Graph Analytics and Algorithms Case Study

Group Members:

Dharineesh Karthikeyan

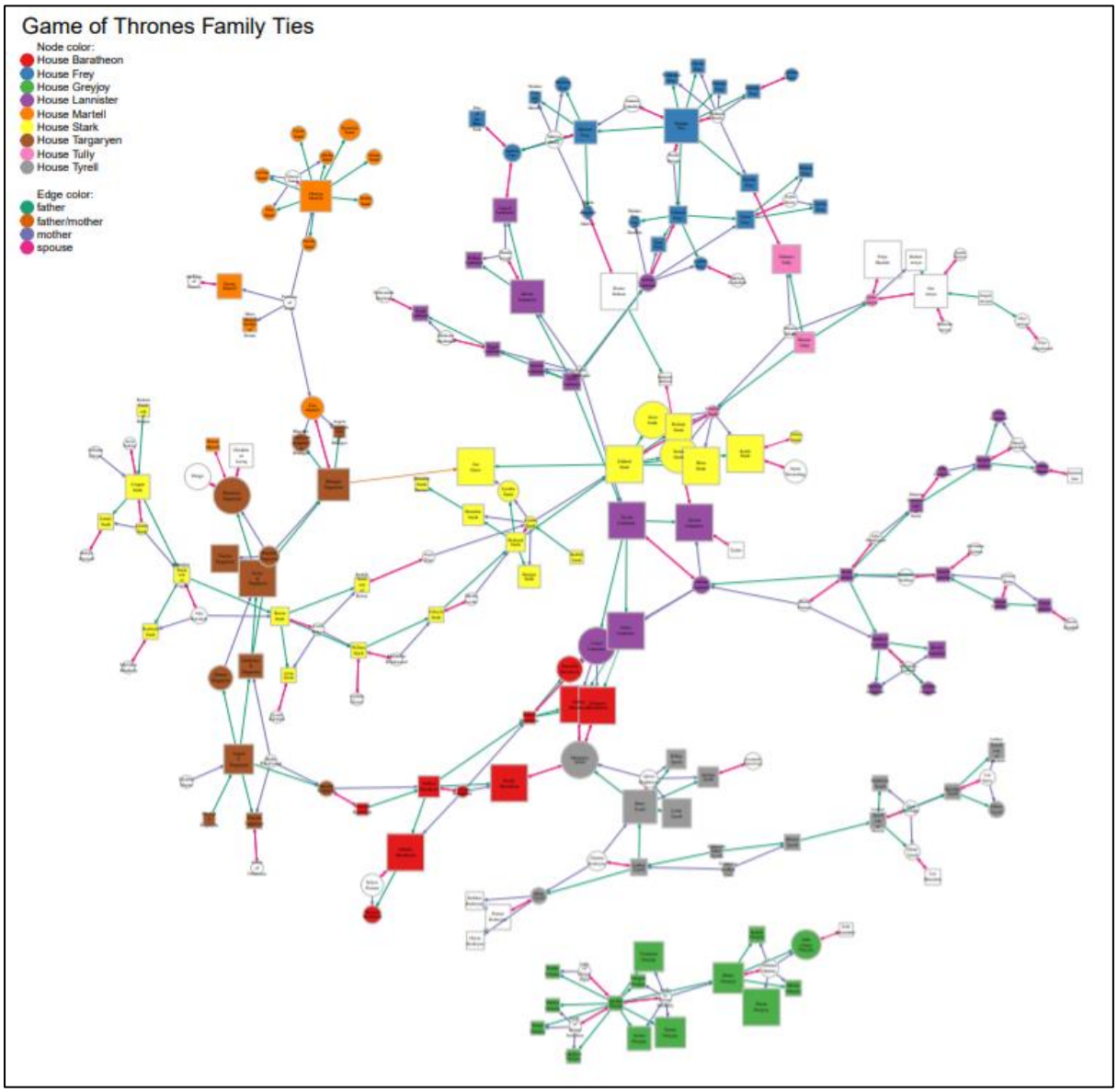
Varshini P.J

About the Network Data:

The dataset and network chosen is of the topic **"Game of Thrones family relationships".**

Game of Thrones is a famous book series and a TV series. (Our data focuses mainly on the data collected from the TV series). Our network is a **family network** of the characters from the series.

