Journal of Cases on Information Technology
Volume 23 • Issue 4

Social Network for Game of Thrones

Manvi Breja, NorthCap University, Gurgaon, India*
Himanshi Bhatia, NorthCap University, Gurgaon, India
Dollie Juneja, Northcap University, Gurgaon, India

ABSTRACT

Along with growing interest and use, the concept of network analysis has taken a new direction to explore data and facts to find existing patterns. The paper highlights the importance of social network analysis in analyzing and mining useful information from the data across various domains. It provides an insight into need, importance, and scope of social network analysis. With the use of social networking tools like NetworkX, data is being represented in the form of graph or network which is then analyzed in a more efficient way making it easier to study the interactions between different persons in Game of Thrones and establishing trends existing in a network. A comparative analysis of various centrality measures such as degree centrality, betweenness centrality, closeness centrality, page rank centrality is performed to explore the features associated to find the most important character of the series based on obtained results.

KEYWORDS

Betweenness, Centrality, Closeness, Connections, Degree, NetworkX, PageRank, Social Network Analysis

INTRODUCTION

Introduction to SNA

Social Network Analysis (SNA) is the process of analyzing relationships that exist between social entities in terms of network and graph theory (Serrat, 2017). The structure of a network consists of social entities that are often people, social groups, organizations and communities (O'Malley & Onnela, 2017). Social network analysis involves the empirical study of how social entities interact with each other within the network. The network structure is formed using nodes which is any entity like people or individual actors or groups and edges that depicts link/interaction/relationship between nodes (Serrat, 2017). The methodical analysis of networks is commonly visualized through sociograms in which the nodes are represented as points and ties are represented as lioness which is further used to visualize various commonly used social structures including meme spread, business networks, disease transmission, sexual relationships, difficult working relationships, collaboration graphs, kinship, information circulation and friendship networks. In social network analysis two nodes are said to be related if they communicate frequently or speak in some way. A social network provides various metrics for understanding networks and the individuals and groups within them. The visual analysis of networks through graphs allows us to measure the strength of the relationships and how information flows between people, groups, communities and other social entities. Thus, social

DOI: 10.4018/JCIT.296252 *Corresponding Author

network analysis is a method that provides a best possible way to visualize entities, their patterns, connections and interactions to share and make effective utilization of their knowledge (Breja, 2017). It can generate graphical representations that reveal individuals in populations that bridge social groups.

Problems Addressed by SNA

There is an increased amount of interactions over the Internet, with the proliferation of technologies and data and so social network analysis has become increasingly relevant in recent years. To handle problems in social domains and real world, Social Networks have ushered in a multidimensional approach. It provides a new angle to problem solving decisions. Social network helps us in analyzing links between individuals; the way they are socialized. The internet imposes new questions and allows a broader perspective for SNA to find solutions to analyze similar patterns in different situations, conditions when networks gravitate around certain things like money, popularity, and advertising, identifying most connection person in network also known as influential nodes, understanding the smooth flow of information between entities, finding key people that form communities, nodes sharing common interests, visualizing the strength of bond and identify clusters inside and outside the community (Bonato et al. 2016).

Applications of SNA

Social network analysis is an emerging field with numerous applications. It is used extensively in a wide range of domains that includes sociology, business, organizational behavior, government, medicine, security, etc. Some commonly used applications (Serrat, 2017) include link prediction, social sharing and filtering, location-based interaction, analysis, community-maintained resource support, recommender systems development, network propagation modelling, user attribute and behavior analysis and data aggregation and mining. In the private sector, SNA facilitates businesses in team building and strategy building and helps to understand the connection between patterns of interaction and business outcomes such as job satisfaction, job performance, likelihood of creation of new ideas and adaptation of new technologies. It helps us to draw underlying facts in a network that are usually hidden when dealing with raw information. It identifies where collaboration breaks down and where decisions are getting bogged down, helping the HR to come up with better decisions (Serrat, 2017, Froehlich et al., 2020, De et al., 2012). The section below addresses certain fields and how social network analysis influence such fields are discussed below:

1) **Telecom:** Social network analysis is used by telecom industries to understand the quantity and location of cell towers to ensure the maximum coverage.
2) **Security:** To ensure security, network analysis is used in counter-intelligence and law enforcement activities. It also finds unusual patterns useful for mitigating fraudulent transactions. By analyzing the flow of money across interconnected banks helps in detecting terrorist's activity.
3) **Supply Chain:** In this domain, social network analysis is used to identify the locations for warehouses and delivery centers. Also helps in identifying optimum rotes for your delivery trucks.
4) **Pharma:** Cutting cost and reducing travel time of the salesmen person by optimizing the salesmen route.
5) **Computer supported collaborative learning:** The most recent application of social network analysis is CSCL. It helps to understand how learners collaborate in terms of quality, length, frequency, and amount as well as on the topic and strategies of communication.
6) **Classroom social communications:** Using social network analysis in classroom social communications.
7) **Marketing Analysis:** Helps you to figure out the most prominent person in the network by. By routing the message through it, advertisers and marketers can estimate the biggest bang for the marketing buck.

8) **Public housing:** It helps us to identify how young people get involved in vital development after experiencing housing instability and transformation.
9) **Business Sales:** In the form of social CRM, trend spotting, loyalty programs, social marketing, direct marketing, reputation monitoring, social network analysis has proven its power on sales and marketing.
10) **Internet Applications**: To understand the behavior of individuals or organizations through social linkages on social sites like Facebook, Twitter and Instagram, social network analysis is applied to social media.
11) **Crime detection:** Social networks have helped drawing hidden facts and patterns which have indirectly been proven effective in prevention of various crimes.
12) **Textual analysis:** Large textual corpora can be turned into networks and then analyzed with the method of social network analysis.

BACKGROUND

Landherr et al. (2010) present a current study on centrality measures in social networks based on its increasing importance in the field of wikis, blogs, online web platforms where individuals interact with each other, based on which it identifies core individuals within a social network. The study shows the comparison of various measures like degree centrality, betweenness centrality, closeness centrality, eigenvector centrality in respect to different types of requirements based on various ongoing researches and findings of several social network analysis literatures. This paper intends to help understand the existing works. Different properties of centrality measures in social networks were used to find the connectedness of a node in the network by analyzing and formulating these measures using an example network against various actors. The analysis shows that betweenness centrality is very unstable when dealing with a varying network structure whereas degree centrality and closeness centrality are slightly affected by the variation in the network structure. Further the work shows how different centrality measures are used with different approaches to analyze the robustness, sensitivity, stability of different models with varying network models when sampled with inconsistent data. The findings indicate how a social network can provide valuable information for different application scenarios when used in a reflective manner.

Chowdhury et al. (2019) through his work describes the deviation of story in books and its adapted movie by analyzing the novels Nastanirh and Harry Potter and the Philosopher's Stone by the means of Network analysis using weighted graph centrality measures to formulate the dissimilarity of the characters in book and their visualized form. A comparison is drawn using network analysis to find out variation in centrality measures, importance of characters and their distribution in each chapter in comparison with the corresponding scenes in visual counterparts. The Mantel test has been performed to find the correlation between these two centrality matrices of the story and adapted film to validate the results. By the means of closeness and betweenness centrality scores the approach identified the amount of divergence that is made in connections of characters in the actual story and its visual form .The graphical visualizations made it easy to understand and study the amount of variations that are made to make it more realistic for the audience.

Breja (2017) presented a model of visualizing properties of social network analysis in the question answer community. The paper implements a Question Answering Community using a new approach using which the system will be able answer questions with the use of information retrieval and natural language processing .The approach helps user to share their ideas/knowledge/views, resolve their problems by discussing with group of people that share common interests of asking questions. The QA community system returns a simple appropriate answer to the question posed by the user, so that the user does not have to scrawl over all the web pages to find one appropriate answer. The author aims to implement the network of people, asking some similar questions in her future work using some existing tools.

Morgan et al. (2019) investigated health behavior in school students when returning from summer holidays. Their analysis explained that Socio-economic inequalities in a range of health risk behaviors as well as mental health and well-being proliferates through childhood and adolescence. The paper aims to examine the importance of summer break experiences on young people's mental well-being on their return to school. The summer holiday experiences include hunger, loneliness, involvement in exercise and time spent with their friends. Relationship between socioeconomic status and mental health and well-being is majorly explained by summer holiday experiences. It was observed that it is important for schools to recognize that school holiday interventions provide a short-term fix. The findings of the analysis clearly indicate that school holiday interventions through reducing loneliness, providing necessary nutritious food and opportunities for social interaction and involvement may offer significant potential to reduce socioeconomic inequalities in young one's mental health and well-being on their return to school. It was concluded that the consequence of socioeconomic inequalities within society is the socioeconomic inequalities in health.

Zhang et al. (2019) provides a systematic review to achieve sustainable employee training by analyzing the similarity in behaviors of corporate training. The study also seeks to find a way to provide an intuitive and visible representation of the similarity by using the methodology of multidimensional scaling. The application of the social network analysis is to quantitatively analyze and describe the corporate social responsibility (sustainability) reports and relationships that exist between training and enterprises, among training and among enterprises. The results were concluded as: (1) SNA can help in reducing the costs of training implementations while achieving the goal of employee training through shared trainings that improves the quality and efficiency of the training at the same time. (2) Shared training procedures provides a platform and opportunity to communicate and promote the knowledge and information exchange among employees both inter-industry and intra-industry providing a diverse environment to improve the creativity and knowledge, making the employees be more sustainable. It was identified that the similarities in corporate employee training is a way to enhance corporate sustainability of human resource management. The future work involves the study over the change in employee training over time. Also, investigation and analyses of the factors that influence the behavior of corporate training.

Bag et al. (2019) designed a predictive model that can help customers in making an effective decision while purchasing in a short period of time. Also, analysis could help e-retailer to take decisions to sponsor a personalized product recommendation at a lower cost. Various influential factors are identified to carry forward the investigation to provide a significant product recommendation to customers. The proposed system is modelled to improve the efficiency and take the needful action for the betterment of the e-commerce system. The improvised version of the model can transfer shoppers into regular customers. The system also identifies that there is a correlation between the price and review attribute with the search of the product. In particular, the research aims to develop a framework that recognizes the search pattern associated with the purchase for a particular attribute. Its goal is to find out the personalized products and to stimulate the effective advertisement or product recommendation flow to the customers. The system's future work includes incorporation of some more influential factors like sales, offers, deals, and discount to improve the model efficiency. Furthermore, the implemented model can be tested on multiple durable goods that consumers looked for on various websites.

Li and Law (2020) investigated contributions of big data research in tourism through network visualization. The authors performed research on datasets from Web of Science, (1) focusing on research from tourism and hospitality, and (2) comparative study of tourism with other computer science disciplines. They utilized network analysis concepts to better visualize the big data research in tourism and hospitality and suggested that still there are some challenges left to achieve in the field of big data such as maintaining privacy, quality of data related to tourism, hospitality and other domains.

MOTIVATION

The field of social network analysis is increasing rapidly from analyzing trends of Facebook, Twitter, LinkedIn to visualizing real world networks that connect and transfer friendships, information, money and power, to analyze their relations about culture, politics, history and several other interesting properties.

Paranyushkin (2019) proposed an open source tool that uses graph detection algorithms on any text to identify main topics and relationships among them represented by different topical clusters. It further elaborates about analyzing the level of bias in discourse and generating insights from the structural gaps in the graph helpful to propose potential new connections between different themes and ideas. This real world example inspired us to work on a project - Social Network Analysis for Games of Thrones. The compound and fascinating narrative dynamics of GoT is a perfect platform lodestone for network analysis and natural language processing, to use two powerful and dynamic tools throughout our analysis. The motivation articulated with the thought to search for some interesting answers - Jon Snow, Daenerys Targaryen or Tyrion Lannister? Who is the most popular and lovable character in TV Series *Game of Thrones* based on R.R Martin's hugely popular book series *A Song of Ice and Fire*. The inquisitiveness to analyze, do mathematics and science show similar results as that concluded in series? The thought to use social network concepts as the basis to compare our results with public polls was the basic approach targeted.

The tools of social network analysis give you a major advantage when studying a population such as:

(1) It helps to find out how one person is connected or disconnected from people, groups or trends in a population that helps analysis much more realistic
(2) Social network analysis allows us to study how individuals segregate their energies between different social groups over time and helps strangers to form distinct friendship groups.
(3) Practical questions that can get quantitative answers and new insights, patterns, trends with social network analysis that just weren't possible before. Social Network Analysis opens an exciting range of new options. For example: Colonial American History, how do age and race affect friendships among students in school
(4) It gives us knowledge about the evolution of the character's importance and their equations over the time span.

