Journal of the Chicago Colloquium on Digital Humanities and Computer Science
Volume 1 Number 3 (2011)

Social Networks and the Language of Greek Tragedy

Jeff Rydberg-Cox, University of Missouri-Kansas City, Department of English and Classical and Ancient Studies Program

Abstract

Using the linguistic dependency treebanks and digitized texts created by the Perseus Digital Library, we are creating social networks for a collection of Greek tragedies that allow users to visualize the interactions between characters in the plays. Because the number of characters who appear on stage in Greek tragedy is limited, most of these social network diagrams fall into a few basic types. The most interesting aspect of these networks are, therefore, the edges that connect the nodes within the graphs. The linguistic data used to label or even create these edges becomes the jumping off point for visualizing and exploring the language of Greek tragedy.

Introduction

Using the linguistic dependency treebanks and digitized texts created by the Perseus Digital Library, we are creating social networks for a collection of Greek tragedies that allow users to visualize the interactions between characters in the plays.[1] Because the number of characters who appear on stage in Greek tragedy is limited, most of these social network diagrams fall into a few basic types. The most interesting aspect of these networks are, therefore, the edges that connect the nodes within the graphs. The linguistic data used to label or even create these edges becomes the jumping off point for visualizing and exploring the language of Greek tragedy.

These social network graphs are designed to chart a middle ground between the emerging distance reading approach adopted by many digital humanists and a close reading approach traditionally adopted by students and scholars in the humanities. As large-scale collections of texts come on-line, one of the most pressing issues to facing scholars in digital humanities is what, exactly, do we do now that vast corpora of primary sources are available in digital form.[2] One approach to emerging

---

[1] There has been no work done on social networks in Greek tragedy, but other scholars have done this sort of work on Shakespeare's plays: J. Stiller, Dr. Nettle, and R. I. M. Dunbar, "The Small World of Shakespeare's Plays," *Human Nature* 14, no. 4 (2003): 397-408; P. Mutton, "Inferring and Visualizing Social Networks on Internet Relay Chat," in *Proceedings Eighth International Conference on Information Visualisation IV* (2004); J. Stiller and M. Hudson, "Weak Links and Scene Cliques Within the Small World of Shakespeare," *Journal of Cultural and Evolutionary Psychology* 3, no. 1 (2005): 57-73; and the literary circles of 18th and 19th century literature: Gillian Russell and Clara Tuite, *Romantic Sociability: Social Networks and Literary Culture in Britain, 1770-1840* (Cambridge, U.K.; New York: Cambridge University Press, 2002). The recent article by Franco Moretti (Franco Moretti, "Network Theory, Plot Analysis," *New Left Review*, no. 68 (2011): 80-102) and the work of David Elson at Columbia exploring the automatic extraction of social networks from 19th century novels texts is the most intriguing current wok in this area (see D. Elson and K. McKeown, "Extending and Evaluating a Platform for Story Understanding," in *Proceedings of the AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II* (2009); D. Elson and K. McKeown, "A Tool for Deep Semantic Encoding of Narrative Texts," in *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations* (2009); D. Elson, Nicholas Dames, and K. McKeown, "Extracting Social Networks From Literary Fiction," in *Proceedings of the 48Th Annual Meeting of the Association for Computational Linguistics*. ACL '10. Association for Computational Linguistics (2010); and D. Elson and K.McKeown, "Automatic Attribution of Quoted Speech in Literary Narrative," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence* (2010).

[2] See Gregory Crane, "What Do You Do with a Million Books?" *D-Lib Magazine* 12, no. 3 (2006) and the accompanying issue of D-Lib.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

vast corpora has been various techniques of distance reading in which quantifiable data such as publication patterns or key words are extracted and visualized. Franco Moretti has pursued this approach in his work on the publication patterns surrounding the emergence of the novel as a coherent genre and the subsequent emergence of genres.[3] While approaches such as these are extremely interesting and valuable, they do not address the questions that a reader might ask as she or he is reading a particular individual literary text. This project aims to find a space between distance reading and close reading; like the distance reading approach, it attempts to discover broad quantifiable patterns within literary texts; like the close reading approach, it attempts to focus on either single literary works or a relatively small collection of literary texts. We hope that quantitative based methods will allow readers to orient themselves within a literary work and make connections between characters. At the same time, we hope that a visualization based approach will make quantitative data about the text more accessible to readers who aren't themselves experts in statistical methods.

Source Data

These social network graphs are based on digital texts and treebanks that have been created at the Perseus Digital Library and released generally under a Creative Commons license.[4] The texts themselves have been part of the Perseus Digital library for many years and are encoded in TEI-conformant XML. The treebanks have also been created by teams working with the Perseus Digital library since 2007. Treebanks are datasets that contain syntactic analyses of the grammatical relationships between every word in a collection of texts along with information about which word depends on which. Teams of scholars and undergraduate researchers have been working on these treebanks since 2007 and they have tagged some 53,000 words of Classical Latin and 192,000 words of Ancient Greek. Note: For larger, higher quality versions of the figures reproduced here, please refer to the *Supplementary Data* section accompanying this article online at http://jdhcs.uchicago.edu

Types of Social Networks in Greek Tragedy

Because the number of characters who appear on stage at any one time is limited in Greek tragedy, their social networks tend to fall into one of four essential types. One type appears in plays where a central character occupies the stage and a sequence of characters appear in-turn to speak to that person, as in Aeschylus' *Prometheus Bound*.

---

[3] Moretti, *Graphs, Maps, Trees : Abstract Models for a Literary History*. Moretti, *The Novel*. Moretti, "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740-1850)".

[4] See D. Bamman and G. Crane, "The Design and Use of a Latin Dependency Treebank," in *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (TLT2006)* (2006); D. Bamman, M. Passarotti, G. Crane, and S. Raynaud, "A Collaborative Model of Treebank Development," in *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories* (December 2007); D. Bamman, F. Mambrini, and G. Crane, "An Ownership Model of Annotation: The Ancient Greek Dependency Treebank," in *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT8)* (2009); and J. Lee and D. Haug, "Porting An Ancient Greek and Latin Treebank." In *Proc. LREC* (2010).

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

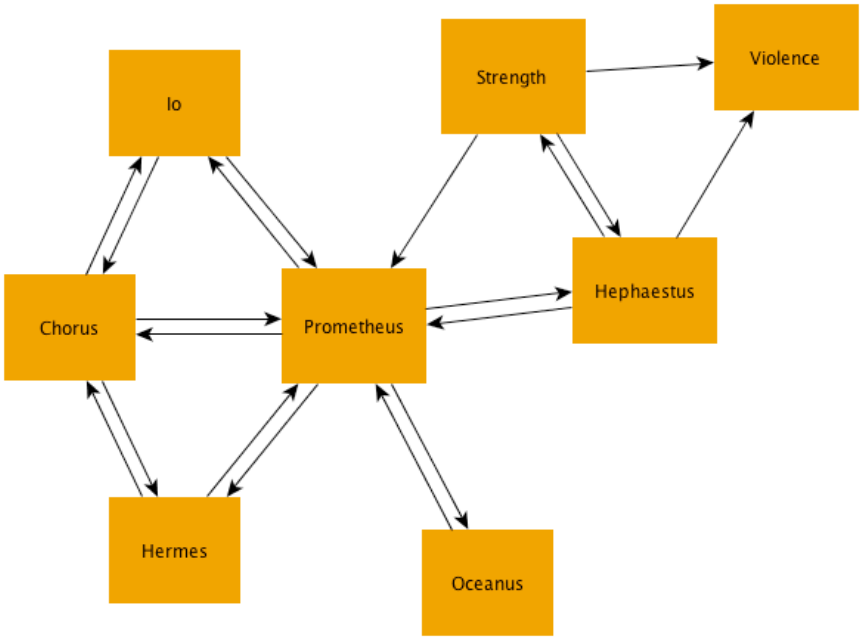

Figure 1. Social network diagram for Aeschylus' *Prometheus Bound.*

The second type occurs when all the characters occupy the stage at essentially the same time and all speak to each other, as in Aeschylus' *Suppliants*.

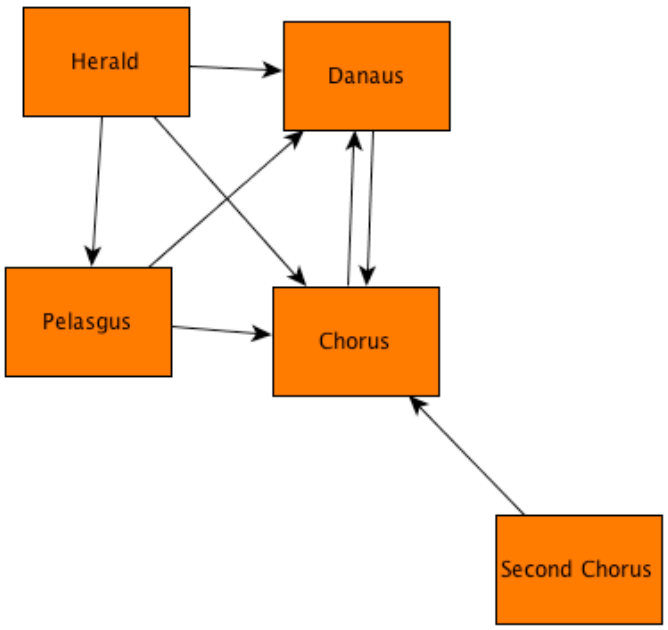

Figure 2. Social network diagram for Aeschylus' *Suppliants.*

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

The third type appears when groups of characters appear on stage in turn and speak to each other with no central character remaining on stage throughout as in Sophocles' *Ajax*.

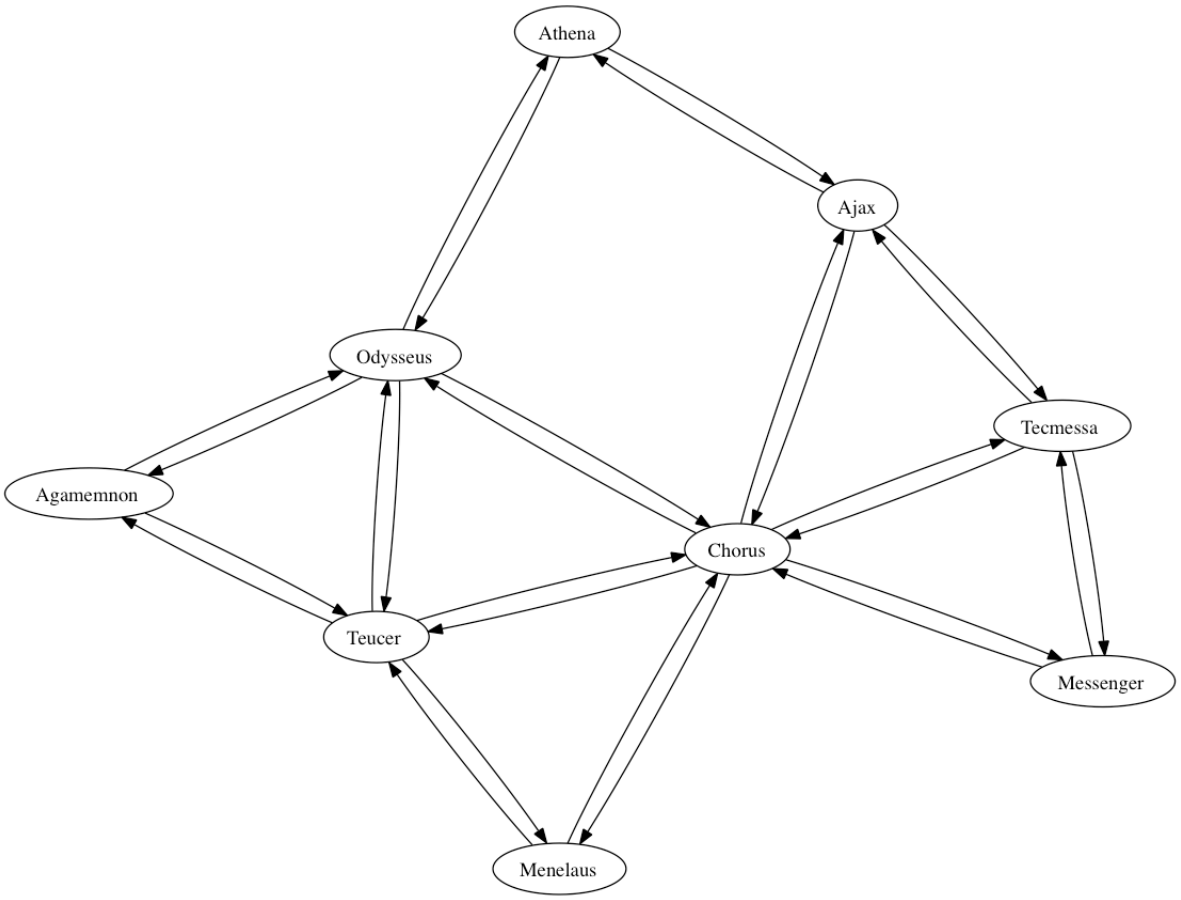

Figure 3. Social network in Sophocles' *Ajax*.

The fourth type appears when there are textual difficulties or anomalies such as the spurious ending to Aeschylus' *Seven Against Thebes* where Antigone and Ismene do not speak to any other characters in the play.

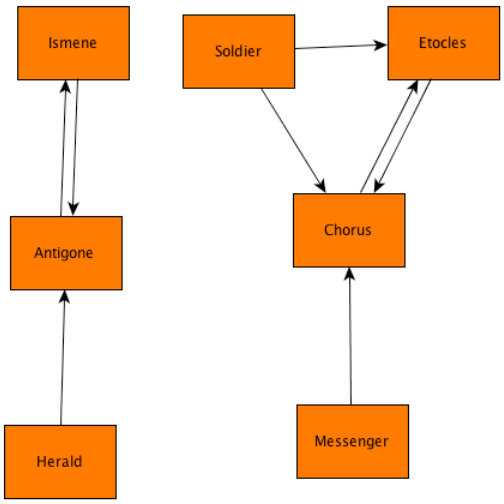

Figure 4. Social network in Aeschylus' *Seven Against Thebes.*

Adding Linguistic Data

Because these graphs fall into a few basic types, the proper method for defining and labeling the edges between the nodes is the most interesting aspect of these visualizations. The social network graph becomes an easily graspable hook to convey other information about the text based on how we label the nodes and the edges. Several evolving prototypes of these social network graphs for tragedies by Aeschylus, Sophocles and Euripides are available on-line at http://daedalus.umkc.edu/VisualExplorer. In these graphs, each web page has a plot summary for the play as a header with a social network diagram such as the one shown below beneath it.

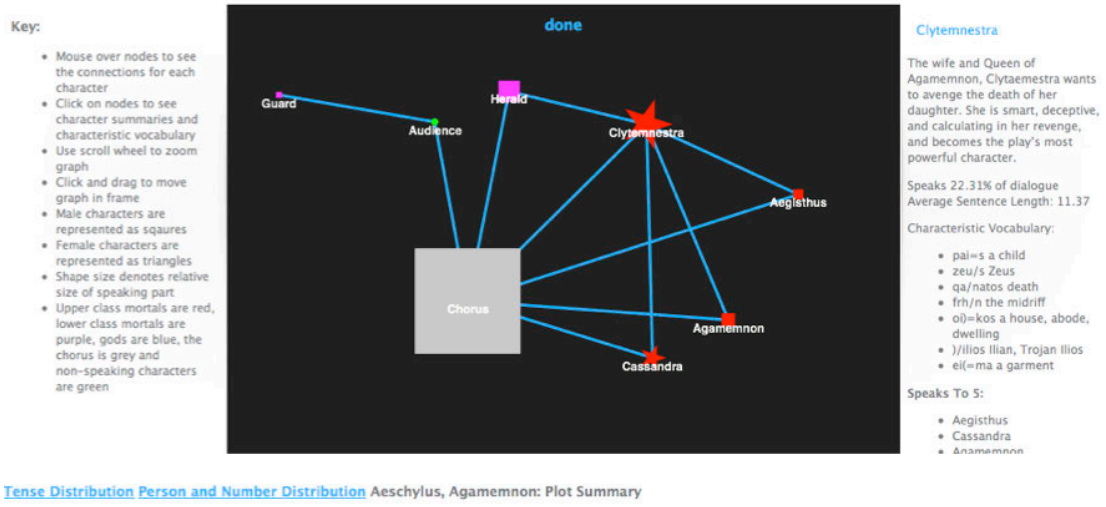

Figure 5. Social network graph for Aeschylus' *Agamemnon.*

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

In this social network, each character is represented as a node in this graph with the size of the node denoting the relative proportion of the dialogue spoken by each character and the color and the shape of the node denoting the gender and social class of each character (upper class mortals are red, lower class mortals are purple, gods are blue, the chorus is grey and non-speaking characters are green). When a user clicks on a node within the graph, character specific data appears in the column to the right. This data includes a custom-written description of the character, data about the average sentence length spoken by that character, key terms associated with the character as calculated using a *TF x IDF* score,[5] and a list of the other characters to whom the character speaks. Additional links at the bottom of each page give access to chart that show the relative distribution of past, present and future verbs among the characters in the play.

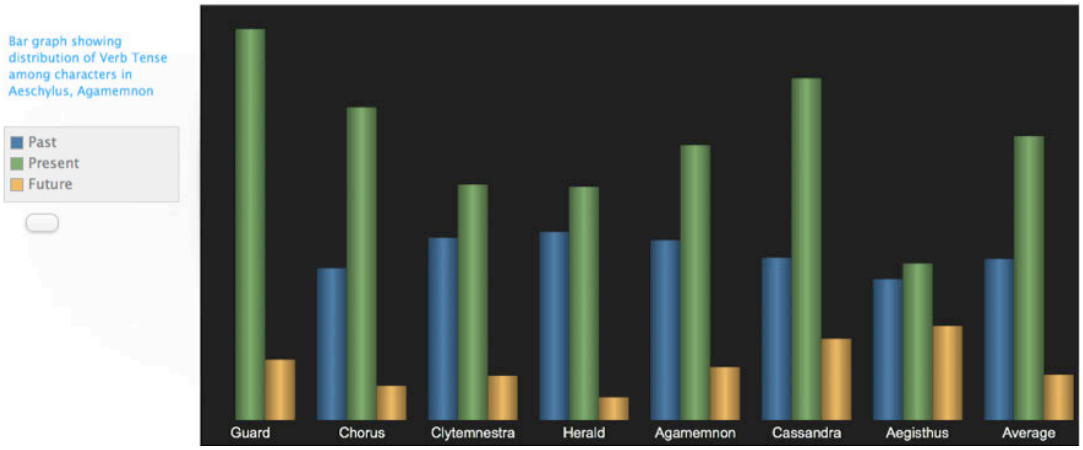

Figure 6. Bar graph showing distribution of verb tense among characters in Aeschylus' *Agamemnon*

A second chart that shows the distribution of verb person and number between the characters in the play.

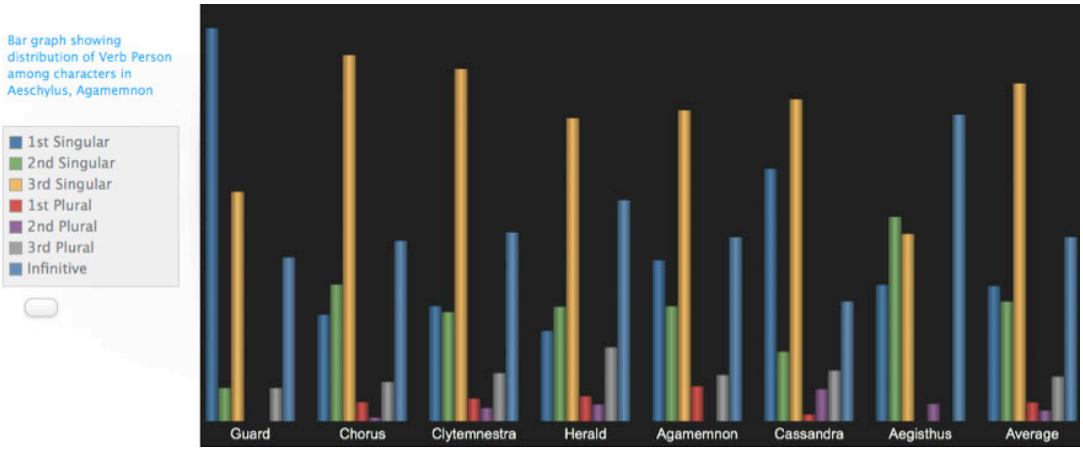

Figure 7. Bar graph showing distribution of verb person among characters in Aeschylus' *Agamemnon.*

[5] For a discussion of the technique used to extract these key phrases, see Jeffrey A. Rydberg-Cox, "Keyword Extraction From Ancient Greek Literary Texts." *Literary and Linguistic Computing* 17, no. 2 (2002): 231-244.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Graphs such as these allow readers to begin to see and consider how quantifiable grammatical features such as these track with literary aspects or plot-lines within the play.

Future Directions

These visualizations continue to evolve and change as we experiment with other visualizations that might be useful to readers working with these texts. This work is moving in several different directions. First, the graphs as shown above provide only raw frequency percentages with no indication of the statistical significance of the variations between the different speakers. We are working on visualizations to integrate t-scores into this graph so that users can see which frequencies fall outside of the expected range. Second, we are exploring other types of data that can be introduced in these graphs such as a vocabulary correlation metric that expresses the degree of overlap between the words used by the two characters and a chart that plots the words shared by each speaker pair allowing readers to see which words are more closely correlated to which character.[6] We are also working on visualizations that incorporate words as if they are actors within the social network. For this visualization, the words most closely associated with each character are calculated as a TFxIDF score with the top five words for each character included in the social network as the object of a social relationship with its speaker thereby serving as intermediaries between the characters in the plays.

[6] See J. F. Burrows, *Computation Into Criticism: A Study of Jane Austen's Novels and An Experiment in Method* (Oxford [Oxfordshire]; New York: Clarendon Press; Oxford University Press, 1987), which constructs these sort of graphs for the very common words in associated with characters in Jane Austen's novels. There are many models for the types of linguistic data that can be graphed in this interface. In addition to Burrows' foundational work, we are looking to the corpus-based approaches to linguistic variation in Douglas Biber, *Variation Across Speech and Writing* (Cambridge [England]; New York: Cambridge University Press, 1988); Douglas Biber, *Dimensions of Register Variation: A Cross-Linguistic Comparison* (Cambridge; New York: Cambridge University Press, 1995); Douglas Biber, Susan Conrad, and Randi Reppen, *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge; New York: Cambridge University Press, 1998); Susan Conrad and Douglas Biber, *Variation in English: Multi-Dimensional Studies* (Harlow, England; New York: Longman, 2001); Randi Reppen, Susan M. Fitzmaurice, and Douglas Biber, *Using Corpora to Explore Linguistic Variation* (Amsterdam; Philadelphia: J. Benjamins, 2002); and Douglas Biber, Ulla Connor, and Thomas A. Upton, *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure* (Amsterdam; Philadelphia: John Benjamins Pub. Co., 2007).

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

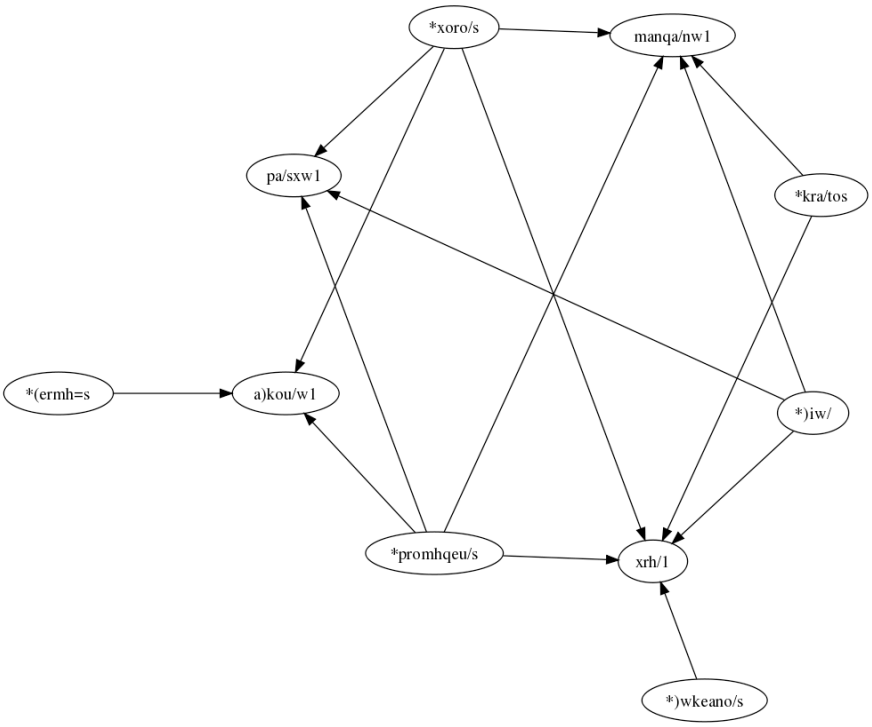

Figure 8. Graph showing connections between the characters in Aeschylus' *Prometheus Bound* with the verbs.

This is a tantalizing possibility that faces some practical difficulties. If we include all words spoken by a particular character, we get an unintelligible graph because there are too many nodes but if we focus on the characteristic vocabulary of each character as defined by the TFxIDF score, there are very few overlaps between characters.[7] Hand-selected lists of words that are interesting to an individual reader seem to provide the most promising results, but this approach does not scale broadly unless it were built as an interactive search facility that suggested candidate words and provided an interactive browsing system to readers engaged in computationally assisted close reading.

