5-2017

Generation and Analysis of a Social Network: Hamlet

Preston Evans

Generation and Analysis of a Social Network: *Hamlet*

An Undergraduate Honors College Thesis
in the
Department of Computer Science
College of Engineering
University of Arkansas
Fayetteville, AR

By

Preston Evans
pdevans@uark.edu

May, 2017
University of Arkansas

This thesis is approved by:

Thesis Advisor:

_____

Susan Gauch, Ph.d

Thesis Committee:

_____

Dr. Jia Di

_____

Dr. Qinghua Li

Abstract

This paper examines the generation and analysis of a social network produced from Shakespeare's play *Hamlet.* An XML file of *Hamlet* was parsed to extract the characters within the play and also identify when the characters appeared within the same scene. After parsing the speakers and the connections between characters, a network graph was generated that displayed all the characters in *Hamlet,* represented by nodes, and edges that represented the connections between characters as measured by their scene co-appearance. The results of the network graph were then compared to a published social network for *Hamlet* created by hand. The two social networks showed strong similarities in character centrality but also showed differences in the number of character nodes and edges. In addition to the case study, we present a suite of tools that provide a framework for computational analysis of future plays.

Contents

1 Introduction

1.1 Background

Social networks have commonly been used as a key method in sociology and the examination of interconnecting relationships between people. By observing these relationships, analyses can be made on several social networks in order to determine patterns or behaviors demonstrated between two people or different groups of people. The results of these analyses can then be used to make accurate observations of relationships such as which person has the fewest connections to other people or which person appears to have the strongest centrality in the entire social network. These can be used to identify individuals on the fringes of society and those whom have the most potential influence on others.

Our work investigates the application of social network analysis to dramatic literature. We consider dialogue between characters as one way of measuring their inter-connectedness. However, plays consist of not only the dialogue and interactions between characters but also stage directions that are used as guidelines for a theatrical performance. The dialogue between the characters is important in determining the relationship between the speaker and the characters listening to the speaker. In order to retrieve this information, parsing techniques can be used to separate the stage directions from the characters' dialogue. Also, a play's text will often announce who is speaking and to whom he/she is speaking. This allows for a more accurate separation of character interactions and stage directions.

By parsing of character information from plays, social networks can be constructed from this information in order to demonstrate the relationships of characters. By utilizing the techniques of social networking analysis with the information of character dialogues in plays, new analyses can be discovered about the connections between a play's characters as well as the patterns demonstrated across genres of other plays.

1.2 Goals

This paper attempts to present a social networking graph that displays the interconnecting relationships between the characters in Shakespeare's play: Hamlet. The resulting graph will provide insight into the centrality or importance of different characters based on their number of interactions with other characters. The graph will also present relevant information such as number of characters represented as nodes, number of interactions between all characters represent as edges, and the node degree of important characters.

Along with the social network graph, this paper demonstrates a framework for computational analysis of plays that was created to parse the necessary information from a suitable file format of Shakespeare's play: Hamlet. The program successfully identifies speaking characters as opposed to stage directions, creates connections between a speaking character and those to whom he/she was speaking. It also produces these connections in a format from which a social network graph can be generated as input to the Gephi social network analysis and visualization tool.

For future goals, we plan to apply the same techniques used to create a social network graph for Hamlet on other Shakespeare plays. Ideally, the program demonstrated in this paper could be used to parse other Shakespeare plays in order retrieve their character relationships and generate new social network graphs. By doing so, further observations and comparisons could be examined to identify commonalities and differences plays in different genres, e.g., tragedies, or histories, or comedies. Additionally, we could study trends in dramatic social networks over time and across different playwrites.

2 Related Work

2.1 Social Network Analysis

The understanding of social network analysis methods and tools is important in focusing on how social networks should be represented and analyzed. Carl Butts discusses many approaches to social network analysis including social network graph representation and information gathering [6]. Campbell, Dagli, and Weinstein have also produced a work pertaining to social network analysis methods with heavy emphasis on community detection[7]. Freeman's work looks into the use and problems that arise when analyzing the centrality of social networks[9]. These works provide insight into how social networks should be represented, how a social network should be analyzed, and what key features should be highlighted from a social network.

