|  |
| --- |
| Theatrical Genre Prediction Using Social Network Metrics |
| Manisha Shukla1, Susan Gauch1 and Lawrence Evalyn2 |
| 1Department of Computer Science and Engineering, University of Arkansas, Fayetteville, AR, USA  *2Department of English, University of Toronto, Toronto, ON, Canada*  mshukla@email.uark.edu, sgauch@uark.edu, lawrence.evalyn@mail.utoronto.ca |

Keywords: Social Networks, Genre Prediction, Relationship Mining, Social Network Analysis, Network Theory.

Abstract: With the emergence of digitization, large text corpora are now available online which provide humanities scholars an opportunity to perform literary analysis leveraging the use of computational techniques. In this paper, we apply network theory concepts in the field of literature to explore correlations between the mathematical properties of the social networks of plays and the plays’ dramatic genre. We generated character interaction networks of 37 Shakespeare plays and tried to differentiate plays based on social network features captured by the character network of each play. We were able to successfully predict the genre of Shakespeare’s plays with the help of social network metrics and hence establish that differences of dramatic genre are successfully captured by the local and global social network metrics of the plays. Since the technique is highly extensible, future work can be extended for fast and detailed literary analysis of larger groups of plays, including plays written in different languages.

# 1 INTRODUCTION

In literary studies, the three key areas of research could be defined as philology (the study of words), bibliography (the study of books as objects), and criticism (the evaluation or interpretation of literary meaning). Particularly since the advent of New Criticism, “the basic task of literary scholarship has been close reading of texts” (F. Moretti, 2011), which builds textual interpretations from precise study of specific words. Computational approaches to literature offer an alternate methodology for the work of literary study without close reading. “Distant reading” (F. Moretti, 2011) takes many forms, including statistical topic models (M. L. Jockers and D. Mimno, 2013), character profiling (Flekova and I. Gurevych, 2015), character frequency analysis (G. Sack, 2011), and sentiment analysis (M. Elsner, 2015), as mentioned in Grayson et al., 2017. For computational methods to produce new literary insights, they must provide information about literary texts which is not easily accessible by reading them and must do so for more texts than it is feasible for a person to read. Our paper presents a distant reading method which may aid in the task of literary criticism using network graph analysis on social networks generated from the scripts of plays.

Social network analysis is well-established as a way to study social groups. Some scholars have applied social network analysis to literary works for e.g. plot analysis (Grayson et al., 2017), or for discovering character communities (D. Watts, 2001), wherein nodes represent characters, and edges represent interaction between pairs of characters for plot analysis. Because these graphs are handmade for a very small number of plays, however, almost no work has been done to study the ability of mathematical properties of network graphs to predict literary features. We address this gap by exploring correlations between the mathematical properties of networks and dramatic genre.

In this paper, we study the social networks of Shakepeare’s plays to establish a correlation between social network metrics and genre identification. Using character networks of Shakespeare’s plays we found that combinations of some of the global and local network metrics (D. Watts, 2001) were able to successfully distinguish plays belonging to different genres. This work has been used for literary analysis of the ambiguous genre of Shakepeare’s “problem plays” (L. Evalyn, S. Gauch, and M. Shukla, 2018).

# 2 Related Work

## 2.1 Social Network Analysis

A social network graph is a set of vertices and edges (called a sociogram) where vertices represent social actors and edges represent social relations among the vertices. However, a social network is more than just a set of vertices and lines, as its structure contains implicit information about the social actors and their relationships. The graph representation of a social network offers a systematic and mathematical method for investigating these structures. Social network analysis is the process of investigating social network structures and ties through the use of network and graph theory concepts.

As S. M. Billah and S. Gauch observe, “Social network analysis (SNA) is not a formal theory, but rather a wide strategy for investigating social structures” (S. M. Billah and S. Gauch, 2015). These strategies borrow core concepts from sociometry, group dynamics, and graph theory (D. Watts, 2001) (J. Scott, 2000) (S. Wasserman, and K. Faust, 1994).