Finally, we are also working on expanding this approach to other works in other languages and other genres. This work is also very preliminary, but the early visualizations are very intriguing. If, for example, we look at the *Iliad* and the *Odyssey* where we find a much broader range of characters, the graphs of character relationships at even a very broad level provide us with insight into the nature of these texts. If, for example, we build a similar social network for the *Iliad* and focus on those characters who speak to Achilles or Hector, the initial graph is very interesting and provokes questions about the possible differences in the language used between the characters at the center of this graph as compared to the language used by the characters on the borders.[8]

[7] This is not surprising given definition of the TF*IDF metric.

[8] Hilary Susan Mackie, *Talking Trojan : Speech and Community in the Iliad* (Lanham: Rowman & Littlefield Publishers, 1996) explores some of these questions.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

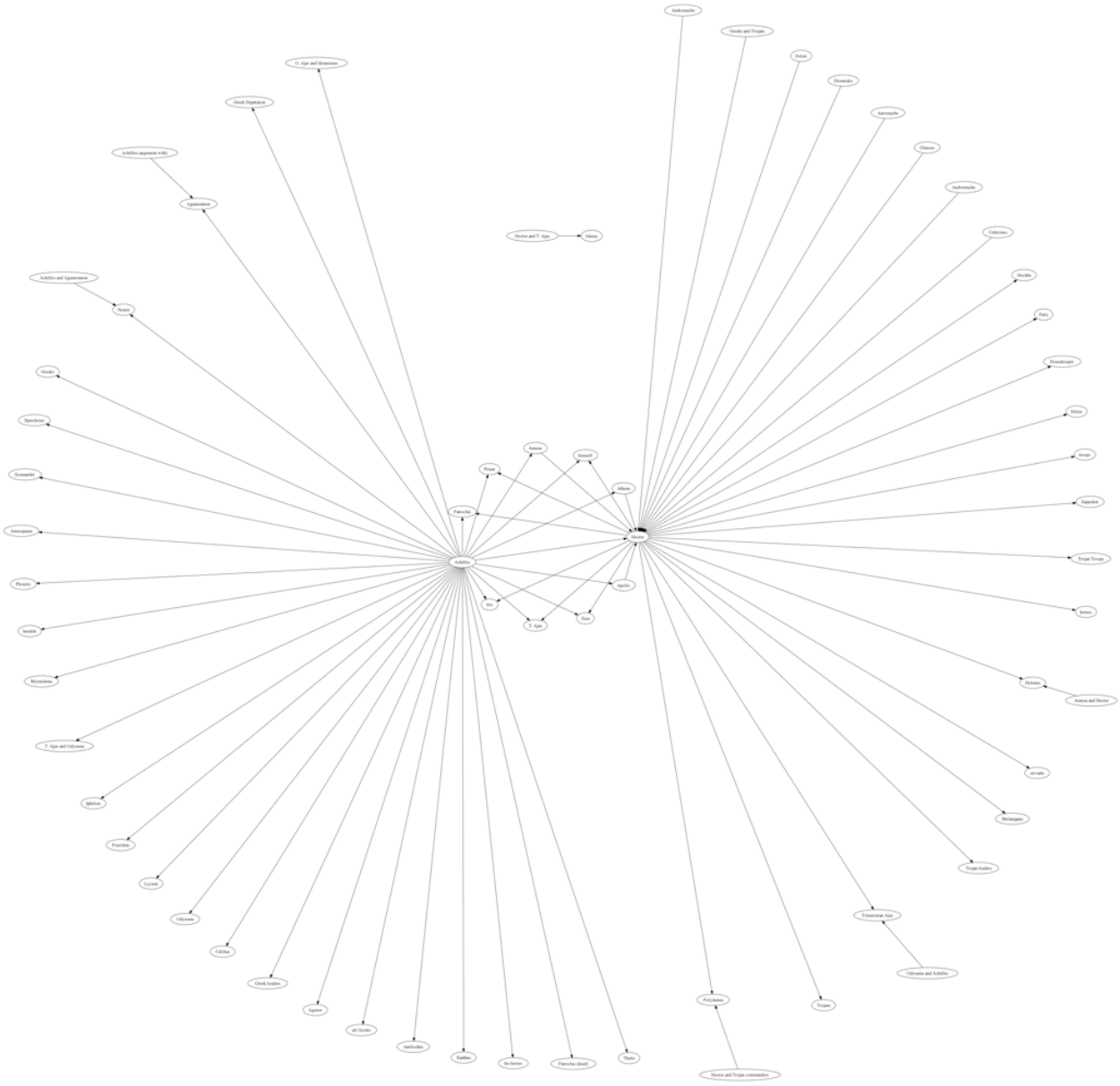

Figure 9. Graph showing those who speak to Achilles and Hector in Homer's *Iliad*

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Bibliography

Bamman, D., and G. Crane. "The Design and Use of a Latin Dependency Treebank." In *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (TLT2006)* (2006).

Bamman, D., F. Mambrini, and G. Crane. "An Ownership Model of Annotation: The Ancient Greek Dependency Treebank." In *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT8)* (2009).

Bamman, D., M. Passarotti, G. Crane, and S. Raynaud. "A Collaborative Model of Treebank Development." In *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories.* (2007).

Biber, Douglas. *Dimensions of Register Variation: A Cross-Linguistic Comparison.* Cambridge; New York: Cambridge University Press, 1995.

—. *Variation Across Speech and Writing.* Cambridge [England]; New York: Cambridge University Press, 1988.

Biber, Douglas, Susan Conrad, and Randi Reppen. *Corpus Linguistics : Investigating Language Structure and Use.* Cambridge; New York: Cambridge University Press, 1998.

Biber, Douglas, Ulla Connor, and Thomas A Upton. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Amsterdam; Philadelphia: John Benjamins Pub. Co., 2007.

Burrows, J. F.. *Computation Into Criticism: A Study of Jane Austen's Novels and An Experiment in Method.* Oxford [Oxfordshire]; New York: Clarendon Press; Oxford University Press, 1987.

Conrad, Susan, and Douglas Biber. *Variation in English: Multi-Dimensional Studies.* Harlow, England; New York: Longman, 2001.

Crane, Gregory. "What Do You Do with a Million Books?" *D-Lib Magazine D-Lib Magazine* 12, no. 3 (2006).

Elson, D. K., and K. R. McKeown. "A Tool for Deep Semantic Encoding of Narrative Texts." In *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations* (2009).

—. "Automatic Attribution of Quoted Speech in Literary Narrative." In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence* (2010).

Elson, D. K., and K. R. McKeown. "Extending and Evaluating a Platform for Story Understanding." In *Proceedings of the AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II* (2009).

Elson, D. K., Nicholas Dames, and K. R. McKeown. "Extracting Social Networks From Literary Fiction." In *Proceedings of the 48Th Annual Meeting of the Association for Computational Linguistics* (2010).

Lee, J., and D. Haug. "Porting An Ancient Greek and Latin Treebank." In *Proc. LREC.* 2010.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Mackie, Hilary Susan. *Talking Trojan: Speech and Community in the Iliad*. Lanham: Rowman & Littlefield Publishers, 1996.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London; New York: Verso, 2005.

—. "Network Theory, Plot Analysis." *New Left Review* , no. 68 (2011): 80-102.

—. "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740-1850)." *Critical Inquiry* 36, no. 1 (2009): 134-158.

—. *The Novel*. Princeton: Princeton University Press, 2006.

Mutton, P.. "Inferring and Visualizing Social Networks on Internet Relay Chat." In *Proceedings Eighth International Conference on Information Visualisation IV* (2004).

Reppen, Randi, Susan M. Fitzmaurice, and Douglas Biber. *Using Corpora to Explore Linguistic Variation*. Amsterdam; Philadelphia: J. Benjamins, 2002.

Russell, Gillian, and Clara Tuite. *Romantic Sociability : Social Networks and Literary Culture in Britain, 1770-1840*. Cambridge, U.K.; New York: Cambridge University Press, 2002.

Rydberg-Cox, Jeffrey A.. "Keyword Extraction From Ancient Greek Literary Texts." *Literary and Linguistic Computing* 17, no. 2 (2002): 231-244.

Stiller, J., and M. Hudson. "Weak Links and Scene Cliques Within the Small World of Shakespeare." *Journal of Cultural and Evolutionary Psychology* 3, no. 1 (2005): 57-73.

Stiller, J., D. Nettle, and R. I. M. Dunbar. "The Small World of Shakespeare's Plays." *Human Nature* 14, no. 4 (2003): 397-408.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago