Corpora and Complex Networks as Cultural Critique: Investigating Race and Gender Bias in Graphic Narratives

Alexander Dunst
dunset@mail.upb.de
University of Paderborn, Germany

Rita Hartel
rst@uni-paderborn.de
University of Paderborn, Germany

Introduction

This paper reports on efforts to integrate cultural critique into a DH project that analyzes a corpus of book-length comics, or graphic narratives. We argue that the analysis of issues such as gender, race, and class should be central to digital scholarship that aims to become accessible to the public and appear relevant to the humanities at large. Therefore, cultural criticism needs to be integrated into digital projects from the very beginning. Our research takes up calls to "design for difference" and to develop visualizations that "enact [the] humanistic properties" of complexity and contradiction (McPherson and Drucker, 242). Part I looks at the construction of a corpus of graphic novels, memoirs, and non-fiction. Basing our corpus on academic databases, international comics awards, literary histories, and online booksellers provides insight into institutional gender and racial biases, as well as the opportunity to address them. Part II takes up Drucker's criticism of network analysis as reductive and static. We present networks of a pilot corpus that pay attention to social inequality and replace reductive edges with distinct forms of communication. Part II exemplifies our intention to make DH scholarship relevant to a wider public. The broad appeal of comics presents an ideal point of entry for people who might not otherwise be interested in digital research. We apply the popular Bechdel Test (proposed by Alison Bechdel in a comic strip but used mainly to study film and TV), to highlight the male bias of graphic narratives.

Corpus Analysis of Institutional Gender and Racial Bias

The traditional canon of literary studies has long been criticized for its exclusion of female, non-Western, and minority authors. As a much younger field, comics studies lacks the extensive canons and bibliographies produced by literary historians. This does not mean that similar biases are absent, however. As part of a larger project, we have built a monitor corpus of book-length comics by drawing on sources that include academic databases (JSTOR, MLA), international comics awards, literary and cultural histories of comics, news media coverage, and Amazon.com (Dunst et al.). Of 220 titles included in the corpus by fall 2016, 84 per cent were written by male authors and 73 per cent were identified as white. Biases are unevenly distributed:

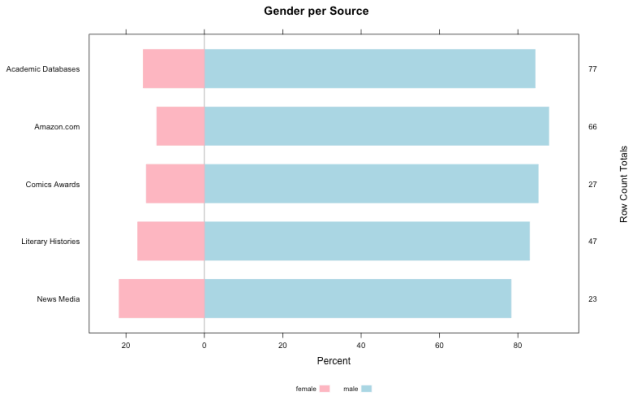

Figure 1: Gender of issue's author grouped by book's source

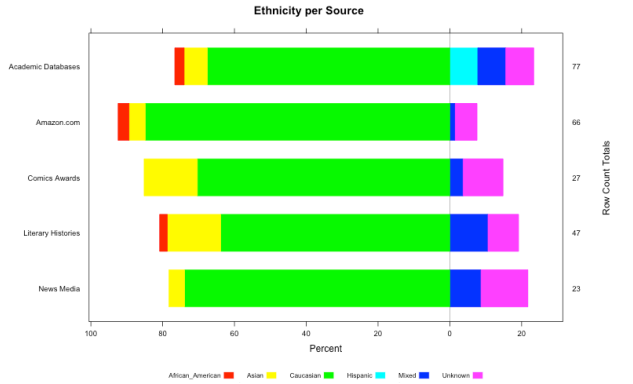

Figure 2: Ethnicity of issue's author grouped by book's source

The absence of reliable bibliographies means that the size, gender and racial make-up of the population remains uncertain. Yet given the considerable differences between sources, institutional biases appear likely. To address these existing biases, two steps were undertaken. A survey sent to comics scholars (five female, five male) asked them to suggest between five and ten graphic narratives written by women that should be included in the corpus. Of a total of 53 suggestions by nine

respondents, nine volumes were listed by more than one scholar and 12 had already been included in the corpus, while 14 fell outside of the sampling frame. 16 new works were added, bringing the ratio of female author to slightly less than 22 per cent. The second step includes a comparison of the monitor corpus and collections held at the Library of Congress and the Billy Ireland Cartoon Library at Ohio State University. By checking authors in these collections against a list of names that were assigned genders by the US Social Security Administration, we compare their gender make-ups and will potentially add to our corpus.

