Classifying Shakespeare’s Plays with Network Graphs: Dramatic Genre in Character Speech Interactions

Genre, as one of the eternally difficult questions of literary study, is a ubiquitous subject for computational “distant reading” of literature. In setting out many of the “perennial doubts [that] plague genre theory” (1), Daniel Chandler notes as one source of difficulty that “[p]articular features which are characteristic of a genre are not normally unique to it; it is their relative prominence, combination and functions which are distinctive” (2). Often, the relative prominence of these distinctive features is obscured by the richness of individual literary texts. By introducing a layer of abstraction, distant reading allows for a new angle of study. This new angle of study enables both macroanalysis (Jockers) and what Hoover terms “microanalysis,” which uses “style variation within a single work or author as a way of better understanding authorship and style” (2016).[[1]](#footnote-1)

Our work carries out a microanalysis of dramatic genre, taking the widely-studied works of Shakespeare as our case study, as the prototype for future macroanalytical study. Work on Shakespeare has a particularly long history of genre studies. Scholars have been interpreting and revising the taxonomy of “comedy,” “history,” and “tragedy” since the First Folio collected Shakespeare’s plays into those categories. We work within the tradition that examines First Folio designations. We employ distant reading to confirm some ways that genre has been conceived in these plays, but also to reconsider them, and to suggest new tools to think about the problem cases.

Our mode of abstraction is the network: we model each individual play as a web of speech passing between characters who share a stage. Ours is not the first project to apply social network analysis to literary study.[[2]](#footnote-2) Where we differ from much past work is in our precise focus on linguistic encounters in the space of the stage, rather than rougher metrics like appearance within the same scene. To draw the edges between nodes that make up our networks, we pay close attention to who is on stage, and who can (probably) hear them speak[[3]](#footnote-3) — and how often they speak, how much they speak at once, and whether people speak back to them. Our model of each play thus renders the speech act as the fundamental “unit” of a play’s script[[4]](#footnote-4), and our core hypothesis is thus that genre, in drama, will be expressed in the relations between characters.

Examining network graphs requires, first, a substantial body of network graphs. Many scholars have handmade theatrical network graphs, both as wholly artistic interpretations (Moretti) and as mathematically-expressed networks (Trilcke et al., Fischer et al., Algee-Hewitt). Like the Digital Literary Network Analysis (dlina) working group, we acquire our networks by parsing plays which have been encoded according to the Text Encoding Initiative (TEI) standards. Our parser takes in a TEI-encoded play, extracts characters for the nodes of a network, and connects nodes via a directional edge each time a character speaks while another character is on stage as a potential auditor. By using directional edges, by weighting character connections based on speech, and by sub-dividing the play into blocks smaller than the unit of the scene based on the configurations of characters present on stage, we create networks at a higher level of detail than prior work which creates edges simply for co-occurrence within a scene. We began by replicating Moretti’s results in order to verify our parser’s accuracy,[[5]](#footnote-5) then generated networks for a full corpus of Shakespeare’s works.[[6]](#footnote-6)

We then used our generated network graphs to test our central question: whether character speech interaction can be used to distinguish between the dramatic genres of tragedy, comedy, and history. We decided that the answer would be “yes” if we could train a classifier to replicate the First Folio’s classifications for Shakespeare’s 36 plays. We tested three classifier algorithms[[7]](#footnote-7) with fivefold cross-validation, using 21 different mathematical features as the metrics for comparison.[[8]](#footnote-8) We found that even one feature could achieve classification accuracy[[9]](#footnote-9) substantially above random chance. Since there are three possibilities, a random guess would be correct 33% of the time, but we achieved 66.43% accuracy using only the networks’ average path length[[10]](#footnote-10). Accuracy increases as more features are considered in combination. With two features, our maximum accuracy is 77.86%.[[11]](#footnote-11) With three features, we reach 83.57%.[[12]](#footnote-12) We evaluated combinations of features up to 7 maximum. Our highest accuracy was 88.93%[[13]](#footnote-13). Compared to the baseline of 33%, this accuracy is sufficient to establish the core premise of our hypothesis, that networks of character speech offer meaningful information about dramatic genre.[[14]](#footnote-14)

The ability of the classifier to replicate the First Folio’s designations with substantially greater than random chance supports our initial hypothesis that the genre of a play is expressed in the relations it stages between its characters. The underlying details that drive the classification can also provide a new lens into the specific genres and character staging of of Shakespeare’s works, particularly in the case of plays which are resistant to this mode of classification. Beginning with the clearest results and moving into the realm of more ambiguous interpretation, we will discuss the implications of our classifications for the comedies; the histories and tragedies; and finally, the plays whose First Folio designations have been challenged.

Figure 1: Network graphs of all First Folio comedies, excluding the two problem plays and three romances.

As Table 1 below suggests, comedies are the most straightforward networks to identify. Whereas histories and tragedies often have similar values across our metrics, comedies tend to be distinct. Their smaller casts are more interconnected, which can be visually observed in Figure 1 above, and can be measured in their networks’ higher density, lower eccentricity, and shorter path length.[[15]](#footnote-15) All of these measures indicate that most characters in a comedy will eventually speak to most other characters. Although comedies often have multiple subplots, we find that these separate stories do not result in highly-separated networks. (See, for example, 2, *A Midsummer Nights’ Dream*, in which distinct “worlds” of the play are nonetheless connected through the bridging character of Bottom.) We theorize that comedic networks are strongly shaped by their final “resolution” scenes, which typically bring the cast together for a restoration of social order, and therefore create edges knitting the network together. (See, again, *Midsummer*, and also Figure 3, *The Comedy of Errors*, which features two sets of identical twins who avoid each others’ presence until the final moments of the play.) Of course, the idea that comedy is about social connection is not new, but that it can be mapped by the way characters relate on the stage provides a novel lens through which to view it.

Figure 2: *A Midsummer Night’s Dream.*

Figure 3: *The Comedy of Errors.*

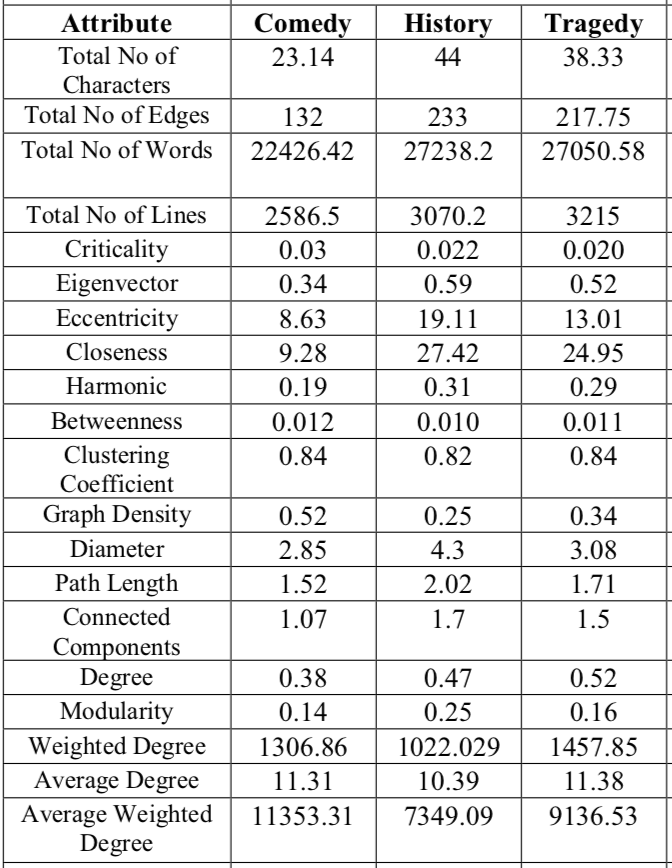


Table 1: The average values per genre for several of the network features we measured.

What Table 1 above also suggests is that histories and tragedies are fairly similar genres. We interpret these results by understanding histories as representing one end of a spectrum which has comedies on the opposite end, with tragedies falling between histories and comedies.

Figure 4: Network graphs of all First Folio histories.

Histories, as can be visually seen in in Figure 4, feature large and highly dispersed networks. Their peripheries are populated with large numbers of very minor characters,[[16]](#footnote-16) such as “First,” “Second,” and “Third” members of groups like soldiers and ambassadors, who each interject briefly in a single scene. (See, for example, Figure 5, *Richard III*, and for an extreme case Figure 6, *Henry VI Part 1*.) Accordingly, density is half that of the comedies. Connections form chains of acquaintance with little overlap, resulting in a high average path length and a high graph diameter. Whereas the directional edges in comedies tended to be reciprocal and balanced, in histories they are often unidirectional, as characters give orders or statements to each other.

Figure 5: *Richard III*

Figure 6: *Henry VI, Part 1*

Figure 7: Network graphs of all First Folio tragedies, excluding the romance *Troilus and Cressida* and the problem play *Cymbeline*.

Tragedies are more formally unpredictable, especially visually, as seen in Figure 7. Some have dense cores with small peripheries that wouldn’t be out of place among the comedies (Figure 8, *King Lear*; Figure 9, *Romeo & Juliet*). *Othello*, in particular, is almost claustrophobically small (Figure 10). Others, especially the Roman plays, look like histories with their enormous peripheral casts (Figure 11; Figure 12). On average, their mathematical features in Table 1 place them closer to, but distinct from, the histories.

Figure 8: *King Lear.*

Figure 9: *Romeo & Juliet.*

Figure 10: *Othello.*

Figure 11: *Antony and Cleopatra*.

Figure 12: *Coriolanus*.

What might unite these tragedies is the increased likelihood of a “tragic hero” who appears in the network graph as a distinct, singular centre of the network. These central nodes are typically connected to a large number of other nodes which may be only sparsely connected to each other, with a disproportionately high number of heavily weighted edges connecting outward (indicating that they have given long speeches) with little to no reciprocation. In other words: a Hamlet. “Hamlet” nodes can be seen in *Hamlet*, of course, and in *Macbeth*, *Timon of Athens*, and *King Lear*. In *Antony & Cleopatra*, it seems that Antony and Cleopatra are in competition to see who gets to be the “Hamlet” node; *Julius Caesar* and *Othello*, too, might show two characters competing for the role. But this pattern is not unique to tragedies: it appears in the histories *Richard III* and *Henry V* as well, and in the *Measure for Measure* and *The Tempest*, both designated as comedies in the First Folio. This seems to be a case, then, when Chandler’s “relative prominence” comes into play. A third of the First Folio’s tragedies show a distinct “Hamlet” node, and if we include the plays with “competing Hamlets” fully 58% of the First Folio tragedies show this pattern. In contrast, only 20% of the histories show it, and only 13% of the First Folio’s comedies, both of which now have their status as comedies disputed.

Fig 13: Network graphs of all the “problem plays” and romances.

As our discussion of the tragedies suggests, genre is never simple. Even in evaluating accuracy, we really determine the classifier’s ability to predict how the First Folio would have labelled a work, not the work’s “true” and objective genre. Shakespearean scholars by no means universally accept the First Folio’s assignments. Accordingly, once we had trained our SVM classifier, we queried three sets of ambiguous texts — what are now called the “Roman plays,” the “problem plays,” and the “romances” — to see what social network analysis might offer these debates. Looking closely at these plays’ networks illustrates the applicability of network-based close reading, as we will demonstrate by examining *Cymbeline* and *The Winters’ Tale* in more depth.

The “Roman plays” are four tragedies whose ancient Roman setting places them closer to tragedy’s origin in history. We thought our classifier might re-classify them as histories; however, as Table 2 shows, only *Antony & Cleopatra*, with only one classifier, was designated a history. Instead, this re-classification experiment suggest that closer attention ought to be paid to the small, dense network which makes up *Titus Andronicus*.

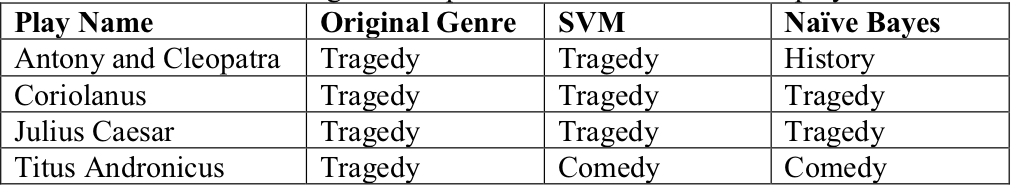


Table 2: classifications for the “Roman plays.”

The “problem plays” and the “romances” are both sets of plays whose contents have proved challenging to scholars, motivating re-assessment of their generic form. As Table 3 shows, both classifiers generally concur with the First Folio’s classification of the problem plays, with the exception that the SVM classifier identifies *Troilus and Cressida* as a comedy rather than a tragedy. The romances (Table 4) are slightly more ambiguous.

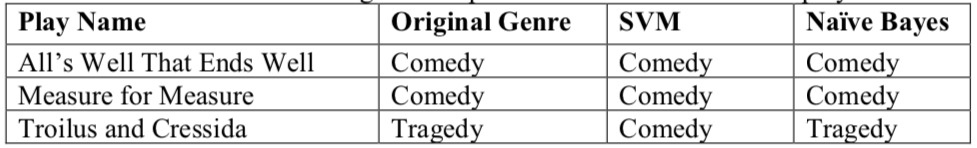


Table 3: classifications for the “problem plays.”

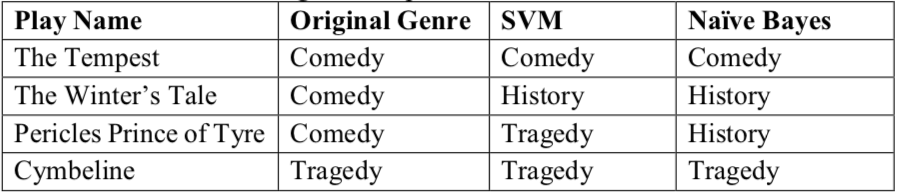


Table 4: classifications for the romances.

The classifiers concur with the First Folio regarding *The Tempest*, but neither classifies *The Winter’s Tale* or *Pericles[[17]](#footnote-17)* as a comedy. Moreover, the classification of *Cymbeline* (Figure 14) as a tragedy, even though this classification concurs with the First Folio, constitutes a surprise.

Audiences since Samuel Johnson in 1765[[18]](#footnote-18) have complained of *Cymbeline*’s failures as a tragedy (among other things, it has a happy ending), and Barbara Mowat refers to the Folio’s identification of the play as “The Tragedie of Cymbeline” in the index, title page, and running heads, as “misleading paratextual matter” (130). Our classification, in contrast, supports Valerie Wayne and Richard Proudfoot’s assessment that although “[i]t is not a tragedy as that form is most widely understood today or as Johnson understood it in the eighteenth century, […] it did sufficiently meet Renaissance definitions” of that genre (Wayne 403).

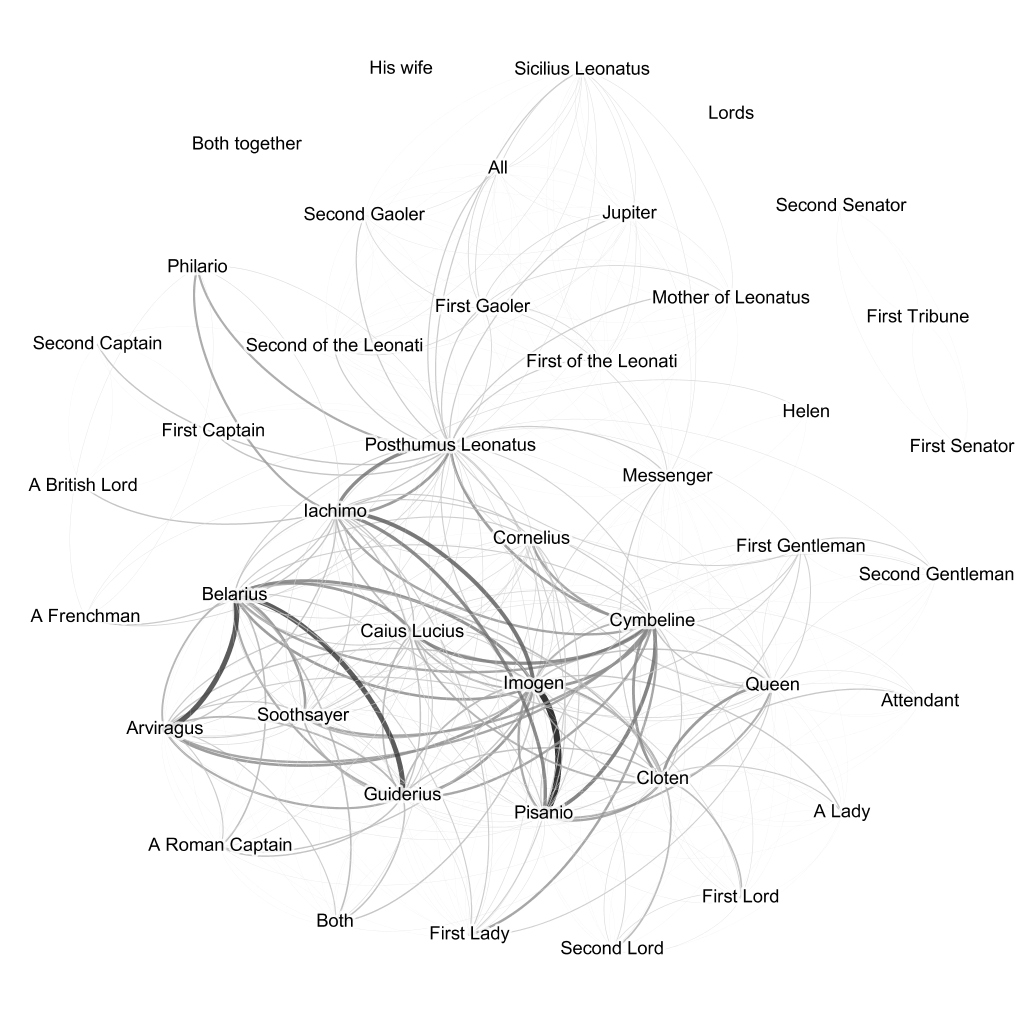


Figure 14: network graph of character interaction in *Cymbeline*.

More strange is the classifiers’ identification of *The Winter’s Tale* as a history.The play earns its designation as a romance due to its two-part structure: it begins much like a tragedy, with Leontes at its centre. Halfway through, the character of “Time” comes on stage to declare that sixteen years have passed, and the play changes to a pastoral comedy which brings the play to a happy resolution. These two theatrical dynamics are visible in the social network graph, as seen in Figure 15: Leontes radiates unreciprocated and unbalanced edges, looking like a Hamlet of his world. But the characters who enter in the latter half of the play (Perdita, Florizel, the Old Shepherd, the Clown, Autolycus, Dorcas, Mospa)[[19]](#footnote-19) connect to each other in a dense, comedic structure. The two substructures within the play are bridged by Leontes’ connections with Polixenes and Camillo[[20]](#footnote-20). As a result, the play’s network as a whole is neither a comedy nor a tragedy. The periphery of the network is filled both with the lords, ladies, and servants who attend Leontes’ court and with the various gentlemen, shepherds, and shepherdesses who join the pastoral festivities. These two sets of attendants, one from each half of the play, likely drive the classifiers’ identification of the overall network as a “history”.

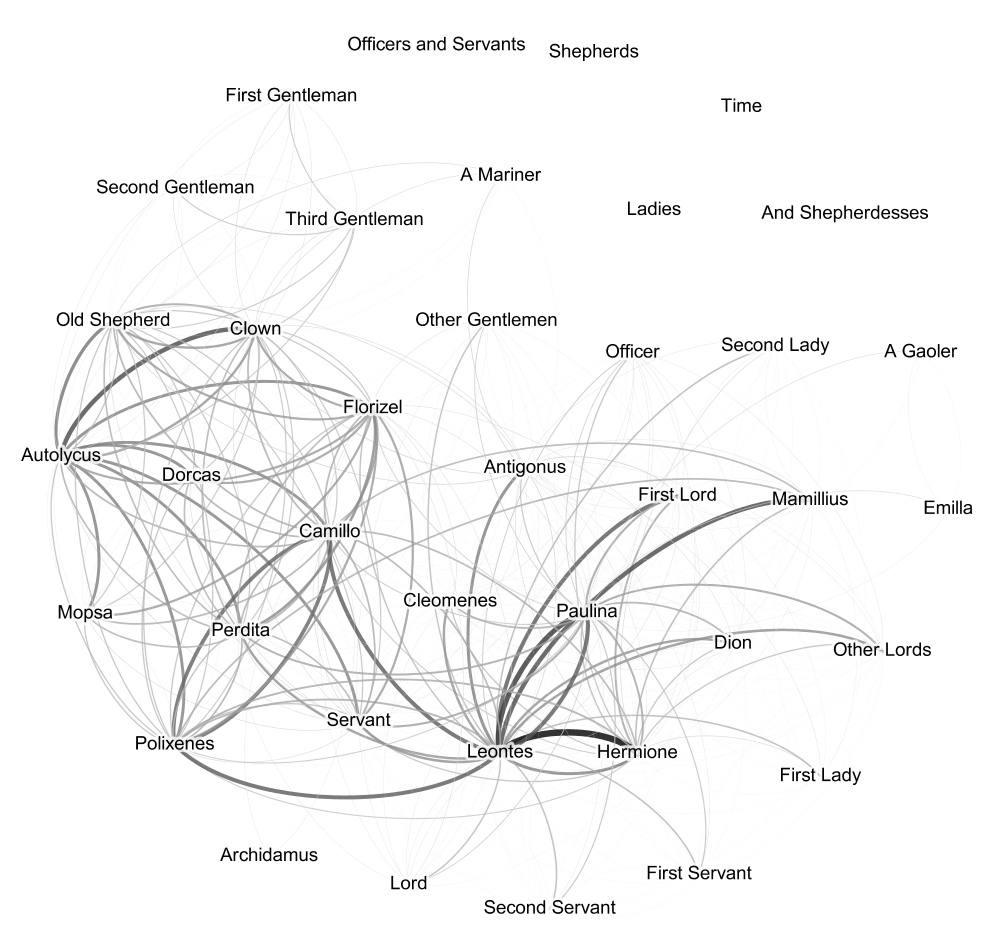


Figure 15: network graph of character interaction in *The Winter’s Tale*.

We treat none of our classifications as definitive declaration of the plays’ “real” genres, but network analysis gives us a new way to think about their genres. Thirty-six plays is a small scale for this approach: since the parser is highly extensible and can be used with any plays encoded in TEI, future work need not be restricted to the Early Modern period. It need not even be restricted to works written in English. The purpose of these tools is to confirm some ways that we have thought about these texts, but also to reconsider them and to open up new vocabulary for the problem cases. Genre, we argue, plays out in who people stand on a stage with, and who they talk to and listen to. Our networks of the well-studied works of Shakespeare can provide a baseline against which to contextualize genre analysis of works for which there is far less critical conversation.

1. Our work was initially motivated by a question we will not answer here, about generic change in eighteenth century tragedies; before we could examine the heroines in these plays, we realised, we needed to establish as a baseline that linguistic encounters in the space of the stage, in and of themselves, constitute meaning. [↑](#footnote-ref-1)
2. Most obviously, Moretti uses handmade network graphs to examine the plots of three Shakespearean tragedies, and to contrast the structure of chapters in English and Chinese novels (2011). 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 (2003). Alberich et al. (2002) and Sparavigna (2013) also discuss the interplay between social and narrative constraints on networks. [↑](#footnote-ref-2)
3. Identifying characters who are on stage but *cannot* hear a given speech remains elusive: the natural language processing necessary for the task will rely on period-specific language norms that cannot be generalized trans-historically like the rest of our methods. [↑](#footnote-ref-3)
4. Our emphasis on the speech act as, in itself, performative, is indebted to J.L. Austin’s How to Do Things with Words. [↑](#footnote-ref-4)
5. The existence of well-crafted graphs for comparison partly motivated our selection of Shakespeare as our first subject. In general, our graphs differ from Moretti’s primarily by including characters whom he excludes: we have nodes for minor characters like “Servant,” for example, and we include all the players of the play-within-the-play in *Hamlet*. Because our edges are directional, we are also able to represent details of character interaction which Moretti regretted leaving out, such as the fact that Horatio addresses the Ghost, but the Ghost speaks only to Hamlet. Despite these differences, our network graphs generally support Moretti’s literary readings. [↑](#footnote-ref-5)
6. Specifically, we generated networks for the 36 plays included in the First Folio, and (to contribute to scholarly conversations about Shakespeare’s “romances”) *Pericles, Prince of Tyre*, for 37 total plays. All play texts came from Northwestern University’s “WordHoard Shakespeare.” [↑](#footnote-ref-6)
7. We tested *k*-nearest neighbors (KNN), a support vector machine (SVM) with one vs one classification, and a Naïve Bayes classifier. KNN, we soon determined, was too vulnerable to the over-representation of comedies among Shakespeare’s works, so we eliminated it from further experimentation. The SVM and Naïve Bayes had highly similar accuracy, so remaining tests were carried out with both. The SVM’s maximum accuracy was typically a few percentage points higher than Naïve Bayes. [↑](#footnote-ref-7)
8. These features are a combination of extracted play features (such as the number of lines), node features averaged across all nodes (such as average eigenvector centrality), and graph features (such as graph density). [↑](#footnote-ref-8)
9. Accuracy here meaning, of course, “agreement with the First Folio.” [↑](#footnote-ref-9)
10. This figure was achieved by the SVM classifier. The maximum single-feature accuracy achieved by the Naïve Bayes classifier was 63.57%, examining only the number of lines in each play. [↑](#footnote-ref-10)
11. Evaluating plays based on a pair of features is the only case in which Naïve Bayes outperforms SVM. Naïve Bayes achieves 77.86% by considering lines & average eigenvector. SVM achieves 72.5% with four different pairs: harmonic & diameter; harmonic & path length; graph density & diameter; and graph density & path length. [↑](#footnote-ref-11)
12. SVM achieves 83.57% accuracy with words, characters, & lines. Naïve Bayes achieves 80.71% with lines, graph density, & degree. [↑](#footnote-ref-12)
13. This result was achieved with the SVM using five features. The features were a combination of play features (words, lines), node features (closeness), and network features (density and degree). [↑](#footnote-ref-13)
14. We anticipate that a larger testing corpus would substantially improve peak accuracy. As part of our experimentation, we tested excluding the various “problem plays” and romances from our corpus, in case these works were too ambiguous in genre to contribute meaningfully to the classification. However, instead of improving our accuracy, reducing our corpus to the most “standard” works substantially reduced accuracy, evidently because the sample was too small to generalize from. Thirty-six plays appears to be near the minimum corpus size for meaningful classification with our methods. [↑](#footnote-ref-14)
15. Algee-Hewitt has compared the Gini coefficient of eigenvector centrality and betweenness centrality for plays from the 16th to 20th centuries. We calculated only the averages for these values, which does not allow for a direct comparison to his findings, since similar averages may reflect very different underlying distributions of values, and a metric like the Gini coefficient is necessary to measure that distribution. However, the strong correlation of genre with graph density suggests that calculating these Gini coefficients would be worthwhile future work. Our findings suggest that, when Algee-Hewitt observes the disappearance of “periphery” roles, “as casts get smaller and actions take place among an increasingly more tightly knit set of characters” (), this change more specifically indicates a reduction or fundamental alteration in “history” plays. [↑](#footnote-ref-15)
16. Early Modern theatre companies remained generally constant in size, so these additional characters likely indicate increased doubling (or tripling, etc) of roles, i.e., having one actor play multiple roles. Future work could use network analysis to identify roles which can and cannot be doubled, and perhaps trace the decline of doubling as a theatrical practice after the Restoration. For more on Early Modern doubling of theatrical roles, see David Bevington’s *From Mankind to Marlowe: Growth of Structure in the Popular Drama of Tudor England*. My thanks to Arlynda Boyer for illuminating conversations about Shakespeare’s doubling. [↑](#footnote-ref-16)
17. *Pericles* is not, of course, a First Folio play, and has not formed part of our corpus for the other stages of this project, but was graphed and classified in service of this question. Its “original” designation as a comedy comes from general critical consensus, rather than the First Folio. [↑](#footnote-ref-17)
18. Every discussion of *Cymbeline*, and particularly of its genre, must duly quote Johnson’s judgment that “This play has many just sentiments, some natural dialogues, and some pleasing scenes, but they are obtained at the expense of much incongruity. To remark the folly of the fiction, the absurdity of the conduct, the confusion of the names and manners of different times, and the impossibility of the events in any system of life, were to waste criticism upon unresisting imbecility, upon faults too evident for detection, and too gross for aggravation” (307). We hereby discharge this solemn duty. [↑](#footnote-ref-18)
19. A brief description of these characters’ roles highlights the alteration in the play’s tone: two young lovers, three comic figures, and two shepherdesses. [↑](#footnote-ref-19)
20. His childhood friend, and a loyal servant who safeguarded his daughter Perdita, respectively. [↑](#footnote-ref-20)