MAIN FOCUS OF THE ARTICLE

The purpose of the work presented in this paper is to explore the features associated with the show- Game of Thrones, its popularity, various communications between the houses across various seasons, topological analysis of its social network of characters and helping in deeper understanding of the properties of social network analysis and graph theory. It proposes a visual analysis of networks through graphs to measure the strength of relationships and how information flows between people, groups, communities and other social entities.

The work tries to overcome the shortcomings in the existing works and present a more accurate practical study by considering variation of characters popularity over a series of seasons using graph theory and its concepts. In the end the results are validated by performing a comparative study with the public polls stating the accuracy of the derived results.

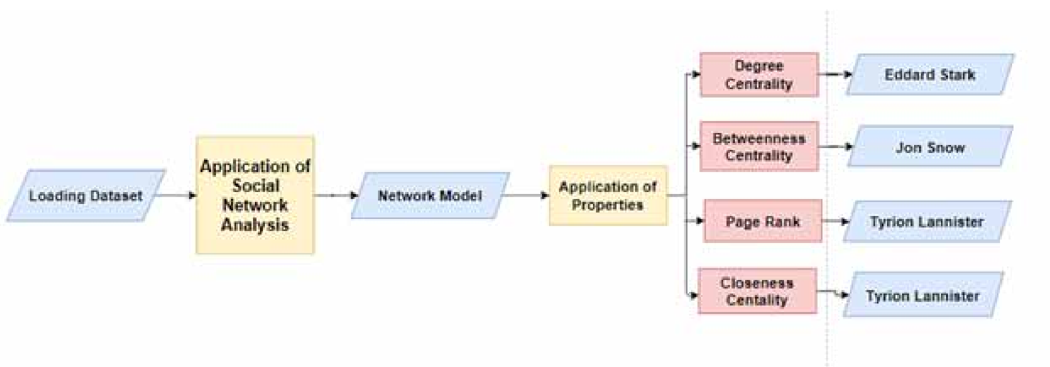
Game of Thrones, a popular TV series which is based on the novel series entitled ***"A Song of Fire and Ice",*** written by George R. R. Martin appealed a million audiences. The plot of the different seasons of Game of Thrones revolves around the following four most important houses: Starks, Lannister, Targaryen and Baratheon.

To enhance the influence of our analysis, we further extended our analysis to find the ranking of various characters and find the most liked character of the series and explore how this ranking gets evolved over the series. It provides an insight into the trend of liking changes, discovering patterns, showing relations between various houses vary and to what extent.

METHODOLOGY

The proposed approach starts with loading the dataset and then creating the social network. Various properties of social network analysis are applied on the data to formulate the connectedness score of the characters in the series. The key characters along with the prevailing interactions are identified, and a network is formed between them based on the interactions that exist between those actors. Figure 1 represents the proposed architecture, the approach followed along with its resulting output when analyzed after the application of different centrality measures.

Figure 1. Proposed Architecture of Games of thrones network analysis



The task was implemented in NetworkX which is a package in Python used for designing network, network algorithm creation and analysis. NetworkX is a Python library for studying graphs and networks. It provides data structures with parallel edges and self-loops to represent several types of networks, or graphs, including simple graphs, directed graphs, and graphs (Hagberg et al., 2008). NetworkX is a powerful tool for methodical calculations as it is easy to use and provides flexibility. A network graph helps us to reveal patterns and detects anomalies in our data. With NetworkX it is possible to draw small and complex networks. It uses the functionalities of matplotlib library to provide easy and interactive visualizations. Unlike many other tools, it can handle data on a scale relevant to modern problems.

It makes visualization easy and provides hidden insights, patterns, outliers that text output results fail to provide. It provides features like building of network models, developing new network algorithms, network drawing, and much more. The nodes and edges used here represent the source and connection between these sources and targets. With NetworkX we can develop quick visualizations of graphs. It can even support large volume datasets and is capable of designing and analyzing complicated networks as well.

Weights are generally assigned to the links. The edge can be directed or undirected depending on the user but in this case we are dealing with an undirected graph. The result of application of this method results in the creation of a network which is being analyzed further. An undirected graph is created showing the connections between the characters.

Loading the Data

The dataset named 'A Song of Ice and Fire 50 edges ONA' was taken from Kaggle. The dataset comprises different files for different seasons containing the list of characters (Nardaci, 2019). It contains the score of how the characters are linked to any other character over the series, also the number of interactions that exist for a particular character. Figure 2 represents the data structure that is used to perform the analysis. The DataFrame comprises of the following columns: Source, Target, Type (Directed or Undirected) and weight of the edge. Source and target are the characters which are represented by the nodes which are connected by an edge that depicts the number of interactions among the characters.

Figure 2. Snapshot of Dataset

	Source	Target	Type	weight
0	Addam-Marbrand	Jaime-Lannister	Undirected	3
1	Addam-Marbrand	Tywin-Lannister	Undirected	6
2	Aegon-I-Targaryen	Daenerys-Targaryen	Undirected	5
3	Aegon-I-Targaryen	Eddard-Stark	Undirected	4
4	Aemon-Targaryen-(Maester-Aemon)	Alliser-Thorne	Undirected	4

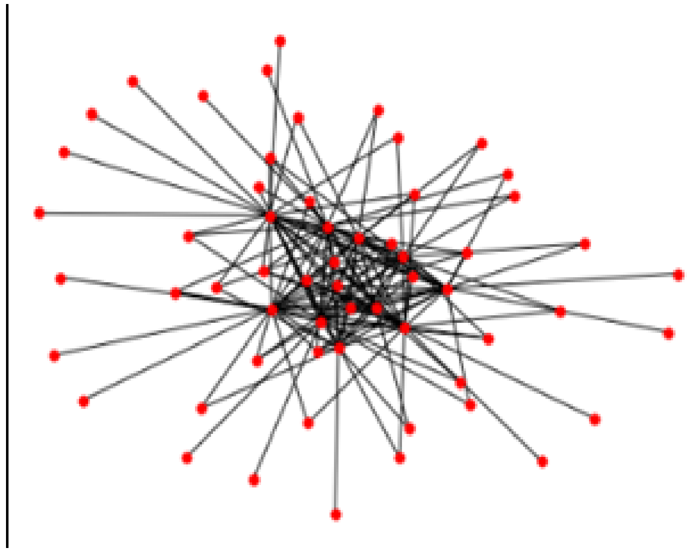
To create a network the first step is to create an empty graph using NetworkX. Post that nodes and edges are added to the empty graph using inbuilt methods of add_edge() and add_node().We can add one node with the method add_node().As soon as the nodes and edges are connected on the graph wwe generate the path graph with linearly connected nodes. Now our social network graph is ready to be analyzed.

The social network, shown in Figure 3 consists of 59 vertices and 183 edges. The nodes are representing the show's characters. The vertices correspond to the relationships between those characters. The multifaceted structure of the network reflects the linking plotlines of the story (Bag et al., 2019). If there is an existing edge between the nodes it does not mean that they have a direct connection or interaction between them; it may be due to mutual connections. Algorithm 1 describes the workflow of the algorithm used for the analysis.

Centrality Measures

Centrality metrics calculate a node's relevance for different relevant meanings or analysis. The importance of a node in a social network can be derived from how central a node is. Centrality measure is an important consideration in social network analysis when the central nodes are not easily visually identifiable. There exists different metrics that measure the importance of a node in a network (Beveridge & Shan, 2016). Pedroche et al. (2016) stated that centrality measures are also used to rank nodes based on their position in the social network. We first measured the importance based on the number of neighbors i.e. the connections of a particular node in the network. Measures like degree centrality, betweenness centrality, closeness centrality and page rank were used to determine importance of a particular node in the network. Betweenness centrality is most preferred of all these when using because you can easily change the betweenness centrality algorithm to measure the other three centrality metrics, but here we will be using all the metrics and then compare the results. The

Figure 3. Social Network of Game of thrones where nodes are represented as red nodes connected by edges



Algorithm 1: Proposed Algorithm for implementing SNA properties on Game of Thrones

Input: Nodes: Characters of show, Edges: Connections between characters
Output: Relationship between characters, Most popular character, Variation of popularity of characters over seasons
Step 1: Consider nodes n_1 to n_i and edges v
for each node i¬1 to i¬n
Calculate degree centrality to get significance of a node taking into consideration number of connections,
Calculate betweenness centrality for analyzing communication between nodes
Calculate degree centrality importance of a node,
Calculate pagerank centrality to rank the nodes
Step 2: Generating graphs on the basis of centrality measures score calculated
Step 3: Comparing the scores of different centrality measures to find the most popular node over the series
Step 4: Validating results after comparing them with public polls

higher the value of these measures (Hagberg et al, 2008) greater is the importance of that node in a network, except in closeness centrality where smaller values have more importance.

Degree Centrality

It is a measure that tells us the extent to which a node is connected to other nodes in the graph (Hansen et al, 2011; Hexmoor, 2015; Beveridge & Shan, 2016; Lv et al., 2019). It can be defined as the number of connections a node retains indicating significance of a node within a network. It is a property of social network analysis which represents the number of edges from a node. The major drawback of this property is that it does not consider indirect contacts. Since, degree centrality takes only direct connections into account when calculating the score, so when the distance between two nodes increases the score remains unaffected. A node with a degree centrality of 15 would mean that the node has fifteen connections whereas a node having only one edge would have a degree centrality of 1.

Degree centrality is a good measure of a node's overall connections, but it won't necessarily mean a node's role in linking others or how central it is to the main group. There is a belief that the more connections a node have the better, but it is not the number of connections that matters in the case of social networks, but whether those connections would lead to even more connections with the nodes. It is a useful measure to determine popular nodes or individuals.

Figure 4 shows the result of the top 10 centrality score of character nodes. Figure 6 compares the degree centrality of characters and as observed in the results from Figure 5, Eddard-Stark has the highest centrality score of 0.269 making him the most loved character of the first season. Similarly we have Tyrion Lannister, Jaime Lannister for the seasons 2, 3 and 4. Also in Figure 5 it can be seen that Ned Stark is connected to the highest number of people.

Figure 4. Top 10 degree centrality score of the characters

```
Season 1: [('Eddard-Stark', 0.2693877551020408), ('Robert-Baratheon', 0.20408163265306123), ('Tyrion-Lannister', 0.18775510204
081636), ('Catelyn-Stark', 0.17551020408163268), ('Jon-Snow', 0.1510204081632653), ('Robb-Stark', 0.14285714285714288), ('Sans
a-Stark', 0.14285714285714288), ('Bran-Stark', 0.1306122448979592), ('Cersei-Lannister', 0.12244897959183675), ('Joffrey-Barat
heon', 0.12244897959183675)]
Season 2: [('Tyrion-Lannister', 0.2054263565891473), ('Joffrey-Baratheon', 0.1821705426356589), ('Cersei-Lannister', 0.1666666
6666666666), ('Arya-Stark', 0.15503875968992248), ('Stannis-Baratheon', 0.1434108527131783), ('Robb-Stark', 0.1356589147286821
6), ('Catelyn-Stark', 0.12790697674418605), ('Theon-Greyjoy', 0.12403100775193798), ('Renly-Baratheon', 0.12015503875968991),
('Bran-Stark', 0.11627906976744186)]
Season 3: [('Tyrion-Lannister', 0.19536423841059603), ('Jon-Snow', 0.17218543046357615), ('Joffrey-Baratheon', 0.1655629139072
8478), ('Robb-Stark', 0.16225165562913907), ('Sansa-Stark', 0.15894039735099338), ('Jaime-Lannister', 0.1490066225165563), ('C
atelyn-Stark', 0.12582781456953643), ('Cersei-Lannister', 0.12582781456953643), ('Arya-Stark', 0.12251655629139073), ('Stanni
s-Baratheon', 0.10264900662251655)]
Season 4: [('Jaime-Lannister', 0.23443223443223443), ('Cersei-Lannister', 0.21978021978021978), ('Brienne-of-Tarth', 0.1025641
0256410256), ('Tyrion-Lannister', 0.09523809523809523), ('Margaery-Tyrell', 0.09157509157509157), ('Sansa-Stark', 0.0879120879
120879), ('Tommen-Baratheon', 0.0879120879120879), ('Samwell-Tarly', 0.07326007326007326), ('Stannis-Baratheon', 0.07326007326
007326), ('Petyr-Baelish', 0.0695970695970696)]
Season 5: [('Jon-Snow', 0.1962025316455696), ('Daenerys-Targaryen', 0.18354430379746836), ('Stannis-Baratheon', 0.148734177215
18986), ('Tyrion-Lannister', 0.10443037974683544), ('Theon-Greyjoy', 0.10443037974683544), ('Cersei-Lannister', 0.088607594936
70886), ('Barristan-Selmy', 0.07911392405063292), ('Hizdahr-zo-Loraq', 0.06962025316455696), ('Asha-Greyjoy', 0.05696202531645
5694), ('Melisandre', 0.05379746835443038)]
```