Gender and Interaction Types in Semi-Automatic Networks

Network analysis has steadily grown as an area of research since Franco Moretti's visualization of Shakespeare's *Hamlet* (Moretti). Scholars have focused on automatic extraction and statistical analysis of data from novels, plays, and intellectual networks (Elson, Dames & McKeown; van de Camp & van den Bosch). Recent efforts include computing main characters and operationalizing dynamic networks (Jannidis et al; Karsdorp & van den Bosch; Xanthos et al). While these networks answer some of Drucker's criticism, the approaches remain reductive. Limiting interactions to undifferentiated edges appears particularly unsatisfying for visual media, in which communication takes visibly different forms: characters may look at and touch each other, or appear together in a panel. Despite recent advances, computer vision has trouble recognizing non-perspectival drawings and applying OCR to handwritten comics fonts remains fraught with difficulty (Dunst et al; Rigaud 2013 & 2015). As the automatic extraction of network data is some way off for comics, we focus on networks that are semantically enriched via manual annotation to engage with the central questions posed by cultural studies. The following network (Figure 3) of Karasik and Mazzuchelli's *City of Glass* combines different types of interaction with gender assignations:

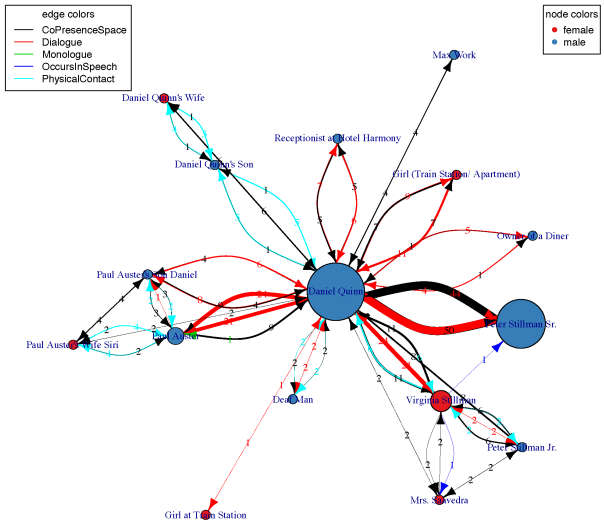

Figure 3: Interactions and gender assignations in City of Glass

These networks and the SSA name database allow us to study the relation between authors' gender and its fictional representation. Male characters are consistently more central to works by male authors. Notably, female characters show higher betweenness centrality in narratives written by women, as Figure 4 shows.

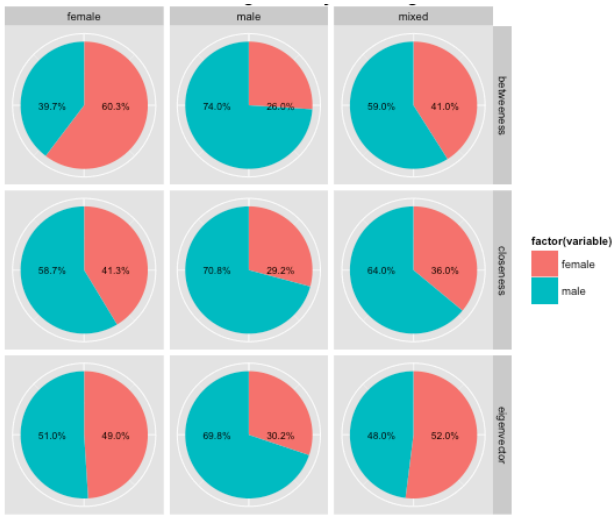

Figure 4: Centralities of character's genders by author's gender

Semi-Automated Bechdel-Test for GNML-Annotated Graphic Narratives

Efforts to automate the Bechdel Test have been limited to plays and film scripts and led to poor results (Lawrence; Agarwal et al). Three conditions need to be met to pass the test: 1. At least two female characters appear in the story. 2. There is at least one conversation between women. 3. The conversation is **not** about a man. Our XML-annotation language GNML, an extension of John Walsh's CBML, allows for automatically checking if graphic narratives fail

