

Networks of Stories, Structures and Digital Humanities

Almila Akdag Salah

almila.akdagsalah@vks.knaw.nl
KNAW Royal Netherlands Academy of Arts and
Sciences, Netherlands

Wouter De Nooy

w.denooy@uva.nl
UVA, Faculty of Social and Behavioral Sciences,
Netherlands

Zoe Borovsky

zoe@ats.ucla.edu
UCLA, ATS, LA, USA

In March 2004, Slashdot, the online news aggregator 'for nerds', cited a web page featuring animated network visualizations of the relationships between characters in Shakespeare's plays.¹ The Shakespeare site began getting hits at a rate of 250,000 hits per hour. Programmer Paul Mutton created the diagrams by feeding Shakespeare's plays into an IRC (Internet Relay Chat) bot designed to visualize social networks. The Slashdot story promised that the diagrams would allow users to "see Shakespeare in an entirely new light".²

Social network analysis, the mapping of relationships as networks, has truly provided new insights in the social and life sciences with increasing participation from mathematicians and computer scientists. The idea that there are laws that govern networks--such as the notion that everyone is at most six steps away from any other person on earth, and that networks evolve in predictable ways--has led to remarkable discoveries in fields from sociology to biology. These insights furthered publications such as *Six Degrees* (by sociologist Duncan Watts)³ presenting these discoveries (and the science behind them) to a more general audience. In addition, online communities such as Facebook, Twitter, MySpace, etc., have demonstrated the utility and power of network theory in our daily lives.

With popular awareness of these tools and techniques, combined with the promise of new insights into ever-increasing amounts of data, network analysis has significant appeal to digital humanities scholars. However, these investigations call for tools and software to gather and clean large amounts of data, even when the researcher opts for analyzing only a tiny fraction of these vast data sets. Moreover, to interpret these datasets, a good understanding of network analysis, as well as other visualization means is necessary. Overcoming the hurdle of handling complicated technological tools and acquisition of the necessary expertise in network theory both demand investment on the side of researchers, a risky investment that only a handful of humanities scholars are willing to make.⁴ Typically, researchers familiar with the tools perform these analyses, posing the questions and interpreting the results. In this session, we will first provide a brief overview of social network analysis and related tools. Then, we will focus on three examples of how we, as humanities scholars, have made use of network analysis tools and techniques in our research, both to illustrate the potential of these approaches, and to discuss some common problems and their possible solutions.

The names and affiliation of confirmed authors are as follows:

- Zoe Borovsky, Academic Technology Services and the Center for Digital Humanities, University of California, Los Angeles
- Wouter de Nooy, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, The Netherlands
- Loet Leydesdorff, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, The Netherlands
- Andrea Scharnhorst, Virtual Knowledge Studio of the Netherlands Royal Academy of Arts and Sciences, The Netherlands
- Almila Akdag Salah, Virtual Knowledge Studio of the Netherlands Royal Academy of Arts and Sciences, The Netherlands

Notes

1. <http://www.jibble.org/shakespeare/>

2. http://tech.slashdot.org/story/04/03/11/0151256/Tracking-Social-Networking-In-Shakespeare-Plays?art_pos=24 Mutton also presented his work at a conference on Information Visualization in 2004. See <http://www.cs.kent.ac.uk/pubs/2004/1931/content.pdf>
3. Watts, D.J. (2003). *Six Degrees: The Science of a Connected Age* 1st ed., W. W. Norton & Company
4. Social network analysis is a more popular tool among social scientist than humanities scholars. A search at Web of Science for the search term returns around 150 hits for social sciences, whereas if the search is refined to arts and humanities only, it has less than 30 results.

Dickens' Double Narrative: Network Analysis of Bleak House

Zoe Borovsky

zoe@ats.ucla.edu

Academic Technology Services and the
Center for Digital Humanities, University of
California, Los Angeles

My presentation extends Masahiro Hori's collocational analysis of Dickens' style (2004)¹ using network content analysis tools. Because DH2010 will be held in London, and since Hori devotes a chapter in his book to Dickens' novel, *Bleak House*, I will focus the presentation on this novel. My goals are three-fold: First, I will show how network analysis provides additional insights and metrics for studying the collocational patterns in texts. Secondly, I will demonstrate how this analysis provides new insights into the gender issues in Dickens' novel. Finally, I will extend my analysis to the field of Digital Humanities and reflect on how it, too, can be read as a "double-narrative" with a deep structure similar to the one Dickens depicts in *Bleak House*.

Network Analysis has been applied to many different types of systems: sociological, biological, ecological, as well as the Internet. In addition, human language has been examined for evidence of network behavior—the laws that have been shown to govern other complex systems. Beginning with the 2001 publication of "The Small World of Human Language".² I will present a brief overview of scholarship on network analysis to provide some background. Typically, humanities scholars have used

network analysis to model networks described in texts (e.g. the characters in Shakespeare's plays).³ However, the main purpose of this paper is to demonstrate the implications of these more theoretical studies upon text analysis.

Having provided the context for my analysis, I will show how Wordij,⁴ a suite of network content analysis tools designed to analyze unstructured texts, can be used to extend Hori's analysis of the collocates⁵ of the two narrators in Dickens' text. Against the prevailing view that the first-person narrative of the central female character, Esther, is plain and boring, Hori compares the unusual collocates of both narrators. His careful analysis reveals that Esther's narrative is "linguistically experimental and satisfactorily creative" (Hori 2004, p. 206). My analysis builds upon Hori's and illustrates that through the course of the novel, the narrative of the anonymous third-person narrator becomes "tainted" with collocates that have led scholars to judge Esther's narrative as overly sentimental and boring.

Because Wordij allows us to analyze texts as a network, with links created between collocates, we can use it to calculate the distance between concepts that are connected through other links. Using the network diagrams produced from collocates of both these narrative sections, I will demonstrate how this type of "semantic map" can be used to compare the distance between words. As examples, I compare the distance between the words "man" and "woman" in each narrator's section. Opticomm (one of the programs in Wordij) measures the distance between nodes (the words) by counting the number of links between words and calculating the distance or path and the average pair frequency.

For example, one of the shortest paths from the word "man" to "woman" in Esther's text takes the following route: man→said→woman. Opticomm calculates the "distance" using the frequency of the pairs. The pair "man→said" occurs 27 times ($1/27 = .0370$), and the pair "said→woman" occurs 19 times ($1/19 = .0526$). The sum of the distances equals .0897. The average pair frequency of the path is low: 23. A higher frequency path takes advantage of more frequently used pairs. One of the "hubs" (a word that has many connections) in Esther's

narrative is the word “little”. By adding the word “little” to the path (man→said→little→woman), the distance between “man” and “woman” becomes shorter (.0673) and more frequent (53.6667) [See Figure 1].

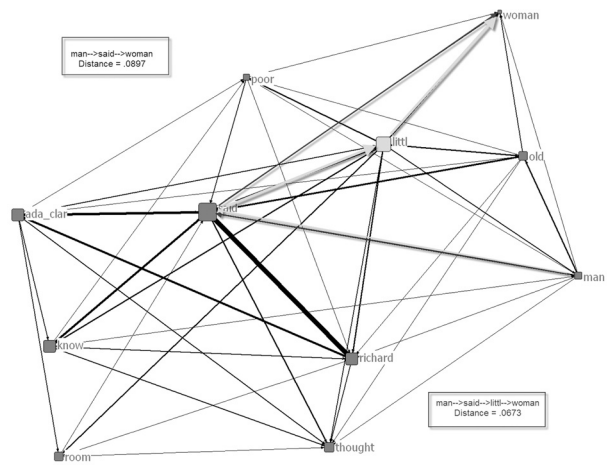


Figure 1: Network of collocates of Esther's narrative, Dickens, *Bleak House*. Visualization is created with NetDraw 2.087 using collocates and distances generated with Wordij. Frequencies of the pairs (shown by weighted lines) are greater than or equal to 18.

The results of this analysis show how the third-person narrator's change in style (employing collocate patterns used in Esther's text) reflects and mirrors the character development in the novel—the rezoning of the hyper-masculine characters: Sir Leicester Dedlock and Mr. George. I will argue that the deep structure of the novel hinges on this “rezoning” of the masculine— a plot structure that is remarkably similar to the occult horror films analyzed by Carol Clover in her book: *Men, Women and Chainsaws: Gender in the Modern Horror Film*.⁶ Like occult films, I will present that the aim of *Bleak House* is a reconstruction of the masculine, a process that typically takes place almost invisibly, as the text foregrounds the excesses of the female characters. Thus the narrative technique, one that “opens up” to the sentimental expressions characteristic of Esther's narrative, reflects the deep-structure “rezoning” on the periphery of *Bleak House*; George's Shooting Gallery and Sir Leicester's Chesney Wold are radically transformed in the novel.

Finally, I will draw comparisons between the cultural conditions that evoke these types of structures and double narratives, and reflect on how the field of Digital Humanities can be read in this light. Scholarship in Digital Humanities

is a “double-narrative” brought about by the rezoning of hard and soft sciences, a project similar to the reconstruction undertaken by Dickens' *Bleak House*.

Notes

1. Hori, M. (2004). *Investigating Dickens' style*, Palgrave Macmillan.
2. Cancho, R.F.I. & Solé, R.V. (2001). The Small World of Human Language. *Proceedings of the Royal Society of London. Series b, biological sciences*, 268, 2261--2266.
3. Stiller, J., Nettle, D. & Dunbar, R. (2003). The small world of Shakespeare's plays. *Human Nature*, 14(4): 397-408.
4. Danowski, J., WORDij, Chicago: University of Illinois at Chicago. Available at: <http://wordij.net/>
5. Collocates, or word-pairs, have been defined and analyzed in various ways. Hori provides a useful short history in his book. Wordij generates collocates or pairs of words that occur within a user-specified span of surrounding text. See Danowski, J. (2009). Network analysis of message content. In Krippendorff, K & Bock, M. (eds), *The content analysis reader*. Sage Publications, pp. 421-430 for a detailed description of how Wordij generates and ranks collocates.
6. Clover, C.J. (1993). *Men, Women, and Chain Saws: Gender in the Modern Horror Film*. Princeton University Press.

Network analysis of story structure

Wouter de Nooy

w.denooy@uva.nl

Amsterdam School of Communication Research (ASCoR), University of Amsterdam, The Netherlands

Relations and structure play an important role in the humanities. Language is structured at the word, sentence, text levels, and beyond. Literary texts contain implicit references to one another and artists are linked by lines of influence or artistic descent. In some branches of structuralism, meaning is inextricably intertwined with the relations among linguistic elements (signifiers). In this paper, I extend the idea that structure is the carrier of meaning to stories. I submit that story structure conveys meaning. Patterns of interactions among the characters of a story are meaningful in the sense that they represent roles with moral connotations.

Structure quickly becomes complex if the number of elements and relations increases.

Then, network analysis is an indispensable tool. In a network, entities such as characters in a story are represented by nodes (vertices) and lines embody links between these characters. There is no limitation to the kinds of links that can be represented by lines as long as each link connects two entities. Durable ties, for example, family relations among characters, can be expressed in this way just as easily as incidental events or interactions. Lines can be symmetric, for example, being siblings, or asymmetric, such as an action from one character directed towards another character in the story. Lines can have attributes and they can be of different kinds, for example, positive versus negative lines (this is called a signed network). Finally, each line can be time-stamped to indicate the onset and expiration of ties. With this formalization of structure as a network, we can use computers to quickly and correctly identify structural patterns. Software that is able to analyze large and complicated networks is widely available.

