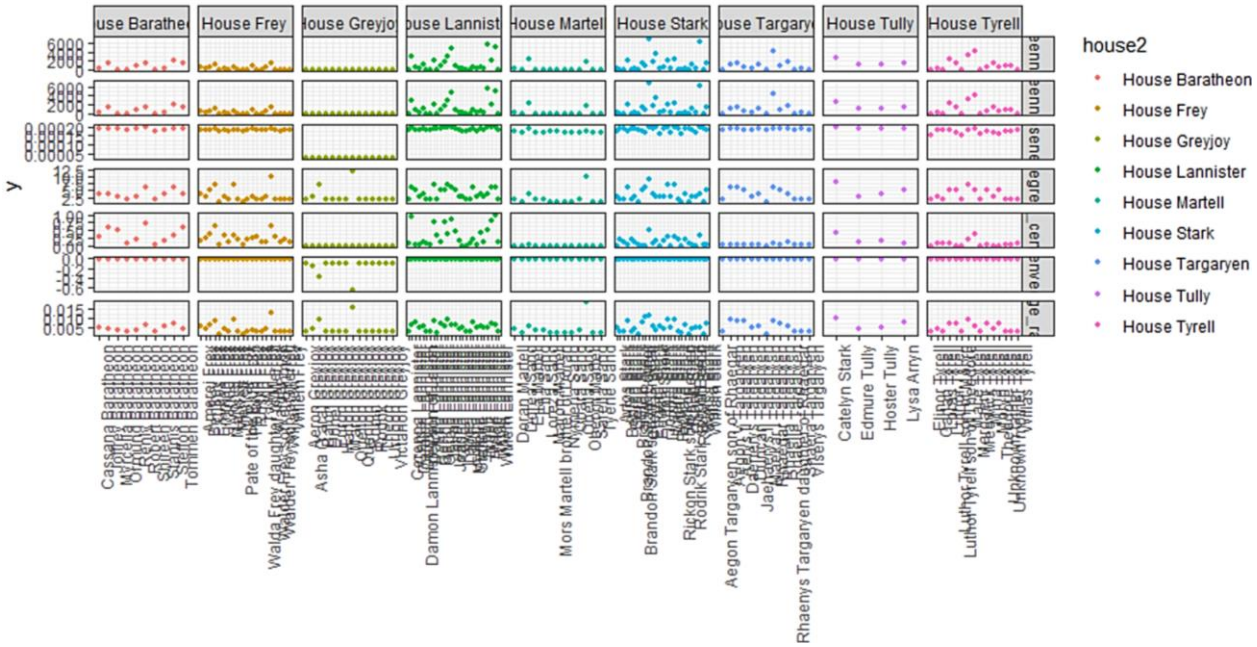
We can see above that the Greyjoy's have been scored with the highest eigenvalues as they have a highly connected family tie.s

| | name | eigen_centrality |
|---|---|---|
| Tywin Lannister | Tywin Lannister | 1.0000000 |
| Cersei Lannister | Cersei Lannister | 0.9168980 |
| Joanna Lannister | Joanna Lannister | 0.8358122 |
| Jeyne Marbrand | Jeyne Marbrand | 0.8190076 |
| Tytos Lannister | Tytos Lannister | 0.8190076 |
| Genna Lannister | Genna Lannister | 0.7788376 |
| Jaime Lannister | Jaime Lannister | 0.7642870 |
| Robert Baratheon | Robert Baratheon | 0.7087042 |
| Emmon Frey | Emmon Frey | 0.6538709 |
| Walder Frey | Walder Frey | 0.6516021 |

From the above 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.

## Summary of Centrality Results –

By exploring different methods of measuring Centrality, we saw different relevance of different nodes. Using all this information and comparing all these measures we can decide which characters are the most important in Game of Thrones.




From the above plot showing all characters from the major houses, 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.

## References –


- Game of Thrones — Wikipédia (wikipedia.org)
- Game of Thrones - Official Website for the HBO Series - HBO.com
- Centrality Measure - an overview | ScienceDirect Topics
- Centrality - Wikipedia
- Social network analysis: Centrality measures (cambridge-intelligence.com)
- Degree Centrality - an overview | ScienceDirect Topics
- Closeness Centrality - an overview | ScienceDirect Topics
- Betweenness Centrality - an overview | ScienceDirect Topics
- Pagerank Algorithm - an overview | ScienceDirect Topics
- Eigenvector Centrality - an overview | ScienceDirect Topics
- Game of Thrones: Network Analysis | Kaggle
- https://www.kaggle.com/mylesoneill/game-of-thrones