criteria 1 and 2, and aids evaluation of whether the remaining narratives pass criteria 3. We decided not to rely on error-prone full automation but to use semi-automatic processes that aid human annotators/analyzers in retrieving quality results. GNML annotations contain information on all character occurrences, the gender of a character, and their interactions. As discussed by Agarwal, even sophisticated machine learning approaches lead to unreliable results in deciding whether a conversation centers on a man. Therefore, for criteria 3, we simplify decision-making by providing the annotator with a ranked list of dialogues, based on the number of male names or male personal pronouns per conversation. Significantly, even if a conversation's focus on a male character could be identified automatically, the test would still be error-prone. Conversations may span several panels or pages and automatic separation of these conversations remains difficult.

Conclusions and Future Research

We present corpus metadata and semantically-enriched networks of a widely popular but understudied medium that is only beginning to attract attention by DH researchers. These methods are used to analyze gender and racial biases and suggest ways in which DH can appeal to scholars in cultural studies and the wider public. Future work includes expanding pilot studies to cover our entire corpus by integrating advances in OCR and computer vision and thus working towards fully-automatic extraction and analysis. In the meantime, our networks may function as conceptual models exploring how humanistic forms of complexity can be introduced into network analysis. Analyzing and addressing racial biases against minority authors presents hurdles of a different sort. A repeat of our survey for minority authors appears unproblematic but assigning racial identity to names or authors could be viewed as unethical and, given the international nature of our corpus, would have to consider the complex relationship of racial, national, and regional identities.

Bibliography

Agarwal, A., Zheng, J., Kamath, S., Balasubramanian, S., and Ann Dey, S. "Key Female Characters in Film Have More to Talk About Besides Men: Automating the Bechdel Test". *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*. 830-40.

Bechdel, A. (2014) "The Rule". http://dykestowatchoutfor.com/wp-content/uploads/2014/05/The-Rule-cleaned-up.jpg, accessed: 25. 10. 2016.

Drucker, J. (2016) "Graphical Approaches to the Digital Humanities". *A New Companion to Digital Humanities*. Ed. S. Schreibman, R. Siemens, & J. Unsworth. Oxford: Wiley. 238-50.

Dunst, A., Hartel, R., Hohenstein, S., and Laubrock, J. (2016) "Corpus Analyses of Multimodal Narrative: The Example of Graphic Narrative". *Conference Abstracts DH 2016*. 178-9.

Elson, D.; Dames, N. and McKeown, K.R. (2010). "Extracting Social Networks from Literary Fiction". *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. 138-47.

Karsdorp, F. and van den Bosch, A. (2016). "The Structure and Evolution of Story Networks". *Royal Society Open Science* 3: 1-15.

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". *Conference Abstracts DH 2016*. 578-82.

Lawrence, F. (2011) "SPARQLing Conversation: Automating the Bechdel-Wallace Test". Available at http://nht.ecs.soton.ac.uk/2011/papers/12-flawrence.pdf. (Accessed: 25 October 2016).

Moretti, F. (2011) "Network Theory, Plot Analysis". *Stanford Literary Lab Pamphlets* 2.

McPherson, T. (2014) "Designing for Difference." *Differences* 25:177-188.

Rigaud, C., Karatzas, D., van der Weijer, J., Burie, J.C., and Ogier, J.M. (2013). "Automatic Text Localization in Scanned Comic Books". *International Conference on Computer Vision Theory and Applications 2013*.

Rigaud, C., Guérin, C., Karatzas, D., Burie, J.C., and Ogier, J.M. (2015). "Knowledge-Driven Understanding of Images in Comics Books". *International Journal on Document Analysis and Recognition* 18: 199-221.

Van de Camp, M. & van den Bosch, A. "The Socialist Network". *Decision Support Systems* 53 (2012). 761-69.

Walsh, J. (2012) "Comic Book Markup Language: An Introduction and Rationale". *Digital Humanities Quarterly* 6.

Xanthos, A., Pante, I., Rochat, Y., Grandjean, M. (2016). "Visualizing the Dynamics of Character Networks". *Conference Abstracts DH 2016*. 417-19.