If network analysis yields a tool for finding patterns, does it also tell us which patterns to look for? In an analysis of story structure understood as the interactions among characters, it makes sense to borrow from the tradition of social network analysis. In this paper, I use the theory of structural balance, which originated in (social) psychology and was adopted in social network analysis. Based on simple premises like 'the friend of my friend is my friend,' this theory predicts that people favour balanced networks. A network is balanced if the positive (friendship) and negative (antagonism) ties display specific, easily recognizable patterns.

Balance theory has been applied to the affective ties among characters in stories: tales, opera, and movies. The general finding is that balance among the story's protagonists is restored at a story's resolution. Stories develop from unbalanced towards balanced networks. In this paper, I focus on another important role of time. I show that the time order and direction of positive and negative actions between two protagonists defines the roles they play in the story. I use the seminal work of Vladímir Propp—*Morphology of the Folktale* (1928)—which inspired most of the later work on story grammars and narratology. Propp

offers a concise set of propositions on the basic form and plot of a corpus of Russian fairytales, distinguishing between several roles of protagonists, such as the hero, the villain, and the donor. I demonstrate that each role is associated with a unique sequence of positive and negative actions. In other words, one can tell a fairytale character's role from the interaction sequence it is involved in.

A formal focus on structure, which is needed for computerized network analysis, entails that attributes of the characters, such as a hut on chicken legs indicating that a character is a witch, and detailed content of interactions are not taken into account. Instead, the focus is on an abstract schema that can be fleshed out in any number of ways in a story. Notwithstanding their abstract character, I contend that the schemata have moral implications. Attacking the story's hero out of the blue, which is a characteristic of the structural pattern associated with the villain, has a negative moral connotation, which is also linked to the last step in the villain's structural pattern, namely that the villain is defeated by the hero at the end of the story. Undesirable properties of the villain, such as ugliness, may help to reinforce the moral depreciation of the role but they are not essential to it.

Tales and other stories socialize children into a culture of appreciating some sequences of interaction and depreciating other sequences. Authors may play with these conventions by combining structural role patterns with attributes of the character that are usually associated with other roles. In a more fundamental way, authors may turn structural patterns upside down, for example, a woman accepting to live with the man who raped her (a typical villain role) in Coetzee's *Disgrace* or the protagonist in Houellebecq's *Elementary Particles*, who turns his back on his most important 'donor.' Fairy tale structure is probably quite straightforward in comparison to stories read by adults. A systematic analysis of more complicated stories requires formal network analysis and so does the detection of conventional role patterns in sets of stories, differences in this regard between story genres, and development over time.

If the role patterns in stories are instruments of socialization, we may expect to find them

in real-life interactions. Because role patterns and network analysis are purely structural, it is straightforward to compare the structure of interactions within stories to interactions among humans. In this paper, I will apply network analysis to evaluations among a set of literary authors and critics to show that fairytale roles can be distinguished and help to explain what happens. In this way, the social sciences benefit from the analysis of story structure just like social theory and social network analysis may advance the analysis of story structure.

Mapping the Flow of “Digital Humanities”

Alkim Almila Akdag Salah

almila.akdagsalah@knaw.vks.nl

Virtual Knowledge Studio of the Netherlands
Royal Academy of Arts and Sciences, The
Netherlands

Loet Leydesdorff

loet@leydesdorff.net

Amsterdam School of Communications
Research (ASCoR), University of Amsterdam,
The Netherlands

Andrea Scharnhorst

andrea.scharnhorst@vks.knaw.nl

Virtual Knowledge Studio of the Netherlands
Royal Academy of Arts and Sciences, The
Netherlands

In this paper, we propose novel ways of charting out activity in digital humanities research. In particular, we apply a combination of classic bibliometric analysis¹ focusing on citation patterns and a broader text and actor network analysis based on various data sets collected from different sources, such as standardized scientific information databases and web data in general.