Let's try to display the network and see,



Additional information represented in the network are:

- **Nodes:** Characters in the franchise
- **Edges:** Various relationships between the characters (Mother/Father/Son/Daughter/Spouse)
- **Number of Nodes and Edges in the network graph:**

```r
print('Number of nodes')
gorder(union_graph)
print('NUmber of edges')
gsize(union_graph)
```

```
[1] "Number of nodes"
[1] 208
[1] "NUmber of edges"
[1] 404
```

Game of Thrones has a total of 460+ houses, in which our network graph only focuses on the **major 9 houses.**

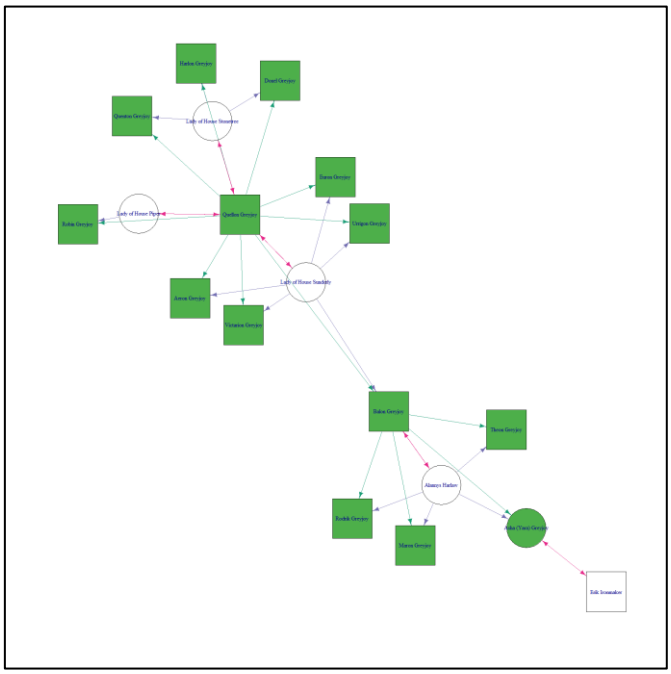
- **Node colour**: Shows the major houses (9 in total). The nodes with no node colour (i.e white) belongs to other less known houses.
- **Node size**: The character's popularity based on ratings
- **Node shape:** Their gender (square for male, circle for female).
- **Edge colour:** Shows interaction type. (Mother/Father/Spouse/etc..)

Network Analysis:

Tools used to perform Analysis: R

Disconnected graph:

When we look for disconnected subgraphs, we can only find one and it is the following:



Inference:

- From the **node colour as green**, we can identify they all belong to the **House Greyjoy family** is being disconnected from the rest of the families and has no ties with the other families.

Centrality:

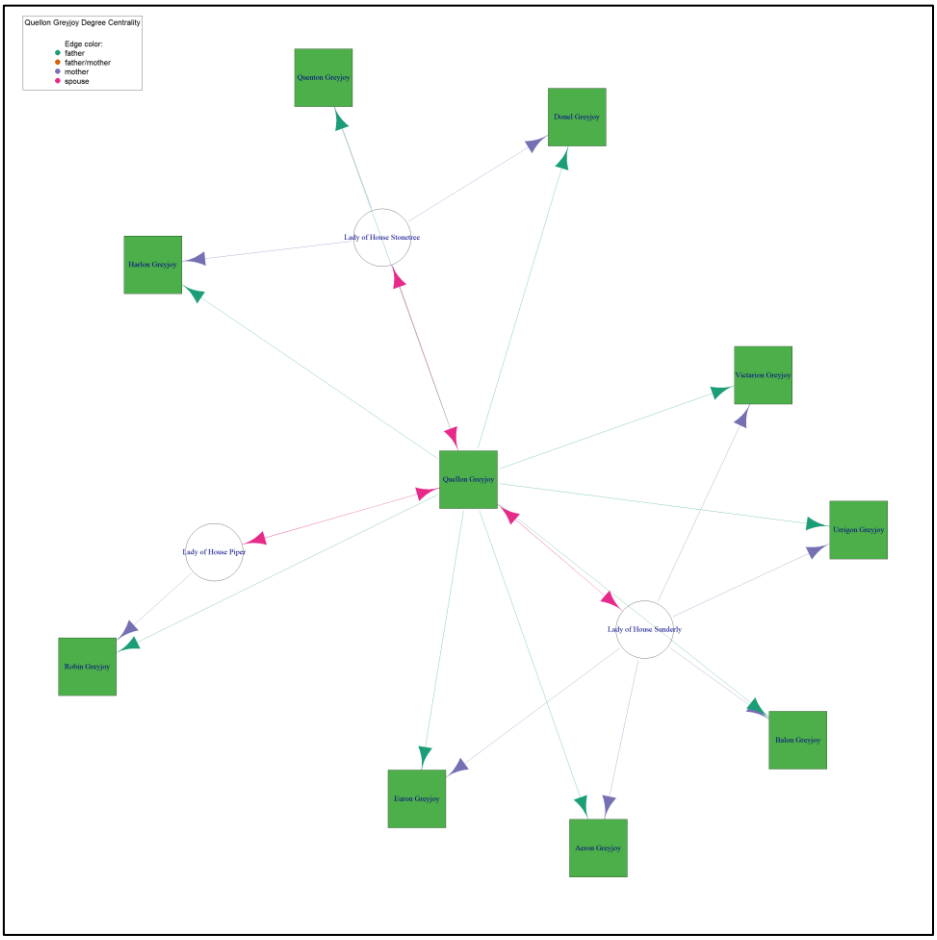
a) Degree Centrality:

