DH Text Submission Guidelines

This is the template for abstract submissions to the annual Digital Humanities conference of ADHO. Please read and observe these guidelines closely in order to facilitate the file conversion for the edition of the Conference Proceedings. The guidelines included here are a subset of more complete instructions you can find online at: https://dh2018.adho.org/.

The template consists of sections with light grey background that cannot be edited and will be processed automatically during the TEI-conversion. The other sections are where you insert the content of your submission.

Below, you will find the three editable main sections of this document:

1. Subtitle: section for an optional subtitle
2. Body: section for the chapters of your paper
3. References: section for the bibliographic references.

Please avoid pasting richly-formatted text into these sections and only use the paragraph styles provided by this template named „DH-[Stylename]“:

* DH-Default
* DH-Subtitle
* DH-Heading1
* DH-Heading2
* DH-Heading3
* DH-Quotation

Also supported are links, lists and tables. Please avoid wide tables with lots of cell content. In addition, we support the following character level styles:

* Bold (to be used, sparingly, for emphasis)
* Italics (to be used for titles)

Please observe the following additional indications:

**Images:** Accepted image file formats are PNG, JPEG and SVG. For PNG and JPEG images the minimum image width is 1000 pixels.

**Quotations** of 5 or more lines should be in the "DH-Quotation" block quote format. For short quotations within the text, single quotation marks should be used.

**Program code or mathematical formulae** should appear as images.

**Notes** should be used only for comments, not for bibliographic citations.

**Bibliographic references** should be made inline using the author's name and year of publication, e.g. (Bloggs, 1990; Bloggs et al., 1991).

**References section:** The works cited should be listed in the "References" section at the end of the article. The list should be in alphabetical order by author. Where an author has more than one publication, they should be arranged in chronological order.

**Citation style:** Please follow the citation style of the DSH journal at:

http://dsh.oxfordjournals.org/for\_authors/index.html. We provide a CSL file for use with reference managers such as Zotero. See the file and instructions here:

https://github.com/computationalstylistics/DHAbstracts\_biblio\_style.

Some examples are given below.

**Biber, D.** (1988). *Variation Across Speech and Writing.* Cambridge: Cambridge University Press.

**Calzolari, N.** (1989). A typology of English text. In Batori, I. S., Lenders, W. and Putschke, W. (eds), *Computational Linguistics.* New York: ACM Press, pp. 510-19.

**Oostdijk, N.** (1988). A corpus linguistic approach to linguistic variation*. Literary and Linguistic Computing,* **3**: 12-25.

**Richardson, S. D. and Braden-Harder, L.** (1988). *The experience of developing a large-scale natural language text processing system: CRITIQUE. Proceedings of the Second Conference on Applied Natural Language Processing,* Austin, TX, February 1988.

**Garfinkel, M. S. and Weiss, S. C.** (1999). In the court of history, Ehlers v. Bush. *Recent Science Newsletter,* **1**(1): 6-7, http://web.archive.org/web/20030604160332/recentscience.gwu.edu/RSN/ (accessed 27 February 2004).

The main title of your contribution – this field cannot be edited

Analyzing Social Networks of XML Plays: Exploring Shakespeare’s Genres

Add a subtitle (optional):

Lawrence Evalyn (lawrenceevalyn@gmail.com), University of Toronto, Canada

Susan Gauch (segauch@gmail.com), University of Arkansas, United States

Manisha Shukla (mshukla@email.uark.edu), University of Arkansas, United States

Author information – this field cannot be edited

Main body – enter your text and images here:

Introduction

Our inquiry considers the speech interactions of characters within plays as a proxy for broad narrative structures. We analyze computationally-generated social networks of 37 plays by Shakespeare to see whether, and how, they can be used to distinguish between Shakespeare’s comedies, tragedies, and histories.

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 (2003). However, as fictions, these networks are shaped not only by sociological principles, but also by narrative structures. Moretti uses social networks to examine the plots of three Shakespearean tragedies, and to contrast the structure of chapters in English and Chinese novels (2011). Alberich et al. (2002) and Sparavigna (2013) also discuss the interplay between social and narrative constraints on networks. We emphasize this distinction to look for specifically literary features of our networks.