In social network analysis of human activities, the nodes can be connected by many kinds of ties, such as “shared values, visions, and ideas; social contacts; kinship; conflict; financial exchanges; trade; joint membership in organizations; and group participation in events, among numerous other aspects of human relationships” (Serrat, 2017). Regardless of the nature of the connection, “the defining feature of social network analysis is its focus on the structure of relationships” (Serrat, 2017). The central assumption in SNA methodologies is that relationships between nodes are of central importance (Serrat, 2017).

Social network analysis has been used in a wide variety of applications, with applications as diverse as disintegration models based on social network analysis of terrorist organizations (D. Anggraini et al., 2015), collaboration of scholars in graduate education (Wang Chuan-yi, Lv Xiao-hong, and Cao Yi, 2016), football team performance based on social network analysis of relationships between football players (Raffaele Trequattrini, Rosa Lombardi, Mirella Battista, 2015), money laundering detection (Rafał Dreżewski, Jan Sepielak, and Wojciech Filipkowski, 2015), and stress disorder symptoms and correlations in U.S. military veterans (Cherie et al. 2017).

## 2.2 Literary Analysis with SNA

Because dramatic performances enact social encounters, social network analysis translates surprisingly well to fictional societies. Stiller et al. have shown that social networks in Shakespeare’s plays mirror those of real human interactions, particularly in size, clustering, and maximum degrees of separation (Stiller J., Nettle D., and Dunbar, R. I. M., 2003).

Surveying the field of literary analysis using SNA, Moretti categorizes several types of analyses: “an empirical, quantitative and hierarchical description of literary characters (Jannidis, F. et al., 2016), corpus-based analyses exploring options for historical periodisation of literature (Trilcke, P. et al., 2015) and types of aesthetic modelling of social formations in and by literary texts (Stiller, J., Nettle, D., Dunbar, R. I. M., 2003) (Stiller, J., Hudson, M., 2005) (Trilcke, P. et al, D., 2016).” Moretti himself uses social networks to examine the plots of three Shakespearean tragedies, and to contrast a few chapters in English and Chinese novels (F. Moretti, 2011). Work following Moretti has focused on historical periodization, as in Algee-Hewitt’s examination of 3,439 plays looking only at the Gini Coefficient of each play’s eigenvector centrality to track changes in ensemble casts from 1500 to 1920 (Algee-Hewitt, M., 2017).

Our project focuses on a novel application, the classification of literary genre. When scaled up to a corpus covering a wider historical time span, our approach to genre could also provide insight on the historic periodization of literature.

Moretti also identifies that, in the application of SNA to literature, “methods for the automated extraction of network data (named entity recognition, co-reference resolution) and their evaluation are of particular importance,” (F. Moretti, 2011), which we accomplish in this paper.

## 2.3 Gephi Toolkit

Gephi is an open source software for graph and network analysis, which allows for fast visualization and manipulation of large networks. As a generalist tool, “it provides easy and broad access to network data and allows for spatializing, filtering, navigating, manipulating and clustering” (Bastian, M.; Heymann, S.; Jacomy, M., 2009). Gephi also calculates a wide range of mathematical features for each graph, which we use as the basis for our mathematical analysis (as discussed in more detail in 3.3).

# 3 Our Design

Our system for identifying genre consists of three building blocks: the Play Parser, the Social Network Generator and the Genre Predictor. Figure 1 shows the main components of the system architecture, which are discussed in more detail in the following subsections.

Play Parser

Social Network Generator

Genre Predictor

Figure 1: Block diagram of our system.

## 3.1 Play Parser

The main purpose of this component is to parse each play to extract basic information such as the total number of characters, the name and role of each character, and the total number of acts and scenes in a play. For each scene, we used our parsed information to determine which characters were present in the scene (using stage directions to account for entrances and exits during a scene), and how many lines and words were spoken by each character. Table 1 shows some details.

Table 1: TEI encoded XML file information.

|  |  |  |
| --- | --- | --- |
| XML Tag | Information contained | Example |
| <casteItem> | Character name and role in the play | <castItem type="role">  <role xml:id="Pedro"> Don Pedro</role> <roleDesc> prince of Arragon </roleDesc>  </castItem> |
| <div> | Act | <div xml:id="sha-man1">  <head>Act 1</head> |
| <div> | Scene | <div xml:id="sha-man101">  <head>Act 1, Scene 1</head> |
| <speaker> | Current speaker | <speaker>Beatrice </speaker> |
| <l>, <ab> | Line | <l xml:id="sha-man101299" n="299">  And tell fair Hero I am Claudio,</l>  <ab xml:id="sha-man201311" n="311"> born to speak all mirth and no matter.</ab> |

## 3.2 Social Network Generator

This component creates each play’s social network graph using the information generated by the play parser described in 3.1. In order to generate files which can be used to create graphs for plays, we used Gephi API to generate gexf files. Each gexf file maps character to a node and communication between characters as an edge. Each character stores as an attribute total number of lines and words spoken by that character in the play. After this mapping, each edge is weighted with the sum of total number of lines and words spoken by the two characters in their shared scenes. These form the basis of the extracted features as shown in Table 2. Once the basic structure is ready, using inbuilt functions of Gephi API we calculate 17 metrics of graphs features. These are represented as network features in Table 2.

Table 2: Features extracted from Shakespeare’s plays. Here n represents a play in graph, c character node in a graph, and e an edge in graph.

|  |
| --- |
| Extracted Features |
| *Features* |
| 1. tot\_characters = total number of characters of n |
| 2. tot\_edges *=* total number of edges of n |
| 3. tot\_lines = total number of lines spoken by c in n |
| 4. tot\_words = total number of words spoken by c in n |
| Network Features |
| 5. Degree = set of adjacent nodes of c in the graph |
| 6. Criticality = A k-critical graph is a critical graph with chromatic number k; a graph G with chromatic number k is k-vertex-critical if each of its vertices is a critical element. |
| 7. EigenVector = A measure of c’s importance in a network based on c’s connections. |
| 8. Eccentricity = The eccentricity of a graph vertex in a connected graph is the maximum graph distance between and any other vertex of. |
| 9. Closeness Centrality = The average distance from a given node to all other nodes in the network. |
| 10. Harmonic Centrality = In a (not necessarily connected) graph, the harmonic centrality reverses the sum and reciprocal operations in the definition of closeness centrality. |
| 11. Betweeness Centrality = Node Betweenness Centrality measures how often a node appears on shortest paths between nodes in the network. |
| 12. Clustering Coefficient = The clustering coefficient, when applied to a single node, is a measure of how complete the neighborhood of a node is. When applied to an entire network, it is the average clustering coefficient over all nodes in the network. |
| 13. Density = Measures how close the network is to complete. A complete graph has all possible edges and density equal to 1. |
| 14. Diameter = The maximal distance between all pairs of nodes. |
| 15. Path Length = The average graph-distance between all pairs of nodes. |
| 16. Connected Components = A connected component of an undirected graph is a maximal set of nodes such that each pair of nodes is connected by a path. |
| 17. Modularity = Measures how well a network decomposes into modular communities. |

### 3.2.1 Extracted Features

As extracted features, we chose to use most simple and easily quantifiable metrics, such as the total number of characters in the play (see Table 2). As our results in 4.3.1 and 4.3.3 demonstrate, despite their simplicity as features, the number of edges and the number of words spoken in a play can play a crucial role in identifying the genre.

### 3.2.2 Network Features

We compute network features of each play from the graph described in 3.2. For features that describe an individual node, such as degree or eigenvector, we calculate the average value of that feature for all the nodes in the graph.

## 3.3 Genre Predictor

The genre predictor is a support vector machine binary classifier which uses one vs one technique for classification. Support Vector Machines (SVMs) are a popular machine learning method for classification, regression, and other learning tasks. Since our classification problem had more than two classes, we combined SVM with One vs One (OvO) classification. This works as follows: choose a pair of classes from a set of *n* classes and develop a binary classifier for each pair. Create all possible combinations of pairs of classes from *n* and then for each pair develop a binary SVM.

For combinations we have:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Since SVM is a binary classifier, *k* = 2.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

In our case, we have three different classes (tragedy, history and comedy); therefore *n* = 3.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Therefore, we need three SVM classifiers.

The final class is assigned to each unseen play based on the class chosen by maximum number of binary SVM classifiers.

By using OvO, our SVM is much less sensitive to the problems of imbalanced datasets, which is particularly helpful given the different sizes of each of our three classes and our small overall sample size (Chang and Lin, 2011).

The model is trained using the extracted and network features of 31 plays in training dataset, listed in Table 3. It is then used to predict the genres of the 6 plays in testing dataset, as listed in Table 4.

# 4 EXPERIMENTS

## 4.1 Dataset

Our dataset is comprised of 37 plays by Shakespeare, in TEI encoded XML files. XML format was chosen as it was much easier to fetch required information from the plays along with maintaining accuracy in the extraction. The dataset was downloaded from the website exist-db.org. Out of 37 Shakespeare plays, 31 were used as a training set as detailed in Table 3 and remaining 6 as a test set as shown in Table 4. The test set comprised of two plays from each genre. For 5-fold cross-validation all 37 plays were used.

Table 3: Training dataset.

|  |  |
| --- | --- |
| **Play** | **Genre** |
| All's Well That Ends Well | Comedy |
| As You Like It | Comedy |
| Cymbeline | Comedy |
| A Midsummer Night's Dream | Comedy |
| Love's Labour's Lost | Comedy |
| The Merchant of Venice | Comedy |
| Measure for Measure | Comedy |
| The Merry Wives of Windsor | Comedy |
| The Taming of the Shrew | Comedy |
| Pericles, Prince of Tyre | Comedy |
| Troilus and Cressida | Comedy |
| Two Gentlemen of Verona | Comedy |
| Twelfth Night or What You Will | Comedy |
| The Winter's Tale | Comedy |
| The First Part of King Henry the Fourth | History |
| The Life of King Henry the Fifth | History |
| The Life of King Henry the Eight | History |
| The First Part of King Henry the Sixth | History |
| The Second Part of King Henry the Fourth | History |
| The Second Part of King Henry the Sixth | History |
| The Third Part of King Henry the Sixth | History |
| The Tragedy of King Richard the Second | History |
| Julius Caesar | Tragedy |
| King Lear | Tragedy |
| Hamlet,Prince of Denmark | Tragedy |
| Macbeth | Tragedy |
| Othello | Tragedy |
| Romeo and Juliet | Tragedy |
| Titus Andronicus | Tragedy |
| Timon of Athens | Tragedy |
| Antony and Cleopatra | Tragedy |
| Coriolanus | Tragedy |

Table 4: Test dataset.

|  |  |
| --- | --- |
| **Play** | **Genre** |
| The Life and Death of King John | History |
| The Tragedy of King Richard the Second | History |
| Timon of Athen | Tragedy |
| Macbeth | Tragedy |
| The Comedy of Errors | Comedy |
| Much Ado About Nothing | Comedy |

## 4.2 Experimental Setup

Our generated network graphs are then used to test our central question: whether the social network of characters in a play can be used as a proxy for features of the play’s narrative content. Can we use social network metrics to distinguish between the dramatic genres of tragedy, comedy, and history? We used 17 different mathematical features as mentioned in Table 2 to test our hypothesis. We first tested how well individual features were able to distinguish between different genres. Our second test comprised of all combinations of pairs of extracted and network features, and the third test used combinations of three features to evaluate the performance. Section 4.3 discusses the result of each test.

## 4.3 Results

The following table shows the calculated average value for each network metric per genre.

Table 5: Average feature value for each genre.

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Tragedy** | **History** | **Comedy** |
| Characters | 59.4 | 62.9 | 43.35 |
| Edges | 870.4 | 924 | 585 |
| Words | 31777.4 | 36544.1 | 33884.88 |
| Lines | 3774.7 | 4102 | 3936.94 |
| Criticality | 0.27 | 0.68 | 0.23 |
| Eigenvector | 0.39 | 0.39 | 0.52 |
| Eccentricity | 2.32 | 3.03 | 2.15 |
| Closeness | 0.56 | 0.52 | 0.62 |
| Harmonic | 0.62 | 0.57 | 0.68 |
| Betweenness | 12.53 | 19.24 | 7.45 |
| Clustering Coefficient | 0.73 | 0.74 | 0.81 |
| Graph Density | 0.48 | 0.45 | 0.61 |
| Diameter | 3 | 3 | 3.3 |
| Path Length | 1.51 | 1.60 | 1.39 |
| Connected Components | 2 | 1 | 1.23 |
| Degree | 28.13 | 28.45 | 26.27 |
| Modularity | 0.14 | 0.20 | 0.12 |

### 4.3.1 Single Feature Accuracy

Our first test attempted to identify genre using only single feature at a time. However, no single feature was independently sufficient to identify the genre. Of the features tested, graph density provided the greatest accuracy (83%) for genre identification.

Table 6: Genre prediction accuracy using a single feature.

|  |  |
| --- | --- |
| **Feature** | **Accuracy** |
| Graph Density | 83.33 |
| Total No. of Characters | 66.67 |
| Total No. of Lines | 66.67 |
| Path Length | 66.67 |
| Connected Components | 66.67 |
| Average Degree | 66.67 |
| Total No. of Edges | 50 |
| Total No. of Words | 50 |
| Average Criticality | 50 |
| Average Eccentricity | 50 |
| Average Closeness Centrality | 50 |
| Average Clustering Coefficient | 50 |
| Modularity | 50 |
| Average Eigenvector Centrality | 33.33 |
| Average Harmonic Centrality | 33.33 |
| Average Betweenness Centrality | 33.33 |
| Diameter | 33.33 |

### 4.3.2 Pair of Features Accuracy

However, when features were used in pairs, the network graphs achieved full accuracy in identifying the genre of Shakespeare plays. Table 7 shows pair of metrics which were able to identify genre with 100% accuracy on our test set.

Table 7: Pairs of features which provided 100% accuracy in genre prediction.

|  |  |
| --- | --- |
| Feature 1 | Feature 2 |
| Edges | Connected Components |
| Lines | Eccentricity |
| Eccentricity | Connected Components |
| Density | Connected Components |
| Modularity | Connected Components |

### 4.3.3 Three Features Accuracy

If we combine three features, the network graphs again achieve full accuracy in genre identification. Table 8 shows the triads which were able to identify genre with 100% accuracy on our test set.

Table 8: Sets of three features which provided 100% accuracy in genre prediction.

|  |
| --- |
| **Feature 1, Feature 2, Feature 3** |
| Characters, Edges, Connected Components |
| Characters, Lines, Eigenvector |
| Characters, Eccentricity, Path Length |
| Characters, Density, Path Length |
| Characters, Connected Components, Degree |

### 4.3.4 Discussion

The relevance of graph density in distinguishing genres is visually obvious when individual comedy and history networks are compared.

Histories feature highly dispersed networks, with large numbers of very minor characters, such as “First,” “Second,” and “Third” members of groups like soldiers and ambassadors (Figure 2).

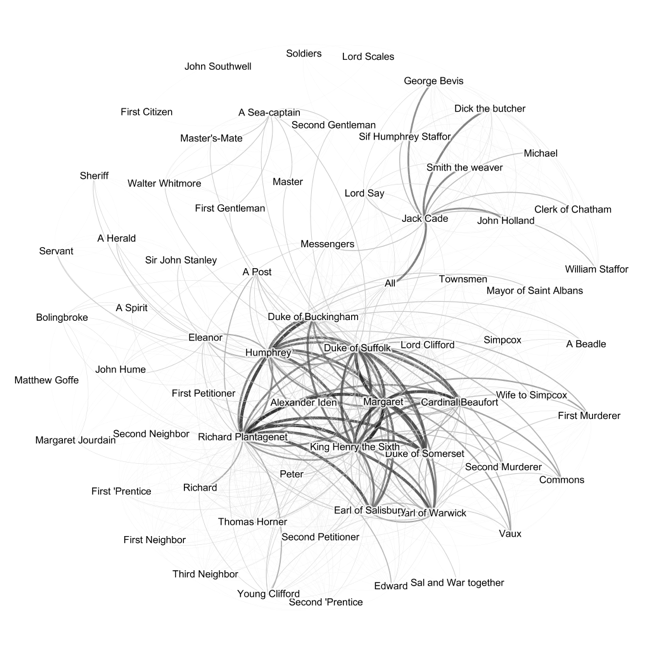


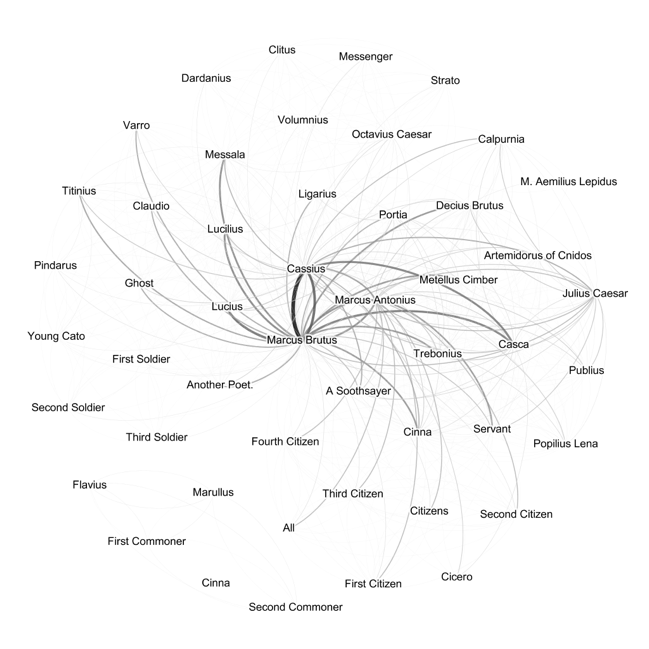
Figure 2: Network graph of *The Second Part of King Henry The Sixth*, a history.

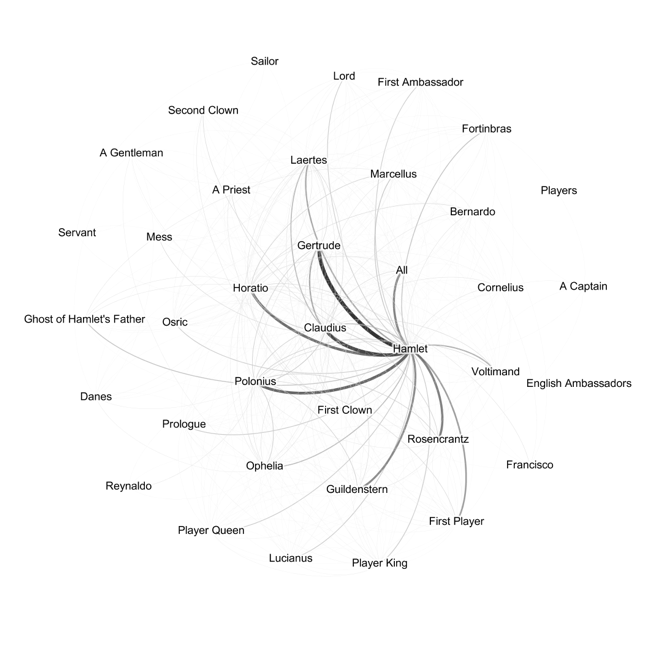
Comedies, in contrast, feature networks with far fewer characters, in which nearly everybody speaks to nearly everybody else at some point (Figure 3).



Figure 3: Network graph of *The Comedy of Errors*, a comedy.

Graph density is insufficient, however, to fully distinguish the tragedies, which feature networks somewhere between history and comedy in their density and show more variety overall (Figures 4 and 5). Therefore, more complex metrics are needed in combination with each other to accurately identify all three genres.

 Figure 4: Network graph of *Julius Caesar,* a tragedy.

Figure 5: Network graph of *Hamlet*, a tragedy.

A comparison of Table 7 and Table 8 shows that the sets of three factors which provide 100% accuracy do not necessarily always include the factors which were able to provide 100% accuracy as pairs. Many of the pairs, for example, include connected components as one of the two identifying features, but only two of the triples include it as a feature, and the triples instead include the number of characters as the most commonly useful feature.

Each metric thus seems to capture a specific kind of information about the play which are more relevant in combination with different other metrics. The number of connected components, for example, is only able to provide 67% accuracy alone, but reaches 100% when paired with edges, eccentricity, density, or modularity. Similarly, the number of characters in the play only provides 67% accuracy alone, but when considered alongside pairs of other features, the combination is more informative.

# 5 CONCLUSIONS

In this paper, we successfully classify plays based on their genre without using the actual words of the plays. Our networks of the well-studied works of Shakespeare can provide a baseline against which to contextualize similar studies of other plays. The network graphs themselves provide a new insight into the plays, revealing the hidden shape of social relationships between characters. The application of mathematical graph analysis to these networks provides a dramatically faster and more scalable way to determine important information about them, in this case their genre.

To apply these findings to literary research, we have explored in more detail the genre attributions of Shakespeare’s romances and problem plays (L. Evalyn, S. Gauch, and M. Shukla, 2018). We have also made the network graphs and selected mathematical features available online at <http://text.csce.uark.edu/clip/sna/> .

# 6 Future Work

Since the parser is highly extensible and can be used with any plays encoded in TEI, future work applying these methods to literary analysis does not need to be restricted to plays that are similar to Shakespeare’s but could be used to compare plays over a long period of time. Future work doesn’t even need to be restricted to plays written in English; one future application in development, for example, will study eighteenth century plays written in English, French, and German. As we develop our website, we will add functionality for others to upload their own TEI encoded plays and download the resulting Gephi file, enabling broad applicability of our methods to new literary research problems.

Future refinements to the social network generator could make edges between nodes directional, to better capture imbalanced relationships between characters; this level of detail was not necessary to distinguish between Shakespeare’s plays, but might be important for different identification tasks. Natural Language Processing (NLP) could also be integrated into the parser to more accurately identify the targets of speech, to capture instances where characters are on stage but cannot hear what is being said or are not being spoken to. These kinds of improvements would reduce “false positives” in the creation of edges between nodes, perhaps enabling better analysis of larger or more complicated groups of literary plays.

# 7 REFERENCES

1. F. Moretti. Network Theory, Plot Analysis. New Left Review, 68:80–102, 2011.
2. <http://showcases.exist-db.org/exist/apps/Showcases/index.html>.
3. M. L. Jockers and D. Mimno, 2013. Significant themes in 19th-century literature. In Poetics, 41(6):750–769.
4. Flekova and I. Gurevych, 2015. Personality Profiling of Fictional Characters using Sense-Level Links between Lexical Resources. In Proc. Conference on Empirical Methods in Natural Language Processing, pages 1805–1816.
5. G. Sack, 2011. Simulating plot: Towards a generative model of narrative structure. In 2011 AAAI Fall Symposium Series.
6. M. Elsner, 2015. Abstract Representations of Plot Struture. LiLT (Linguistic Issues in Language Technology), 12(5).
7. Lawrence Evalyn, Susan Gauch, and Manisha Shukla. Analyzing Social Networks of XML Plays: Exploring Shakespeare’s Genres. In Digital Humanities Conference 2019.
8. Siobhán Grayson, Karen Wade, Gerardine Meaney, Jennie Rothwell, Maria Mulvany, and Derek Greene, 2016. Discovering structure in social networks of 19th century fiction. In Proceedings of the 8th ACM Conference on Web Science (WebSci '16). ACM, New York, NY, USA, 325-326.
9. D. Watts, “Small Worlds: The Dynamics of Networks between Order and Randomness”, Princeton University Press, 2001.
10. Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. ACM Trans. In Intell. Syst. Technol. 2, 3, Article 27 (May 2011), 27 pages.
11. S. M. Billah and S. Gauch. 2015. Social network analysis for predicting emerging researchers. In 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Lisbon, 2015
12. J. Scott, “Social Network Analysis: A Handbook”, 2nd ed., Sage Publications, London, 2000.
13. S. Wasserman, and K. Faust, “Social Network Analysis: Methods and Applications”, Cambridge University Press, 1994.
14. Serrat, O., 2017. Social network analysis. In Knowledge solutions (pp. 39-43). Springer, Singapore.
15. D. Anggraini, S. Madenda, E. P. Wibowo and L. Boumedjout, 2015. Network Disintegration in Criminal Network. In 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, 2015, pp. 192-199.
16. Wang Chuan-yi, Lv Xiao-hong, and Cao Yi. 2016. An empirical study on the collaboration of scholars in graduate education: based on the social network analysis. In Proceedings of the 2016 International Conference on Intelligent Information Processing (ICIIP '16). ACM, New York, NY, USA, Article 36, 7 pages.
17. Raffaele Trequattrini, Rosa Lombardi, Mirella Battista, 2015. Network analysis and football team performance: a first application. In Team Performance Management: An International Journal, Vol. 21 Issue: 1/2, pp.85-110.
18. Rafał Dreżewski, Jan Sepielak, Wojciech Filipkowski, 2015. The application of social network analysis algorithms in a system supporting money laundering detection. In Information Sciences, Volume 295, 2015, Pages 18-32, ISSN 0020-0255.
19. Cherie Armour, Eiko I. Fried, Marie K. Deserno, Jack Tsai, Robert H. Pietrzak, 2017. A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in U.S. military veterans. In Journal of Anxiety Disorders, Volume 45, 2017, Pages 49-59.
20. Stiller, J., Nettle, D., and Dunbar, R. I. M., 2003. The Small World of Shakespeare's Plays. In Human Nature, 14(4): 397–408.
21. Jannidis, F., Reger, I., Krug, M., Weimer, L., Macharowsky, L., Puppe, F., 2016. Comparison of Methods for the Identification of Main Characters in German Novels. In Digital Humanities Conference Abstracts, Jagiellonian University & Pedagogical University, Kraków, pp. 578–82.
22. Trilcke, P., Fischer, F., Göbel, M., Kampkaspar, D., 2015. In 200 Years of Literary Network Data.
23. Stiller, J., Nettle, D., Dunbar, R. I. M., 2003. The Small World of Shakespeare's Plays. In Human Nature 14, pp. 397–408.
24. Stiller, J., Hudson, M., 2005. Weak Links and Scene Cliques Within the Small World of Shakespeare. In Journal of Cultural and Evolutionary Psychology 3, pp. 57–73.
25. Trilcke, P., Fischer, F., Göbel, M., Kampkaspar, D., Kittel, C., 2016. Theatre Plays as 'Small Worlds' Network Data on the History and Typology of German Drama, 1730–1930. In Digital Humanities 2016 Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków, pp. 385–87.

Algee-Hewitt, M., 2017. Distributed Character: Quantitative Models of the English Stage, 1500-1920. In Digital Humanities 2017: Book of Abstracts. Montreal: McGill University and Université de Montréal, pp. 119–21.

1. Elson, D. K., Dames, N., McKeown, K. R., 2010. Extracting Social Networks from Literary Fiction. In Proceedings of ACL 2010. Uppsala, pp. 138–47.
2. Park, G. M., Kim, S. H., Cho, H. G., 2013. Structural Analysis on Social Network Constructed from Characters in Literature Texts. In Journal of Computers 8.9, pp. 2442–47.
3. Agarwal, A., Corvalan, A., Jensen, J., Rambow, O., 2012. Social Network Analysis of Alice in Wonderland. In Proceedings of the Workshop on Computational Linguistics for Literature. Montréal, pp. 88–96.
4. Fischer, F., Göbel, M., Kampkaspar, D., Trilcke, P., 2015. Digital Network Analysis of Dramatic Texts. In Digital Humanities 2015 Conference Abstracts. University of Western Sydney.
5. Waumans, M. C., Nicodème, T., Bersini, H., 2015. In Topology Analysis of Social Networks Extracted from Literature. Plos One, 3 June 2015.
6. Bastian, M., Heymann, S., Jacomy, M, 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In International AAAI Conference on Web and Social Media, North America.
7. https://www.csie.ntu.edu.tw/~cjlin/libsvm/
8. https://gephi.org/developers/
9. https://weka.wikispaces.com
10. https://www.cs.waikato.ac.nz/~ml/index.html