We calculate the degree centrality of each node in the graph and display the top 10 nodes with the highest degree centrality.

	rowname <chr>	degree <dbl>
1	Quellon Greyjoy	12
2	Walder Frey	10
3	Oberyn Martell	10
4	Eddard Stark	9
5	Catelyn Stark	8
6	Emmon Frey	7
7	Genna Lannister	7
8	Merrett Frey	7
9	Balon Greyjoy	7
10	Jason Lannister	7

1-10 of 10 rows

From the above table, we can see **Quellon Greyjoy** has the highest degree centrality. Let's try to take a look at the network connection of Quellon Greyjoy in detail.



Inferences:

- From the above graph, we can see that Quellon Greyjoy has **3 pink edges**, denoting that he had **3 wives**.
- Also, there are **9 outgoing green edges**, denoting that he had **9 children**. All the nodes are **square in shape**, showing that all 9 of his children are **male**.
- Hence, we can see why he has the highest degree centrality as he has the **biggest family among all the other families**.

b) Closeness Centrality:

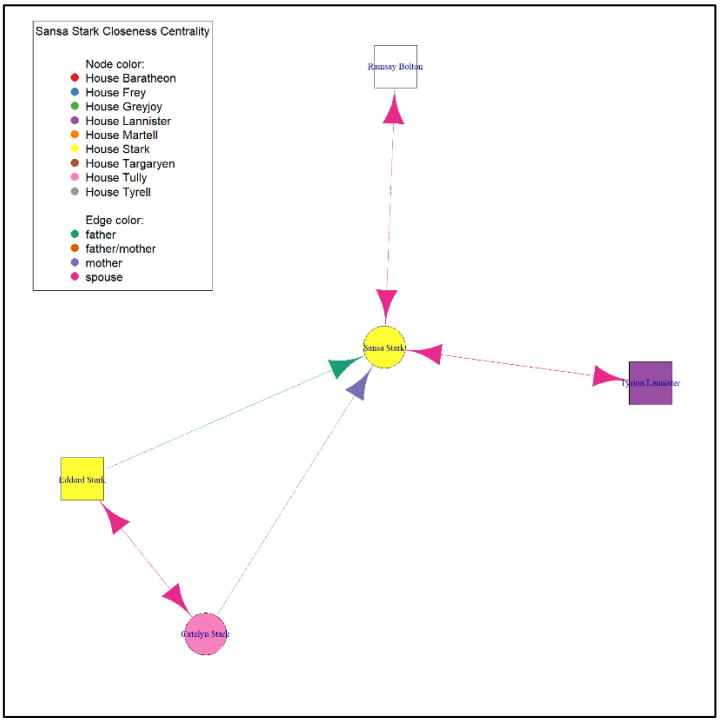
The closeness of a node describes its distance to all other nodes. A node with highest closeness is more central and can spread information to many other nodes.

We calculate the closeness centrality of each node in the graph and display the top 10 nodes with the highest closeness centrality.

	rowname <chr>	closeness <dbl>
1	Sansa Stark	0.0002013288
2	Tyrion Lannister	0.0002012882
3	Tywin Lannister	0.0002011668
4	Joanna Lannister	0.0002005616
5	Eddard Stark	0.0002002804
6	Catelyn Stark	0.0001986492
7	Cersei Lannister	0.0001984915
8	Jaime Lannister	0.0001975894
9	Jeyne Marbrand	0.0001966568
10	Tytos Lannister	0.0001966568

1-10 of 10 rows

From the above table, we can clearly see that the closeness centrality of all the top 10 people are quite close to each other. We can see Sansa Stark has the highest closest centrality. Let's try to take a look at the network connection of the Sansa Stark.



Inferences:

- Sanasa Stark is connected to other people such as Tyrion Lannister, Catelyn Stark and Eddard Stark. From the above, closeness centrality node table we can see that the above-mentioned nodes are also in the top 10 most important nodes based on closeness centrality.
- In addition to this she is married to two people, Tyrion Lannister and Ramsay Bolton, in which Tyrion Lannister is the second most important node according to closeness centrality.
- Hence, from this we can see that the connection she has with other prominent characters and being married to another important person (Tyrion Lannister) explains why Sanasa Stark is the most important node based on closeness centrality.
- Along with this, from the top 10 closeness centrality nodes table we can easily identify that the **Stark and Lannister family** has the dominance in importance and is more connected with the other members of its own families and other families as well.
- **Note:** The characters with highest closeness all surround central characters that connect various storylines and houses in Game of Thrones series.

c) Clustering Centrality:

The top 10 clustering coefficient for each node calculated is as follows:

	name <chr>	clustering_coeff <dbl>
1	Robert Arryn	1
2	Ormund Baratheon	1
3	Selyse Florent	1
4	Shireen Baratheon	1
5	Amarei Crakehall	1
6	Marissa Frey	1
7	Olyvar Frey	1
8	Perra Royce	1
9	Perwyn Frey	1
10	Tion Frey	1

1-10 of 10 rows

Inference:

- As we can see the clustering coefficients are 1 for all of the nodes. This is because the network we are working with is a **family network**, characters with a clustering coefficient of **1** form **triangles with their parents or offspring.**
- The local clustering coefficient of a node in a graph quantifies how close its neighbours are to being a clique (complete graph). Hence, the **maximal clique is 3**, since the largest group of nodes possible is only 3 (Parent/child relationships only)
- Since, most nodes are at 1 since it is a family network. We cannot infer anything about the importance of the characters from this centrality.

d) Betweenness Centrality:

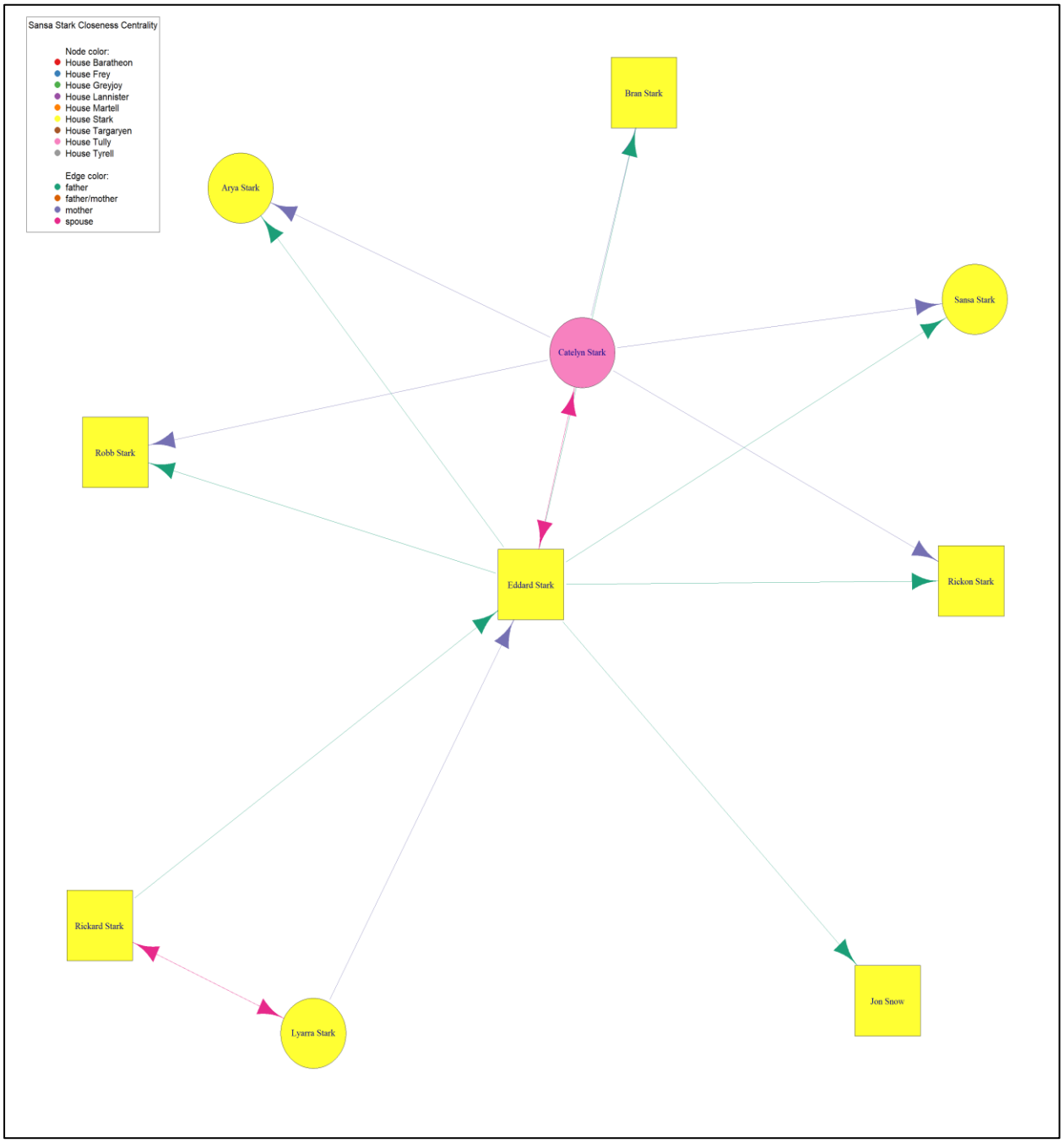
Betweenness describes the number of shortest paths between nodes. Nodes with high betweenness centrality are on the path between many other nodes.

We calculate the betweenness centrality of each node in the graph and display the top 10 nodes with the highest betweenness centrality.

	rowname <chr>	betweenness <dbl>
1	Eddard Stark	6926.864
2	Sansa Stark	6165.667
3	Tyrion Lannister	5617.482
4	Tywin Lannister	5070.395
5	Joanna Lannister	4737.524
6	Rhaegar Targaryen	4301.583
7	Margaery Tyrell	4016.417
8	Jon Snow	3558.884
9	Mace Tyrell	3392.500
10	Jason Lannister	3068.500

1-10 of 10 rows

Taking a deeper look at Eddard Stark's network connections,



Inferences:

- Eddard Stark is the father of Sansa Stark who is the most important node based on closeness centrality and also Sansa Stark is the second most important node based on betweenness centrality also.
- From the last closeness centrality, we studied that Sansa Stark is married to Tyrion Lannister, who is also a very important person based on betweenness centrality. Hence, being the son-in-law also contributes to the centrality of Eddard Stark.
- Thus, **Eddard Stark** is one of the most important characters who has contact with most other families and people.

Result and Analysis:

We studied the different centralities of the graph, in conclusion we found out that:

- Degree Centrality refers to the person who has the biggest family
- Closeness Centrality refers to the person who is more central and has contacts with more important people
- Clustering Centrality here is the same for every person, since it is a family graph and the connections between parent/child relationships is always connected and a maximal clique.
- Betweenness Centrality refers to people who are key connections or bridges between different groups of nodes. In our network, these nodes would be very important because they are likely to pass on information to a wide reach of people.

Who are the most important people in the network?

- From degree centrality, the answer is **Quellon Greyjoy** to be the most important person in the network. This is **cannot be true**, since we saw that the **Greyjoy family is the only disconnected subgraph** from the other families. Hence, they do not have much connection with the rest of the people in the given network
- Studying the closeness centrality, we can see that the **Stark and Lannister family** seems to be the most important families in the network.
- Since the clustering centrality is 1 for every person, it **fails to provide any real information** in our given network.
- Observing the betweenness centrality, we can say **Edward Stark** to be the most important person in the graph.
- Correlating the top 10 nodes of closeness and betweenness centralities we can easily observe and conclude that **Edward Stark, Sansa Stark, Tyrion Lannister and Tywin Lannister** is consistently on top based on centralities. So, these 4 are the most important people in the network. True to the study we end up Stark and Lannister family being most prominent.
- Looking again at the relationships between these four are as follows:
    - Edward Stark is the father of Sansa Stark
    - Sanasa Stark is married to Tyrion Lannister
    - Tywin Lannister and Tyrion Lannister belong to the same Lannister family
- Hence, the four most important people from the graph that we studied are: **Edward Stark, Sansa Stark, Tyrion Lannister and Tywin Lannister.**