Figure 5. The degree centrality of Ned Stark is the highest as he is connected to the highest number of people (Season 1)

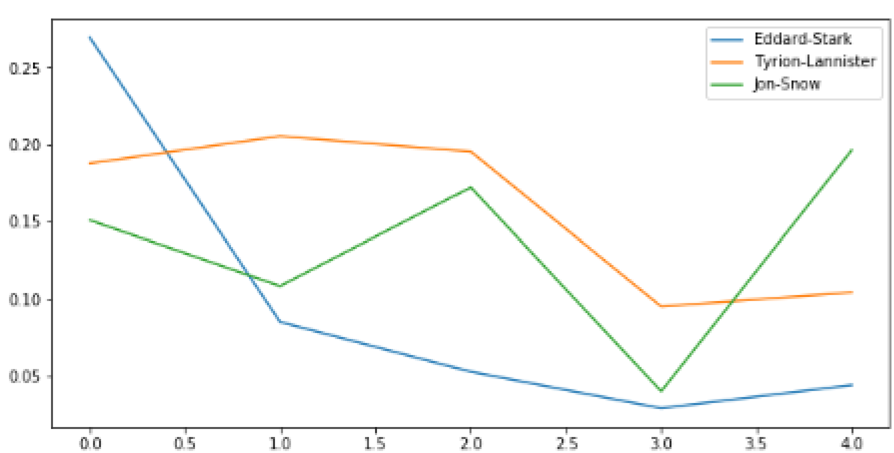
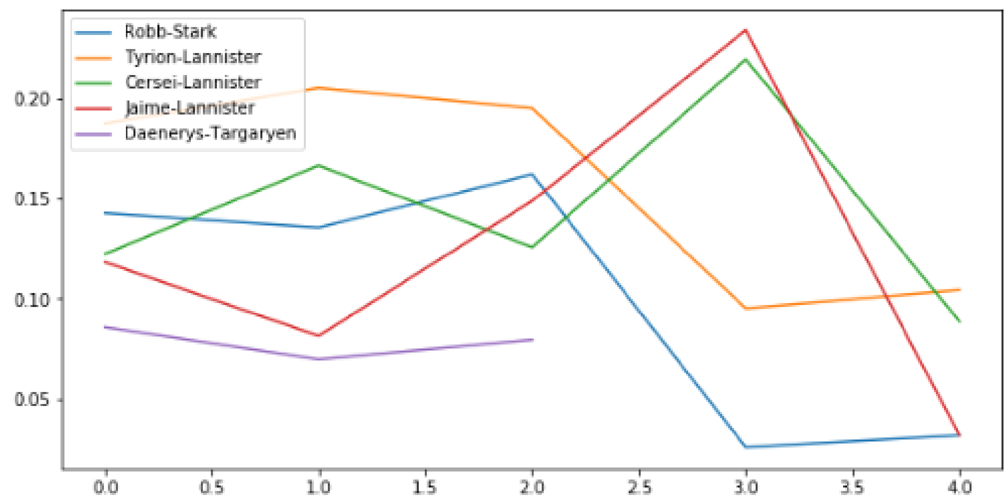


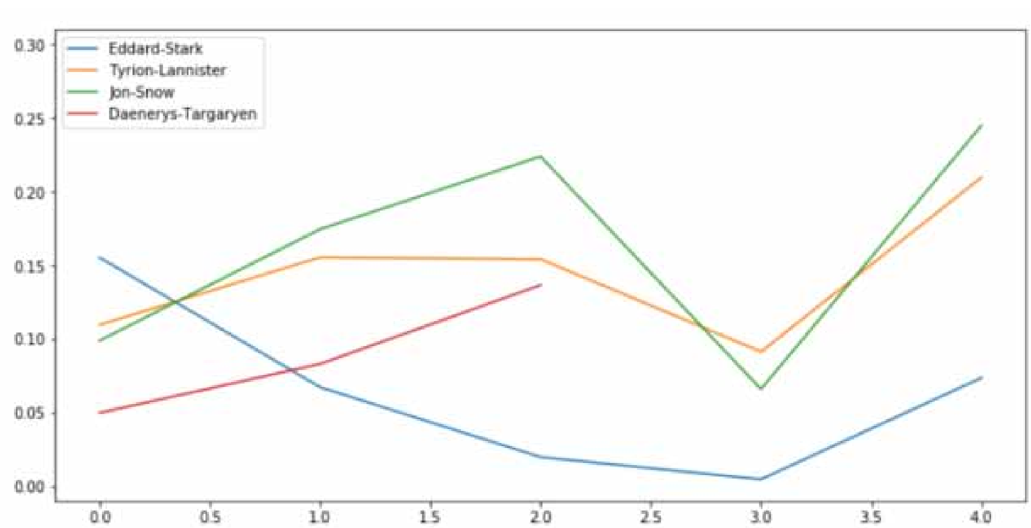
Figure 6. Comparison of the degree centrality of other characters



Betweenness Centrality

It refers to the number of times a node is visited when going from one node to another. Betweenness is useful for analyzing communication. High betweenness centralities mean that the node acts as intermediates between two different networks. From Figure 7, it can be seen that Jon Snow is having the highest score. In comparison to other measures, only this measure provides third ranking to Eddard Stark giving distinctive results.

Figure 7. Betweenness Centrality for First Season



In this measure, mean influence of a person's neighbors is almost identical regardless of the influence of the centered person, so almost the same influential people surround an individual which makes betweenness centrality a reliable measure (Goh et al., 2003). Unlike degree centrality, which is a count, betweenness is normalized as the proportion of all geodesics which includes the vertex under study. Betweenness centrality can be computed by calculating the ratio of total number of shortest paths passing through an edge to the shortest path between pairs of edges for all possible pairs of vertices in the graph. Equation (1) represents the betweenness score of node n.

$$\text{Betweenness centrality} = \text{å} \, \frac{\sigma jk\left(n\right)}{\sigma jk} \tag{1}$$

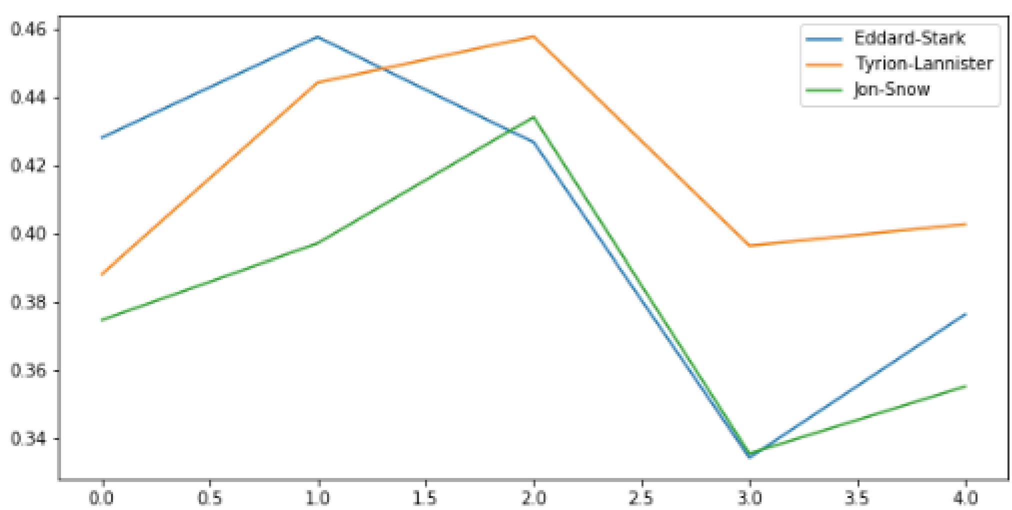
where,
ϭjk(n) is the total number of shortest path passing through n
ϭjk is the total number of shortest path from node j to k

Closeness Centrality

It is the average distance of a node to all other nodes. Nodes with a high score of proximity have the shortest distances to all other nodes. Closeness centrality tests the location of each entity in the network from a different point of view from the other network metrics, measuring the average distance in the network between each vertex that exists in the network (Hansen et al, 2011; Hexmoor, 2015; Beveridge & Shan, 2016; Lv et al., 2019). It is considered that the nodes having a shorter distance to the other nodes would help dissemination of information productively across the network. When the distance to any other nodes is reduced it leads to an increase in the centrality score. A low closeness centrality score means that a person is connected directly from most other people on the network.

In Figure 8, it can be seen that the centrality of Eddard Stark is the highest in the beginning, but as the story progresses Tyrion Lannister steals the spot of being the favorite character leaving Jon Snow as the third most important character.

Figure 8. Closeness centrality



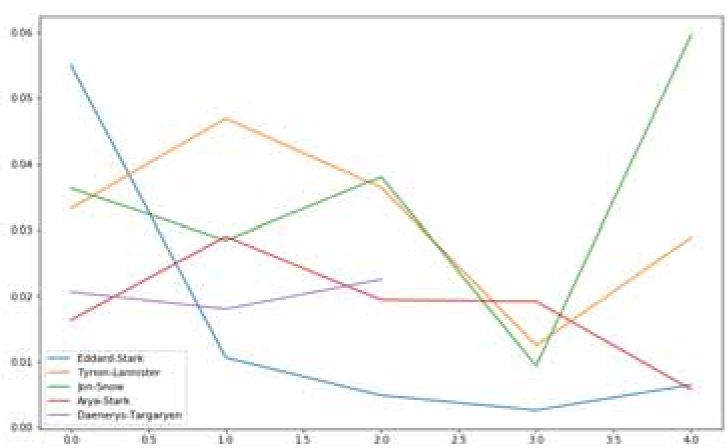
Page Rank Centrality

PageRank is a web-based ranking algorithm developed by google which is used here to assess a vertex within a network by considering the inlinks and outlinks of a node and then formulating a ranking of the nodes based on the score. A high PageRank page has many pages pointing to it (Ding et al., 2009). PageRank algorithm helps us to identify the top most popular nodes and has been proven useful in various situations like visualizing IT network activity, modeling the impact of SEO and link building activities.

Initially the pagerank algorithm sets the value of node as 1 and as the number of links corresponding to a node increases this value increases. In the end the node with maximum links will have maximum PageRank value and is considered as the most popular node.

From Figure 10 we can observe that the PageRank ordering is nearly identical to the degree centrality ordering, except Daenerys who jumps from twelfth place to fifth place. So PageRank correctly identifies Daenerys as one of the most important players, even though she has relatively few connections. It can be observed from Figure 9 though there is a tough fight between Tyrion Lannister and Jon Snow for the top spot but as the series progresses Jon Snow manages to be on the top rank in the end.

Figure 9. Score of Top 5 characters using PageRank Centrality



RESULTS AND DISCUSSIONS

The analysis of the network revealed the connections that exist, while identifying which nodes are at the most important in the network of Game of Thrones in terms of communication. As observed in Table 1, in most of the cases Jon Snow comes out to be the favorite character. A comparison was done to check the score. According to degree centrality, Eddard Stark is the most important character initially, but to find the most important characters in the end according to these measures, a comparison was made. The result of this comparison is shown in Figure 11 where Jon Snow outshines with a score of 0.059572 using PageRank centrality measure.

From the results it can be inferred that Jon Snow has evolved throughout different years and still manages to retain its top position in the end. Figure 12 shows the variation of Jon Snow's Popularity in different seasons over years.

Figure 10. Results of PageRank Centrality over different seasons

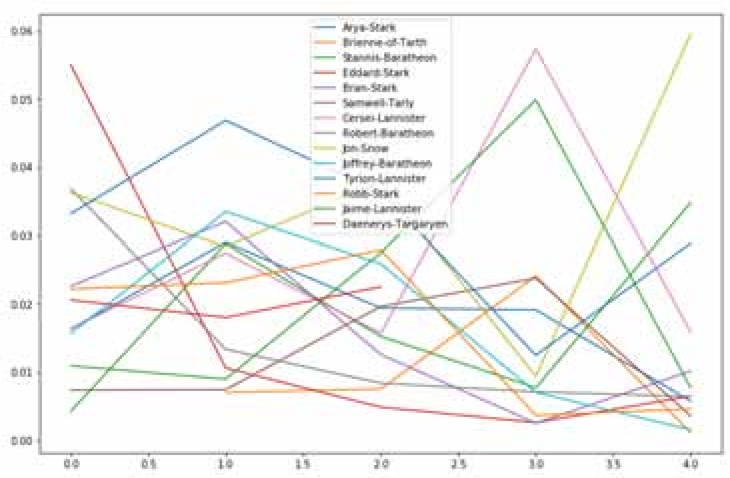


Figure 11. Most popular character using all measures at the end

```
Pagerank                    0.059572
Betweenness_centrality      0.205652
Degree_centrality           0.196203
Closeness_centrality        0.355056
Name: Jon-Snow, dtype: float64
```

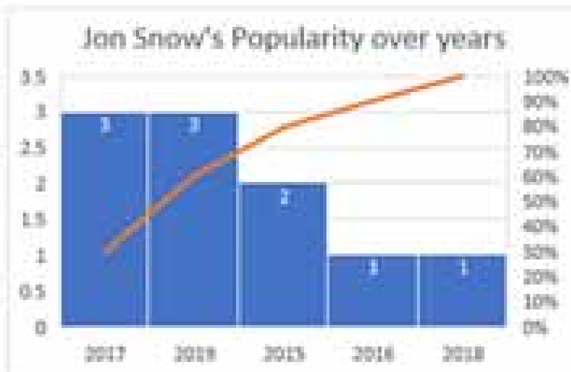
Table 1. Result of different centrality measures

Centrality measure	Score	Result
Degree Centrality	0.354838	Eddard Stark
Betweenness Centrality	0.205652	Jon Snow
Closeness Centrality	0.195364	Tyrion Lannister
PageRank Centrality	0.059572	Jon Snow

COMPARISON OF OUR RESULTS WITH PUBLIC POLL

The comparison of various polls and ranking by common websites has helped us to support the results that we have obtained. Jon Snow is found to be the most favourite character in the majority of the poll that is considered. Jon Snow was voted the most heroic character with a majority percentage of 29.9, and the most deserving character for the throne with 38.2%. Though he does not get the throne in the series but he still owns the throne of the favourite personality of the users. Eight out of ten sources have voted him as the top ranked character of the show. From the comparison of the derived results and the people's views it can be drawn that the result of betweenness centrality and page rank centrality are found to be more accurate than the other two measures which have concluded Jon Snow at the top position, whereas Ned Stark is on the seventh position with a vote count of 35.7% (Aquilina, 2019).

Figure 12. Variation of Jon Snow's Popularity in different seasons over years



CONCLUSION

The analytical model using social network analysis, made it easy to analyze the network of Game of Thrones. The idea was to demonstrate the implementation of network analysis using NetworkX to show adaptation of characters popularity and relationship of characters that are linked to each other and to what extent. After the application of various centrality measures and their comparison with the public polls, it is concluded that Jon Snow is on the top in most of the results by different measures performed and hereby considered the most lovable character of the show. Second place goes to Tyrion Lannister, but we do note that he underperforms in the PageRank centrality test when compared to his communication with other characters. Also, the work was able to find the transformation of Jon Snow's popularity over the seasons which can be further analysed to study the factors affecting his growth. The properties of SNA have helped us to find interesting facts about Game of Thrones, centrality scores of each character and the visualization helped in understanding the deviation of importance of characters throughout.

This quixotic application of network proved to be valuable in understanding the underlying trends in a network using network analysis. Concepts of SNA can further be used to explore various other properties of characters like biased, focused, dispersed and diversified characters in the graph network. A community can be detected which involves characters belonging to the same house and comparable traits which further help to draw interesting conclusions and predictions. The proposed method is further applicable to find association between other related subjects and web series, comparing story and direction in novel based movies, Regional and historical documentaries representations. This paper opens the scope for future research on recent trends,, identifying the structural gaps to enhance the potential recommender systems, measuring the level bias in structure discourse, sentiment analysis of reviews of people about the storyline, creating a directed graph by assigning weights to the edges that can help us to understand the depth of relationship that exists, dark web analysis and link analysis. Other Software Analysis tools like Gephi, NodeXL, Tulip, and Netlytic can also be used for analysis and visualization

REFERENCES

Akhtar, N. (2014, April). Social network analysis tools. In *2014 Fourth International Conference on Communication Systems and Network Technologies* (pp. 388-392). IEEE. doi:10.1109/CSNT.2014.83

Aquilina, T. (2019). *Here are the results of EW's ultimate Game of Thrones poll*. EW.com. Retrieved February 14, 2020, from https://ew.com/tv/2019/05/21/game-of-thrones-poll-results/

Bag, S., Tiwari, M. K., & Chan, F. T. (2019). Predicting the consumer's purchase intention of durable goods: An attribute-level analysis. *Journal of Business Research*, *94*, 408–419. doi:10.1016/j.jbusres.2017.11.031

Betweenness Centrality. (n.d.). Retrieved November 20, 2020 from https://en.wikipedia.org/wiki/Betweenness_centrality

Beveridge, A., & Shan, J. (2016). Network of thrones. *Math Horizons*, *23*(4), 18–22. doi:10.4169/mathhorizons.23.4.18

Bhatia, R. (2018). *Top 7 Network Analysis tools for Data Visualization*. Retrieved February 3, 2020, from https://analyticsindiamag.com/top-7-network-analysis-tools-for-data-visualisation/

Bonato, A., D'Angelo, D. R., Elenberg, E. R., Gleich, D. F., & Hou, Y. (2016, December). Mining and modeling character networks. In *International workshop on algorithms and models for the web-graph* (pp. 100-114). Springer.

Breja, M. (2017, August). *Social network analysis in question answering community*. Paper presented in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). doi:10.1109/ICECDS.2017.8390017

Chowdhury, T., Muhuri, S., Chakraborty, S., & Chakraborty, S. N. (2019). Analysis of adapted films and stories based on social network. *IEEE Transactions on Computational Social Systems*, *6*(5), 858–869. doi:10.1109/TCSS.2019.2931721

De, S. S., Dehuri, S., & Wang, G. N. (2012). Machine Learning for Social Network Analysis: A Systematic Literature Review. *IUP Journal of Information Technology*, *8*(4).

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, *60*(11), 2229–2243. doi:10.1002/asi.21171

Froehlich, D., Rehm, M., & Rienties, B. (2020). Reviewing Mixed Methods Approaches Using Social Network Analysis for Learning and Education. In *Educational Networking* (pp. 43–75). Springer. doi:10.1007/978-3-030-29973-6_2

Goh, K. I., Oh, E., Kahng, B., & Kim, D. (2003). Betweenness centrality correlation in social networks. *Physical Review. E*, *67*(1), 017101. doi:10.1103/PhysRevE.67.017101 PMID:12636633

Golbeck, J. (2015). *Introduction to social media investigation: a hands-on approach. In Syngress*. Elsevier Inc.

Hagberg, A., Swart, P., & Chult, S. D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab (LANL).

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). *Social network analysis: measuring, mapping, and modeling collections of connections. In Analyzing social media networks with NodeXL: insights from a connected world*. Elsevier Inc.

Hexmoor, H. (2015). Network Analysis. In *Computational Network Science* (pp. 15–19). Elsevier Inc. doi:10.1016/B978-0-12-800891-1.00002-0

Landherr, A., Friedl, B., & Heidemann, J. (2010). A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, *2*(6), 371–385. doi:10.1007/s12599-010-0127-3

Li, X., & Law, R. (2020). Network analysis of big data research in tourism. *Tourism Management Perspectives*, *33*, 100608. doi:10.1016/j.tmp.2019.100608

Lv, L., Zhang, K., Zhang, T., Bardou, D., Zhang, J., & Cai, Y. (2019). PageRank centrality for temporal networks. *Physics Letters. [Part A]*, *383*(12), 1215–1222. doi:10.1016/j.physleta.2019.01.041

Mitchell-Guthrie, P. (2016). How Santa's Workshop uses social network analysis. *The SAS Data Science Blog*. Retrieved February 2, 2020, from https://blogs.sas.com/content/subconsciousmusings/2016/12/16/how-santas-workshop-uses-social-network-analysis/

Morgan, K., Melendez-Torres, G. J., Bond, A., Hawkins, J., Hewitt, G., Murphy, S., & Moore, G. (2019). Socio-economic inequalities in adolescent summer holiday experiences, and mental wellbeing on return to school: Analysis of the school health research network/health behaviour in school-aged children survey in Wales. *International Journal of Environmental Research and Public Health*, *16*(7), 1107. doi:10.3390/ijerph16071107 PMID:30925676

Nardaci, R. (2019). A Song of Ice and Fire ONA. *Kaggle*. Retrieved February 2, 2020, from https://www.kaggle.com/ricardocnsoares/a-song-of-ice-and-fire-50-edges-ona

O'Malley, A. J., &Onnela, J. P. (2017). Introduction to social network analysis. *Methods in Health Services Research*, 1-44.

Paranyushkin, D. (2019, May). InfraNodus: Generating insight using text network analysis. In *The World Wide Web Conference* (pp. 3584-3589). doi:10.1145/3308558.3314123

Pedroche, F., Romance, M., & Criado, R. (2016). A biplex approach to PageRank centrality: From classic to multiplex networks. *Chaos (Woodbury, N.Y.)*, *26*(6), 065301. doi:10.1063/1.4952955 PMID:27368791

Serrat, O. (2017). Social network analysis. In *Knowledge solutions* (pp. 39–43). Springer. doi:10.1007/978-981-10-0983-9_9

Thomsen, M. S., & South, P. M. (2019). Communities and attachment networks associated with primary, secondary and alternative foundation species; a case study of stressed and disturbed stands of southern bull kelp. *Diversity (Basel)*, *11*(4), 56. doi:10.3390/d11040056

Xing, W., & Ghorbani, A. (2004, May). Weighted pagerank algorithm. In *Proceedings. Second Annual Conference on Communication Networks and Services Research*, 2004 (pp. 305-314). IEEE. doi:10.1109/DNSR.2004.1344743

Zhang, L., Guo, X., Lei, Z., & Lim, M. K. (2019). Social network analysis of sustainable human resource management from the employee training's perspective. *Sustainability*, *11*(2), 380. doi:10.3390/su11020380

Manvi Breja is an Assistant Professor in Computer Engineering Department. She has completed her B.Tech from Lingayas Institute affiliated by MDU Rohtak, followed by M.Tech from YMCA. She has been awarded with Dean Merit certificate for securing first position in M.Tech. She has around four years of teaching experience as an Assistant Professor. Currently, she is pursuing PhD from NIT Kurukshetra in the field of Information Retrieval and Data Mining. She has also completed NPTEL course on "Social Networks" with rank in top 5% and "Cryptography and Network security" with a score of 90%. Her key areas of interest are Information Retrieval, Data Mining, Natural Language Processing.

Himanshi Bhatia is currently pursuing Bachelor's of Technology in Computer Science from NorthCap University, Gurugram. She has a background in database maintenance and holds keen interest in the area of data analytics and visualisation. During her graduation, she served on various significant positions in National service scheme-NCU chapter like Event head and Creativity head. Also, she has received membership of ASQ. Pronto she'll be joining Real Time Data services, Gurugram and will work as a software quality analyst. Her consultancies with industries during the internship period have been in the areas of machine learning and NLP, emerging technology in furtherance with technological innovation in daily life.

Dollie Juneja is currently pursuing her B.Tech in Computer Science from NorthCap University. She is keen about the new development in the field of text and web intelligence analysis and has strengthened her skills in the tableau - analytics tool. She has pursued various courses in python and machine learning. She has worked on various projects that involve deep knowledge of ML python and NLP. She is a central level representative of Student Chapter of Association of Computing and Machinery (NCU) and student ambassador of The NorthCap University and Connexio. She is a core member of the learning and implementation domain in Google Student's Club and is associated with various social club's like Rotary and Forti fectum. Also she served as an operations head in International Conference in Computational Intelligence and Data Science. Shortly, she will be joining Deloitte.