Recent papers presented at DH2017 sought ways to richly quantify the details of one or two plays (Fischer et al., 2017; Tonra et al., 2017). At another scale, Algee-Hewitt examined 3,439 plays by looking only at the Gini Coefficient of each play’s eigenvector centrality (2017). With our three dozen plays, we attempt to strike a fruitful middle ground in the inevitable balancing act between detail and scale. Each play is considered individually, but at a level of abstraction which allows rapid and direct comparisons.

Creation of social network graphs

Our parser tracks characters present on stage during speech. This approach is highly extensible: it can parse any play that follows TEI P5 guidelines for performance texts. Each speaking character is connected to all characters currently present on stage. These connections are recorded in a network graph, with characters as nodes and shared speech as edges. Edges are directional, and weighted based on the number of lines spoken.  In future, we plan to extend our parser to identify the specific addressees of a character’s speech, allowing us to model more detailed interactions.

To verify that our parser is accurate, we compare our generated network of Hamlet to Moretti’s well-known handmade model of that play (2011). Despite some minor differences in peripheral characters like “Servant”, and our less-minor difference in including the play-within-the-play, the two networks are highly similar. Our network graph supports Moretti’s reading. Our tool also improves on Moretti’s model by adding direction and weight to each connection. Although this level of detail turned out not to be necessary for the basic task of using network graphs to distinguish between Shakespeare’s genres, it may be useful in future work examining a less homogenous corpus of plays, or in work asking different questions about this corpus.

Using networks to identify genre

We then use our generated network graphs to test our central question: whether the social network enacted by a play’s characters can be used as a proxy for features of the play’s narrative content. More specifically, we ask whether social networks can be used to distinguish between the dramatic genres of tragedy, comedy, and history. Using a support vector machine with fivefold validation, we tested 17 different mathematical features of the networks. No single feature was independently sufficient to identify the genre, though graph density came closest (83% accuracy). However, if features are used in combination, the network graphs can indeed achieve full accuracy. One combination of features which does achieve 100% accuracy is edges, words, and degree. We are currently exploring other combinations that might also be capable of accurately identifying genres.

Discussion

History, comedy, tragedy

The potential utility 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, who each interject briefly in a single scene. Connections form chains of acquaintance with little overlap, so even the monarchs have low eigenvector centrality.

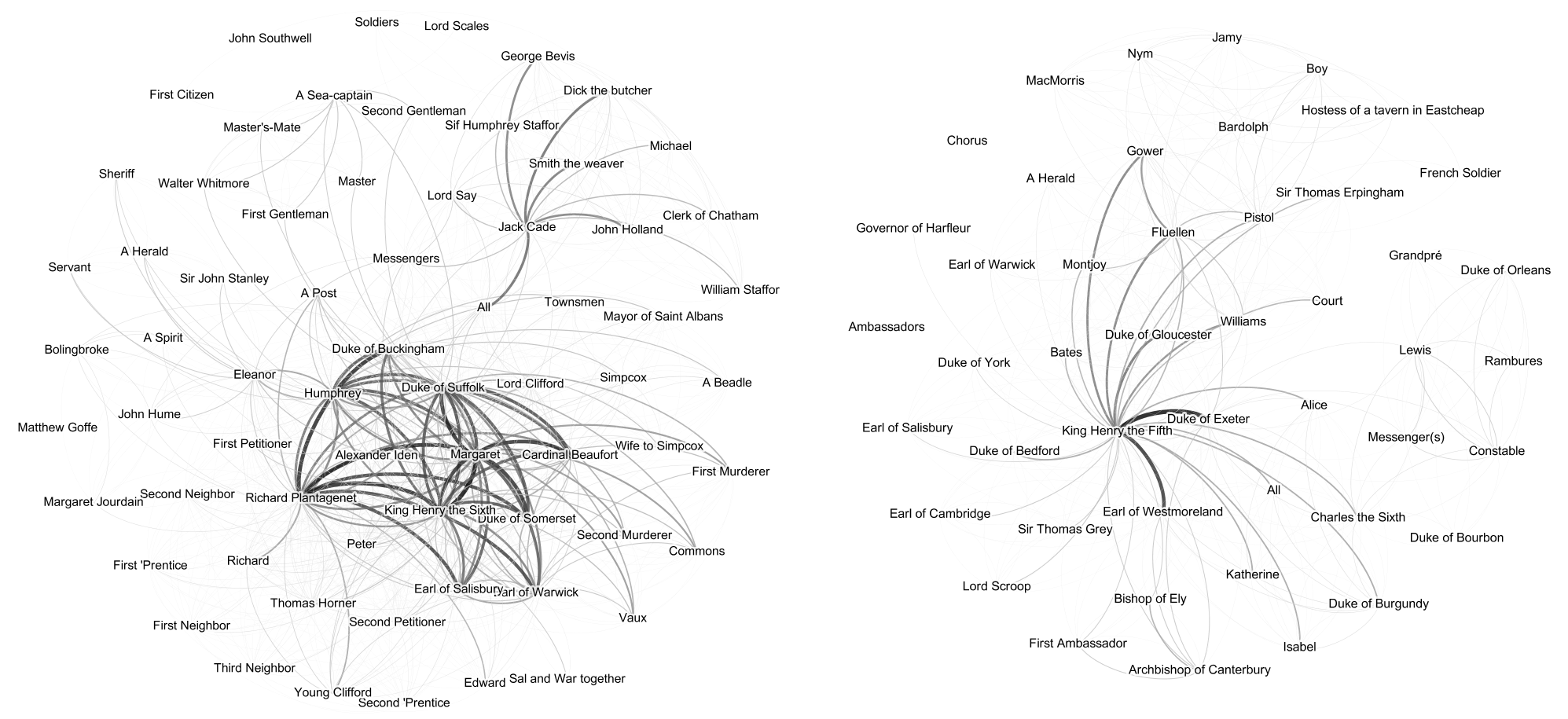


Figure 1: Social network graphs of the histories *Henry VI, Part 2* and *Henry V*.

Comedies, in contrast, feature networks with far fewer characters, in which nearly everybody speaks to nearly everybody else at some point. Although comedies often have multiple subplots, these separate stories do not result in highly-separated networks. We theorize that comedic networks are strongly shaped by the plays’ final “resolution” scenes, which bring together the full cast. The average eigenvector centrality of the characters in comedies is much higher than in tragedies or histories; this suggests that many more of the characters in a comedy are “important,” reflecting a focus on ensemble stories.