"Digital humanities" is the latest term for "humanities computing", almost a century old enterprise that made use of computational methods and tools in humanities research. Today, this enterprise has become a hot topic supported by significant government funding agencies like National Endowment in

Humanities (NEH) in the United States² and the European Framework programs of EU.³ The topic attracts attention from scholars coming from a wide range of disciplines like computer science and art history, and its future is the focus of a heated debate on certain weblogs dedicated to this research initiative. Depending on the research environment, the idea of Digital Humanities (DH) is envisioned in diverse ways; some research institutes or universities view it as a methodology, others incorporate it into their research agendas as a tool for analysis. Others view it as a discourse, investigating how knowledge is produced with new media technologies, while at the same time making use of these technologies in humanities research itself.

Usually, the applications and methodology of DH ask for a collaborative commission, and these interdisciplinary collaborations forge beneficial links that enable the use of technological tools in a way to serve humanities research questions. However, recent findings (Leydesdorff & Akdag Salah, in preparation)⁴ show that the scholarly activities around DH are less broad than one could expect. One might even say that DH falls short of reaching out to an interdisciplinary audience. A bibliometric study of the citation patterns of papers that are tagged in Web of Science's all three databases under the topic of "digital humanities" reveals that the knowledge base of DH is confined to two groups of journals, namely the ones that are dedicated to humanities computing, and the others dealing with library and information sciences (see Figures 1 & 2). This outcome should not be surprising, as research in the arena of humanities computing was centered upon text mining and analysis of humanities literature, and the publication of these research results were confined to specialized journals such as *Humanities Computing*. Beside text analysis, a main focus of DH activities was the digitization of archives and libraries, hence the second group of journals are specialized in this area.

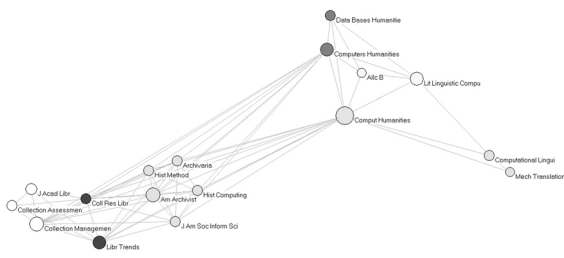


Figure 1: Journal co-citation patterns of 33 documents citing 46 documents about “Digital Humanities” in 2008; threshold 0.5%; cosine \geq 0.0; N = 15, from Leydesdorff & Akdag Salah (forthcoming).

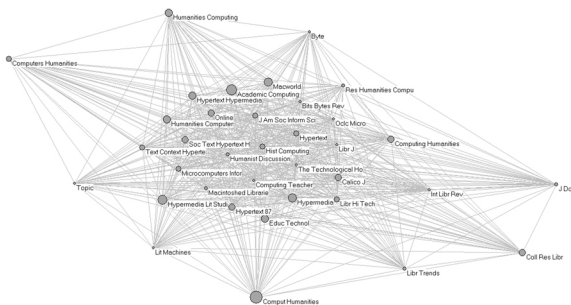


Figure 2: k-core group of 36 journals bibliographically coupled in 46 documents about “digital humanities”; threshold 0.5%; cosine > 0.2, from Leydesdorff & Akdag Salah (forthcoming).

Currently, it is not clear whether DH is in the first formations of either becoming a discipline on its own, or whether it will become merged into the infrastructure of future humanities departments by offering a platform for collaborative research environment enriched by media technologies of today. In either case, our initial findings on DH’s position and its present inability to reach an interdisciplinary audience led us to conclude that policy makers and scholars, both of whom contribute to the development of DH could benefit from an expansion of our analysis.

We assume that DH activities that are very innovative and still in an emergent state might not be fully reflected by the journals sampled in *Web of Science*. We furthermore suppose that a more in-depth analysis that goes beyond journal-journal citation networks will reveal interesting patterns. We will present the results of a social network analysis and a semantic map of DH based on a dataset extracted from papers published on leading DH journals such as *Humanities Computing*, *Calico*, *Literary and Linguistic Computing*, as well as papers retrieved from online publication venues such

as *Text Technology*, *Digital Studies*, *Digital Medievalist*. We believe that online journals will alter the recent findings of the knowledge dissemination of DH to a considerable extend. Here one should not be limited with classical database providers like *Web of Science* or *Scopus*, but also make full use of Google Scholar search-engine. Even then, a bibliometric study of the topic confined to journal-journal citation maps will not deliver a complete picture of the flux of DH. To capture this change and the debate around it, we propose a methodology that will incorporate different type of social networks in rendering and interpreting this new research area.

DH scholars can use virtual research environments for collaboration and production of knowledge, as well as for publication of their results. Hence, information retrieval from online communities, mailing lists, discussion forums and blogs dedicated to DH research is crucial for understanding the state of the art in projects, programs, and research done in the name of DH. A social network based on the shared keywords of such a dataset reveals what topics are of recent interest to DH scholars. Furthermore, this dataset is used in preparing a second network that shows the links of websites and blogs (in and out), which facilitates the drawing of information flow inside DH community, as well the community's relation to other scholarly communities.

A very interesting extension is a map that depicts the connection between institutions and the resources they draw upon. As we have already mentioned, the structural formation of DH is different in every institution. Thus, every DH research environment is built on different premises, and the development of each of these settings is mainly based on internal resources (which are usually delivered in terms of equipment or the availability of technical staff and scholars) and on external support (governmental or private agencies). Today, when the main attention of DH scholars are on combining their past experience and their ideas about DH to create a better platform for humanities departments in general, to depict the flow of resources, financial and otherwise, is as vital as following the flow of information in DH.

Notes

1. 'Scientometrics' or 'bibliometrics' is a research venue specialized in evaluating growth, relations and interactions in scientific fields with the help of citation data collected above all from journal papers. Scientometrics is traced back to the beginnings of the 1900s, but the more official start can be settled to 1964, when Garfield founded Institute for Scientific Information; today it is a product of Thompson ISI and called Web of Science. From the 80's onwards, the research in this area accelerated with the advancement of computers and various combinations of statistical methods used to extract and evaluate information such as citations, cocitations with reference of various bibliometric data. The end-results are usually rendered as so-called 'citation networks' which are a variation of social networks. Now it is a common practice to evaluate a scholar or a journal according to how many times it/he/she is cited. To read more on the history of scientometrics see Katy Börner, Jeegar T. Marus, and Robert L. Goldstone. (2004). The Simultaneous Evolution Of Author And Paper Networks, PNAS 101 (suppl.1): 5266-5273. For more information on citation networks, please see Doreian Patrick. (1985). A Measure Of Standing Of Journals in Stratified Networks. *Journal of the American Society for Information Science*, 8(5/6): 341-363. For a critique of bibliometric analysis see Lindsey D. (1989). Using Citation Counts As A Measure Of Quality In Science Measuring What's Measurable Rather Than What's Valid., *Scientometrics*, Vol. 15, No 3-4: 189-203; Leydesdorff L. (2006). Can Scientific Journals Be Classified in Terms of Aggregated Journal-Journal Citation Relations Using the Journal Citation Reports? *Journal Of The American Society For Information Science And Technology—March*: 601-614.
2. NEH has an office dedicated for Digital Humanities research: <http://www.neh.gov/whoweare/divisions/DigitalHumanities/index.html>
3. One of the FP7 projects is titled PREPARINGDARIAH (Preparing for the construction of the Digital Research Infrastructure for the Arts and Humanities) and is devoted to enriching DH: <http://www.dariah.eu>
4. Leydesdorff L., Akdag Salah, A. (forthcoming). Maps on the basis of the Arts & Humanities Citation Index: the journals Leonardo and Art Journal, and "Digital Humanities" as a topic, *Journal of the American Society for Information Science and Technology*, available at <http://www.leydesdorff.net/ahci/ahci.pdf>