2.2 Utilization of Social Network Analysis in Different Fields of Study

Social network analysis has been utilized across several fields of study such as sociology, economics, and marketing for research purposes and relationship analytics. Grunspan, Wiggins, and Goodreau utilized social network analysis in an education environment to better understand the relationships established between students in the classroom[10]. Borgatti, Mehra, Brass, and Labianca examined how social sciences have employed social network analysis in order to observe interactions and connections between people and social groups[5]. Jackson studied the importance of a social network's impact on socioeconomics and the increasing use of social networks in creating better economic decisions[11]. Because social networks have been shown to be powerful tools in the observation of relationships, the use of social network analytics has risen and expanded across many fields of study.

2.3 Utilization of Social Network Analysis in Literature

Recently social network analytics have been utilized in the study of character relationships and interactions within literature. Agarwal, Corvalan, Jensen, and Rambow generated a social network of the novel *Alice in Wonderland* in order to study and visualize the roles of the different characters through social events within the novel[4]. Kydros and Anastasiadis published a work which analyzed and formed a social network based on character interactions in the Modern Greek novel *The Great*

Eastern[12]. Kydros and Anastasiadis also delved into how the methods and practices of social network analytics have been applied in literature for the purpose of visualizing character relationships[12]. Elson, Dames, and McKeown's paper examined how social networks could be extracted from literary fiction through the detection of character dialogue and interactions[8]. Many researchers are now utilizing social network analytics in order to examine and visualize character relationships and interactions within literature. However, these social networks are generated manually from the manuscripts, a time-consuming and tedious task.

3 Implementation and Methods

3.1 *Hamlet* Data

In order to better parse the plaintext of *Hamlet*, a suitable file format that could easily distinguish characters from scenes and stage directions needed to be used. An XML format of *Hamlet* provided the necessary distinctions of characters resulting in more accurate parsing of the play. The XML format of *Hamlet* used in this paper was retrieved by the public digital library "ibiblio" which is run by School of Information and Library Science and the School of Journalism and Mass Communication at the University of North Carolina at Chapel Hill[1]. The play is annotated with labels for each character whom was the current speaker as well as labels for plaintext that was considered stage directions. An example from the *Hamlet* XML file can be seen in Figure 3.1.

```
<STAGEDIR> Enter Horatio and Marcellus </STAGEDIR>
<SPEECH>
        <SPEAKER>Horatio</SPEAKER>
                <LINE>Friends to this ground</LINE>
</SPEECH>
<SPEECH>
        <SPEAKER>Marcellus</SPEAKER>
                <LINE>And liegemen to the Dane.</LINE>
</SPEECH>
```

Figure 3.1: An XML Snippet from Hamlet

In preparation for future research, we also downloaded all 37 Shakespeare plays from the same site.

3.2 Tools

We selected JDOM [2] as the parsing tool. The open source JDOM libraries used in this paper were downloaded from their website. JDOM was selected as the parsing tool for its user friendly API, ability to be easily implemented into a Java program, and its open source libraries. Using JDOM, we are able to read an XML file and parse the character currently speaking in a given scene

The social network analysis tool used for this paper was Gephi. The Gephi software used for this paper was downloaded from their website[3]. Gephi is an open source network visualization and analysis software. Gephi was selected because it was able to import the file format output by the parsing program. Also, Gephi was able to output the desired statistics of the generated social network for *Hamlet* including number of edges, degree of each node, and node centrality.

Additionally, Gephi has the capabilities to export a social network graph in either PNG, PDF, or SVG formats. Before exporting a social network graph, Gephi can change several settings of the graph including color, font, and scale of nodes and edges. Gephi also has a window to view the data tables of the network graph that can add, delete, and edit node and edge properties including labels and weights. Lastly, Gephi can determine the average degree, average path length, graph density, modularity, and connected components for a network graph.

3.3 Program Implementation

In order to retrieve the necessary character interactions in the play *Hamlet*, a program was designed to parse through an XML file of *Hamlet* and output the connections of the characters in the desired file format. The program utilized the JDOM libraries API in order identify a scene in the play, the character speaking within a scene, and the name of the character.

Initially, the program was designed to parse when a new scene occurred and record all speaking characters in the scene. However, after the program finished parsing and connecting all speaking characters within the scene, it became apparent that some characters had connections to other characters who were not currently on stage during the scene. This information was contained in stage directions for the scene that signaled for some characters to either leave or enter the scene. The program was then redesigned to parse when a new stage direction had occurred and record all speaking characters between each stage direction. By doing so, the program is able to more accurately parse which speaking characters were present in a scene.