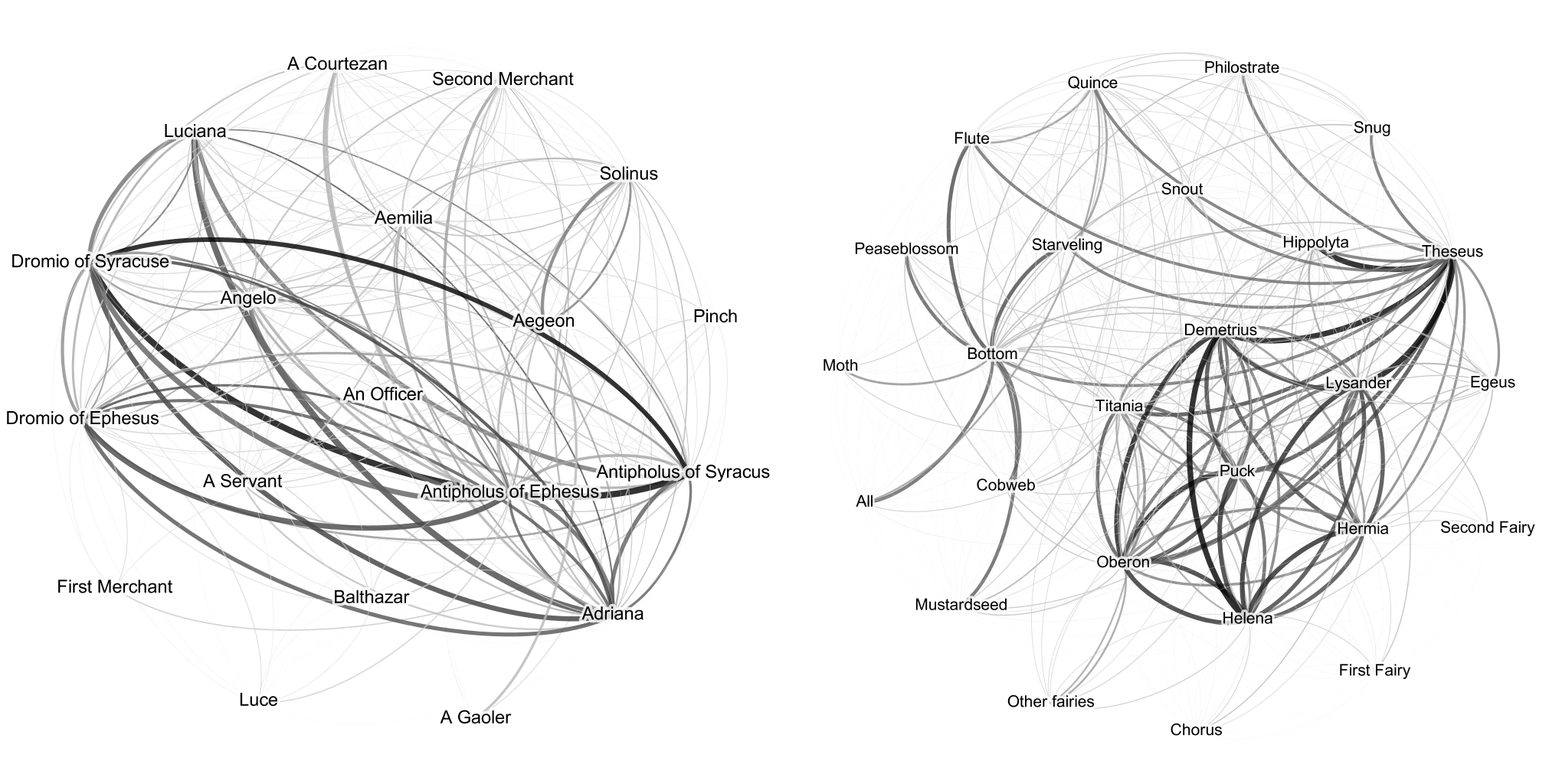


Figure 2: Social network graphs of the comedies *The Comedy of Errors* and *A Midsummer Night’s Dream*.

Graph density is insufficient, however, to fully distinguish the tragedies, which feature networks somewhere between history and comedy in their density. They often have a dense core with a secondary ring of more peripheral characters. What seems to distinguish them is the existence of the central tragic hero, whose influence directly touches more of the network than the protagonists of histories, but whose connections are less interconnected than the ensembles of comedies. These subtleties are better captured, it seems, by the combined metric of “edges, degree, and words.”

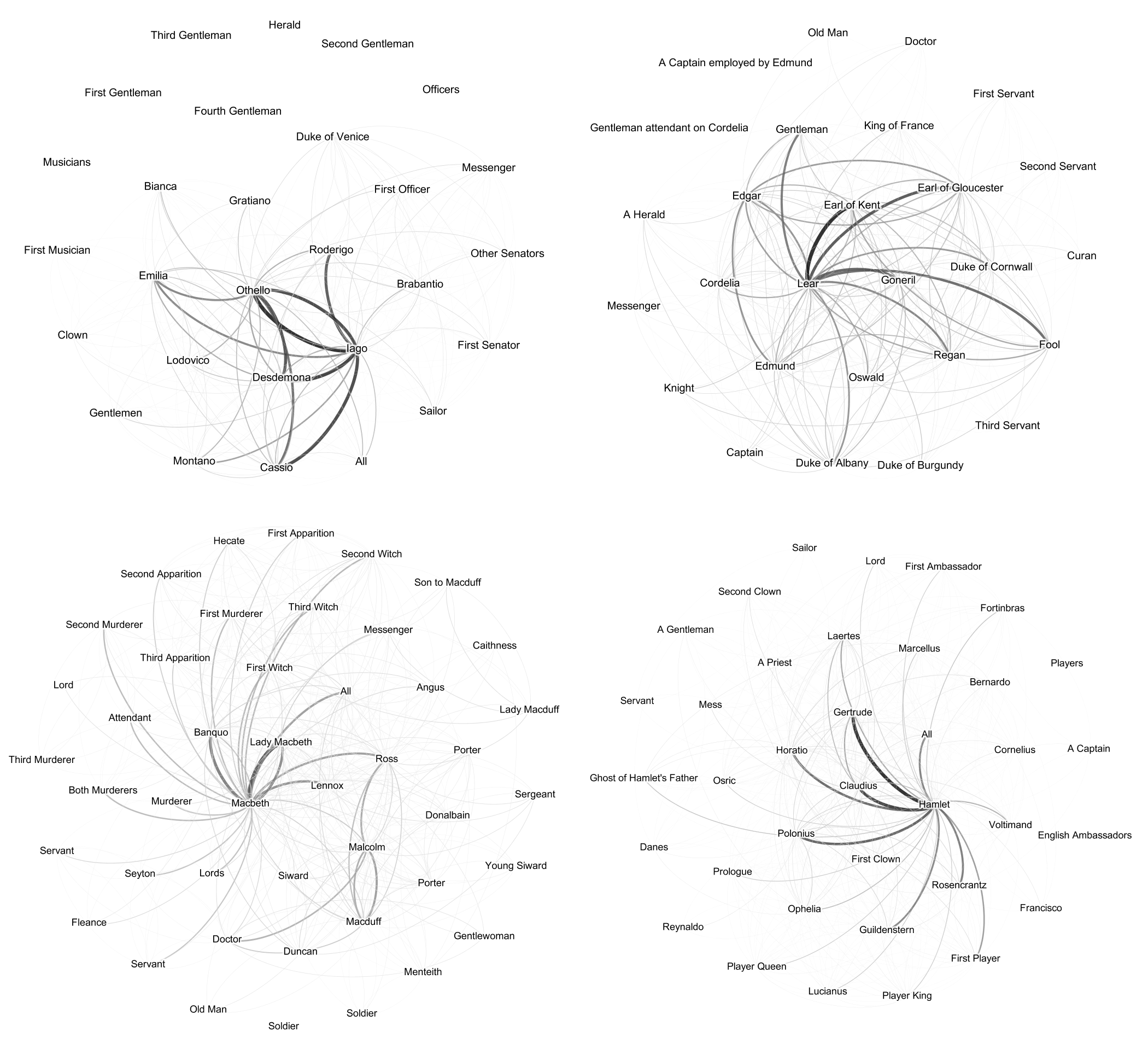


Figure 3: Social network graphs of the tragedies *Othello*, *King Lear*, *Macbeth*, and *Hamlet*.

The "problem plays"

We then use our preliminary identification of each genre’s features to examine Shakespeare’s various contested genres. Training our model only on the plays for which there is strong consensus, we applied it to the “Roman plays,” the “problem plays,” and the “romances” in turn. Of the Roman plays, all but *Antony and Cleopatra* are identified as tragedies by every metric; *Antony and Cleopatra* is identified as a history by “edges, words, and degree” and as a comedy by “degree, modularity, and density”. Of the problem plays, *All’s Well that Ends Well* is always identified as a comedy; *Troilus and Cressida* and *Measure for Measure* are both identified as a comedy by all metrics except for “edges, criticality, and degree”, which identify them as tragedies. The four romances, despite visually unusual networks which support literary arguments that Shakespeare’s writing had grown more experimental at the end of his career, are identified as comedies by every mathematical metric. We treat none of these identifications as definitive declaration of the plays’ “real” genres, but use them to distinguish between plays whose generic ambiguity lies in their subject matter, and plays whose ambiguity lies in their structure.

Conclusion

Our parser successfully and rapidly produces sophisticated social network graphs of TEI plays that can be used to computationally identify theatrical genre in Shakespeare’s plays. Thirty-seven 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. Our networks of the well-studied works of Shakespeare can provide a baseline against which to contextualize analysis of these elements in works for which there is far less critical consensus.

References

Enter your references here:

**Alberich, R.,Miro-Julia, J., and Rosselló., F.** (2002). Marvel Universe Looks Almost Like a Real Social Network. arXiv:cond-mat/0202174v1

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

**Fischer, F., Göbel, M., Kampkaspar, D., Kittel, C., and Trilcke, P.** (2017). Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts. *Digital Humanities 2017: Book of Abstracts*. Montreal: McGill University and Université de Montréal, pp. 437–41.

**Moretti, F.** (2011). Network Theory, Plot Analysis. *New Left Review* 68: 80–102.

**Sparavigna, A. C.** (2013). On Social Networks in Plays and Novels. *International Journal of Sciences*, 2: 20–25.

**Stiller, J., Nettle, D., and Dunbar, R. I. M.** (2003). The Small World of Shakespeare's Plays. *Human Nature*, 14(4): 397–408.

**Tonra, J., Kelly, D., and Reid, L.** (2017). Personæ: A Character-Visualisation Tool for Dramatic Texts. *Digital Humanities 2017: Book of Abstracts*. Montreal: McGill University and Université de Montréal, pp. 627–30.

Finished editing your document?

1. MAKE SURE TO SAVE THIS FILE UNDER A NEW FILE NAME!
2. Go back to the DHConValidator to upload & convert your file and check the output.

Click [HERE](http://adho.org/dhconvalidator-2018popup/DHConvalidatorServices#!converter) to return to the DH ConValidator.