When the program parsed a speaking character, it would record the character's name until a new stage direction was given. After parsing the new stage direction, the program creates connections between all characters that have been previously recorded. A connection between characters is be indicated by the character's name and other character's name separated by a semicolon. An example being "Hamlet;King Cornelius". The program creates these connections until every character had a

connection to every character recorded. After all characters have been connected, the program empties all previously recorded characters and prepares for the next block of speaking characters. This process continues until the end of the play.

After the program finishes creating connections for each speaking character, the resulting output of the program is a list of every character connection in a CSV format. In order to properly import the result of the program to Gephi, the outputted CSV file needs to indicate that the character before the semicolon is the source and the character after the semicolon is the target. The program is able to indicate speaking characters present within a scene and form a connection that can then be processed by a social networking tool.

3.4 Gephi Utilization

Gephi's main purposes for this paper are to generate a social network visualization of *Hamlet*, analyze features of the social network, and export the social network into a suitable format. Gephi was used separately from the parsing program but future implementations could combine the parsing of a play and creation of a social network within the same program.

In order to generate the social network of *Hamlet*, Gephi first imports the corresponding CSV file produced by the parsing program. When the CSV file was imported, Gephi created nodes for each character identified within the CSV file. Next, edges are created between the source and target characters found in the CSV file. Because these connections were not labeled as directed or undirected edges, Gephi automatically sets each connection as a directed edge.

After importing the CSV file, Gephi produces a rough social network for the characters in *Hamlet*. To better visualize which characters were more central in the social network, a tool was used in Gephi to centralize character nodes whom had more edges connected to multiple other character nodes. After several iterations of centralizing the social network, character nodes were given labels and moved around manually in order to provide a better visualization of the social network.

4 Results and Analysis

4.1 Social Network of *Hamlet*

After importing the CSV file into Gephi, nodes and edges were created between each speaking character in *Hamlet*. The social network was then centralized around characters with stronger node degrees and slightly reorganized in order present a more clear visualization of the social network. Figure 4.1 presents the final social network.

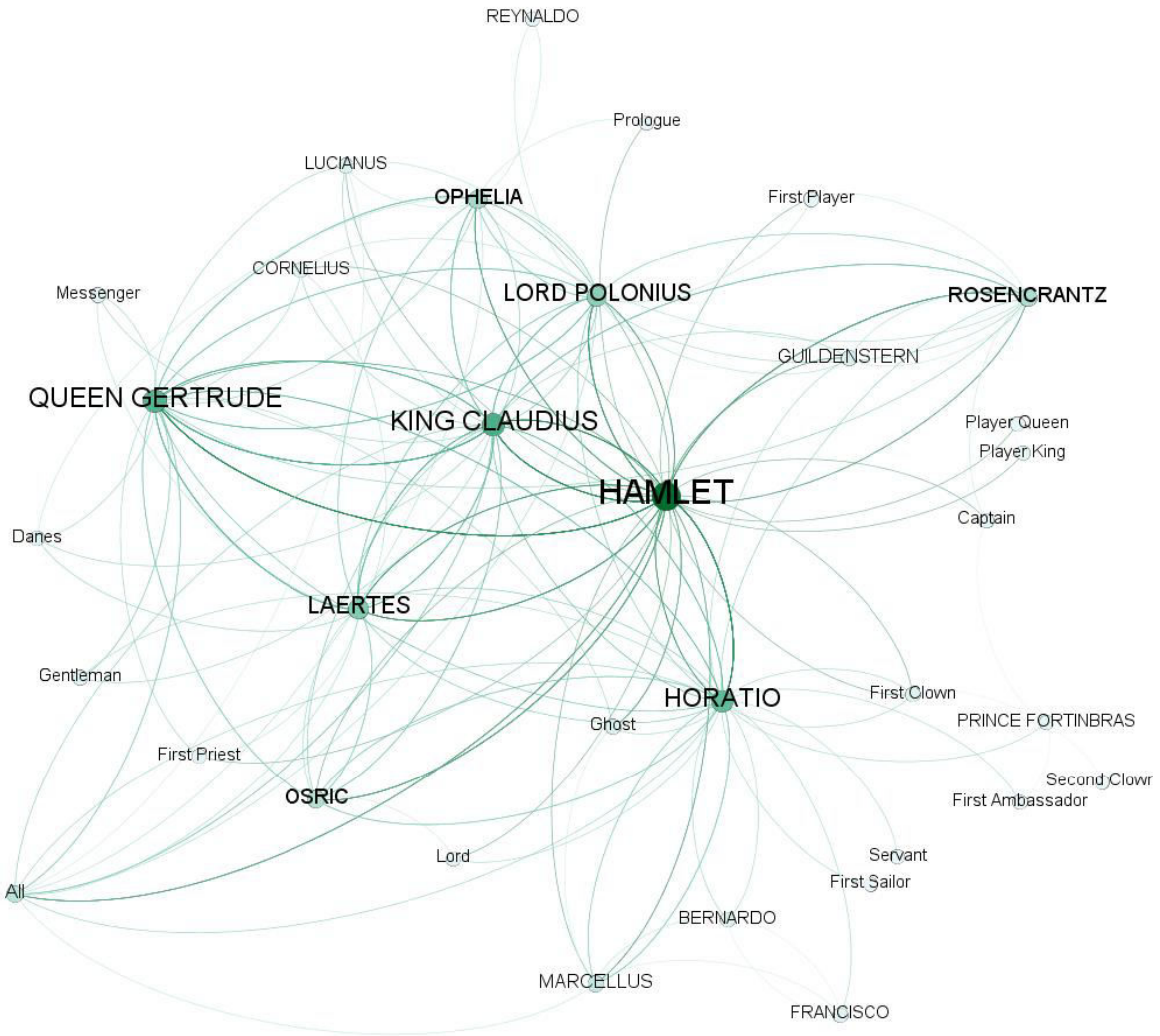

Figure 4.1: The social network for Hamlet.

4.2 Social Network Analysis of *Hamlet*

When analyzing the centrality of the social network, Hamlet, King Claudius, Queen Gertrude, and Horatio appear to be the most central nodes within the social network. This would make sense considering these four characters are rather important in the story. Also, characters with fewer speaking roles were less centralized in the social network that could correlate to their importance within the play. The social network was able to identify the major characters within *Hamlet* based on the frequency the character's speaking parts.

The generated social network indicated that there was a total of 34 nodes representing characters and 244 edges representing speaking connections within the graph. The shade of each node represents the degree of the node, i.e., the darker the shade the higher the degree and vice versa. The thickness of an edge represents the number of times the character spoke or was spoken to by the other character. Figure 4.2 gives a listing of the degrees for each character node.

Figure 4.2: Characters and node degree of characters

Id	Label	Degree
BERNARDO	BERNARDO	6
FRANCISCO	FRANCISCO	4
HORATIO	HORATIO	48
MARCELLUS	MARCELLUS	13
KING CLAUDIUS	KING CLAUDIUS	55
CORNELIUS	CORNELIUS	5
LAERTES	LAERTES	36
LORD POLONIUS	LORD POLONIUS	33
HAMLET	HAMLET	89
QUEEN GERTRUDE	QUEEN GERTRUDE	52
All	All	16
OPHELIA	OPHELIA	25
Ghost	Ghost	6
REYNALDO	REYNALDO	2
First Player	First Player	3
ROSENCRANTZ	ROSENCRANTZ	21
Prologue	Prologue	2
Player King	Player King	2
Player Queen	Player Queen	2
LUCIANUS	LUCIANUS	6
GUILDENSTERN	GUILDENSTERN	8
PRINCE FORTINBRAS	PRINCE FORTINBRAS	3
Captain	Captain	3
Gentleman	Gentleman	4
Danes	Danes	5
Servant	Servant	1
First Sailor	First Sailor	1
Messenger	Messenger	3
First Clown	First Clown	5
Second Clown	Second Clown	1

First Priest	First Priest	3
OSRIC	OSRIC	20
Lord	Lord	3
First Ambassador	First Ambassador	2

According to Figure 4.2, Hamlet has the highest node degree within the social network, where the degree represents the number of times Hamlet spoke to or was spoken to by other characters. This makes intuitive sense considering Hamlet is the main character within the play. Further observations show that King Claudius, Queen Gertrude, and Horatio have similar node degrees of roughly 50, second only to Hamlet. This seems to indicate that they are also characters of primary importance, just behind Hamlet. Those with very low node degrees more than likely had only one appearance within the entire play.

Additionally, the in-degree and out-degree were calculated through Gephi which represented the number of times a character spoke towards another character and a character was spoken at by another character in a scene. For example, the in-degree for Hamlet in this social network is 21 and the out-degree for Hamlet is 68. Thus, Hamlet is spoken to about three times as often as he speaks himself. Also, Gephi calculated that the average path length for the social network is 2.253 which is based on the centrality of every node. Lastly, Gephi calculated the closeness centrality, harmonic closeness centrality, and betweenness centrality for each node in the network graph. Across each of these properties, Hamlet has the highest centrality values with a closeness centrality of 0.74359, a harmonic closeness centrality of 0.827586, and a betweenness centrality of 301.414155.

4.3 Comparisons

For a deeper analysis of the accuracy of the social network of *Hamlet*, it was beneficial to compare a social network of *Hamlet* created by an English professor, Moretti[13], (Figure 4.3) and the social network of *Hamlet* generated through computational parsing of the play's text and network graph manipulation. Upon initial comparisons, there were strong similarities between the two social networks and some differences.

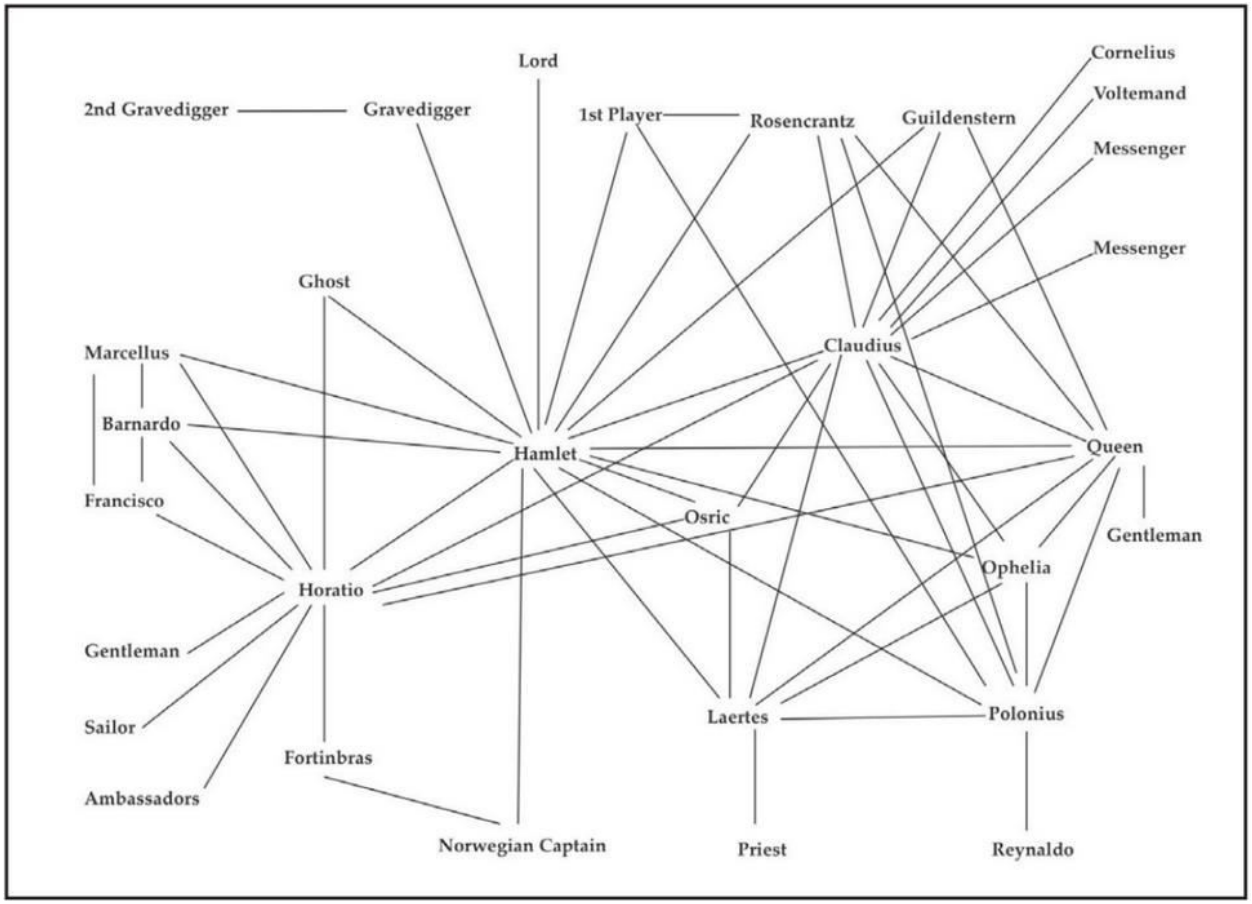

Figure 4.3. Moretti's social network graph for Hamlet

The most important similarity noticed between the two social networks is the central characters. Both social networks show Hamlet as the main central character within the play along with King Cornelius, Queen Gertrude, and Horatio being only slightly behind in terms of node centrality. Also, both social networks identify the same lesser important characters within the play as indicated by their outside presence in the social network and their fewer number of edges. The majority of edges between major and minor characters appear to be also similar in both social networks. Because both social networks identified the major characters within the play, the social network generated in this paper is more valid in terms of creating an accurate social network depiction of *Hamlet*.

Although both social networks were similar with regards to important features, there were some differences between the two social networks. Some character nodes in the social network created in this paper are not present in the social network created by Moretti including “All”, Servant, and “Prologue”. Within the XML file for *Hamlet*, prologue is identified as a speaker and therefore was categorized as a character node. One of the major differences between the two social networks is that the social network graph created by Moretti is undirected while the generated social network graph in this paper is directed. Although the edges are different in type, it does not affect the connections between characters. Also, edges in Moretti’s social network do not visualize the weight of the edges between characters while the social network in this paper does visualize the weight of edges. Other differences include the fact that the characters Gravedigger and 2nd Gravedigger in Moretti’s social network are the same characters First Clown and Second Clown in the social network for this paper.

5 Conclusion

The parsing program is able to determine the speaking characters in the play *Hamlet*, and Gephi is able to generate a social network for *Hamlet* that can identify major and minor characters. However, the social network contains some inaccuracies such as a character node for “Prologue” and potential misconnections between characters. Some of these inaccuracies within the social network could be reworked through more accurate identifications of characters speaking to one another. This paper serves to show that future social networks of plays can be generated with accurate representations of character connections. In addition, we have created a framework for parsing and analyzing social networks for future plays.

References

[1] ibiblio. https://www.ibiblio.org/xml/examples/shakespeare/hamlet.xml

[2] jdom. http://www.jdom.org/

[3] Gephi. https://gephi.org/

[4] Agarwal, A., Corvalan, A., Jensen, J., and Rambow, O. *Social Network Analysis of Alice in Wonderland.* Department of Computer Science, Columbia University, New York, NY, USA. 2012. http://www.aclweb.org/anthology/W12-2513

[5] Borgatti, S., Mehra, A., Brass, D., and Labianca, G. *Network Analysis in the Social Sciences.* 2009. https://www.researchgate.net/profile/Stephen_Borgatti/publication/24004962_Network_Analysis_in_the_Social_Sciences/links/09e4150b425f5a929a000000.pdf

[6] Butts, C. *Social Network Analysis: A methodological introduction.* Department of Sociology and Institute for Mathematical Behavioral Sciences, University of California, Irvine, California, USA. 2008. http://courses.washington.edu/ir2010/readings/butts.pdf

[7] Campbell, W., Dagli, C., and Weinstein, C. *Social Network Analysis with Content and Graphs.* Lincoln Laboratory Journal, Volume 20, Number 1. 2013. https://www.ll.mit.edu/publications/journal/pdf/vol20_no1/20_1_5_Campbell.pdf

[8] Elson, D., Dames, N., and McKeown, K. *Extracting Social Networks from Literary Fiction.* Columbia University. 2010. http://www1.cs.columbia.edu/~delson/pubs/ACL2010-ElsonDamesMcKeown.pdf

[9] Freeman, L. *Centrality in Social Networks Conceptual Clarification.* Lehigh University. 1979. http://leonidzhukov.ru/hse/2013/socialnetworks/papers/freeman79-centrality.pdf

[10] Grunspan, D., Wiggins, B., and Goodreau, S. *Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research.* 2014. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041496/

[11] Jackson, M. *The Study of Social Networks in Economics.* Department of Economics, Stanford University, Stanford CA. 2007. http://web.stanford.edu/~jacksonm/netsocialecon.pdf

[12] Kydros, D. and Anastasiadis, A. *Social network analysis in literature. The case of The Great Eastern by A. Embirikos.* TEI of Central Macedonia, Greece. http://www.eens.org/EENS_congresses/2014/kydros_dimitrios_and_anastasiadis_anastasios.pdf

[13] Moretti, F. *Network Theory, Plot Analysis.* 2011. https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf