Highlighted Gaps: Toward Undogmatic Modeling of Literary Character Networks

Dan Vilenchik and Itay Marienberg-Milikowsky
Ben-Gurion University of the Negev, Israel

Abstract

We present a new computational framework for identifying “hot-spots” in a novel, in form of pairs of characters that the reader may want to pay closer attention to while reading the novel. The crux of our approach is to identify discrepancies between at least two network representations of the set of characters: appearance on stage, semantic proximity, and more.

1 Introduction

The construction and analysis of social networks in literary texts has been one of the most prominent applications of computational literary studies (Elson et al. 2010; Moretti 2011). Since novels and short stories represent a fictional social world at one level or another of complexity, this application seems particularly appropriate.

But like other literary phenomena examined in computational tools, character networks cannot be taken for granted; Automatic-extracted graphs are not just another way of reading, and their measurement cannot ignore conceptual problems (Moretti 2013). Jan Christoph Meister, who suggested the term “Undogmatic Reading” to describe an ideal approach for computational literary studies, wrote once that “A computational philology [...] cannot be concerned with driving out a person’s natural-language intelligence and their desire for ambiguity and obliging literary scholars to communicate in a restricted way with ones and zeros. Rather, its aim must be to make fruitful a fundamental tension: that between the human conceptualisation of text as a synthetic, meaningful communication phenomenon on the one hand, and the digital conceptualisation of text as an information phenomenon on the other” (Meister 2013, trans. by Flüh et al. 2021).

If the text is an information phenomenon, then, ideally, the researcher can simply extract data from it, data which can then be measured as needed. But if the text is a “meaningful communication phenomenon”, connecting people to one another on the basis of interpretation-dependent formulated discourse, then it is much more difficult to define the data to be extracted and measured, if at all. In the digital humanities, various proposals for balancing the two have so far been proposed.

2 Our Contribution

Computational methods, if used correctly, can offer a useful bridge between the territory and the map (alluding to Alfred Korzybski’s famous saying), by proposing systematic reductions (a map) of the reading experience (the territory) in a way that invites the reader to re-explore the territory and to re-invent a map. Our work manifests this idea by experimenting with character network construction and analysis as a test case. We propose an undogmatic operationalization of literary networks, based on three assumptions:

1. Each model is partial, so it is especially worth noting the relationship between it and other models, especially when the models *do not match* one another;
2. The most interesting phenomena *may be* precisely in the places of mismatch; and finally,
3. Identifying and analyzing interesting places is, and should be, a mission for human hermeneutics.

This idea is the concept that underlies TEASER, the product (and philosophy) of our ongoing joint project. TEASER, abbreviation of Text Evaluation and Analysis based on Serial Readings, now under development, is designed to support such an interpretive process: the system is serially fed with one after the other of different models of text, some automatic, some manual. It calculates the relationship between them, and produces a product that highlights points of discrepancy between them, and which serve, therefore, as a teaser for advanced human reading.

3 Results

In our work on certain novels (by one of the leading Israeli writer, Amos Oz, translated into English), we built several alternative networks for each novel: (1) a manual network based on a human reading of the novel, during which any communication between characters was marked as noteworthy connection; (2) an automated network based on the mentioning of different characters in the same paragraph; and (3) an automated network based on similar contextual-semantic relationships between characters, as these are expressed in the word2vec model. This pipeline combines close reading, NLP procedures, and statistics. Next, we overlayed the networks on top of one another. The results obtained were indeed, in some cases, teasing.

Figure 1 is a unified graph for networks in Oz' novel 'A Perfect Peace' (1985 [1982]). At first glance the figure is similar to other illustrations of its kind. But in fact, this is a multi-layered network designed as a heat map: it is based on a measured comparison between the two automatic models mentioned above.

This heat map - which cannot be analyzed here in depth - compares the relationship differences found by the two models: the redder the graph, the greater the gap between the semantic relationship that connects characters, as identified in the word2vec model, and the relationship based on their joint number of appearances. Or, in other words, these are characters who appear in a similar semantic field, but do not tend to get on stage together in the novel scenes. On the other hand, the more the graph tends towards blue-purple, the more the characters tend to appear together, while the semantic connection between them is low. The number that appears in the graph in the links that connect the characters indicates a measure of the strength of the relationship between them in the semantic model.

4 Discussion

Re-analysis of the text in light of the computational findings, in this case, is a task for the human reader. The algorithm weights data that the person cannot weigh, but the computational result acquires meaning only when it is examined in a sensitive reading of the text. This, in short, is what can be illuminated by our approach: The question of any sort of algorithmic reading, is not only what to measure, and how to measure and what does measurement means, but also, *what is encapsulated – hermeneutically – in the relationship between different measurements*, as calculated and represented numerically or visually.

The result is especially interesting when it reveals *gaps* between different measurements; It does not necessarily function as a naïve direct answer to a given question, but as a teaser, as food for thought. The undogmatic modeling approach described here, therefore, might contribute not only to the validation of the computational model for literary study, but also to the understanding of the special hermeneutic potential found in highlighting differences between models. It treats them as potential markers of literary points of interest, which are interesting because they are derived from an encounter between alternative perspectives – mathematically-oriented and literary-oriented.

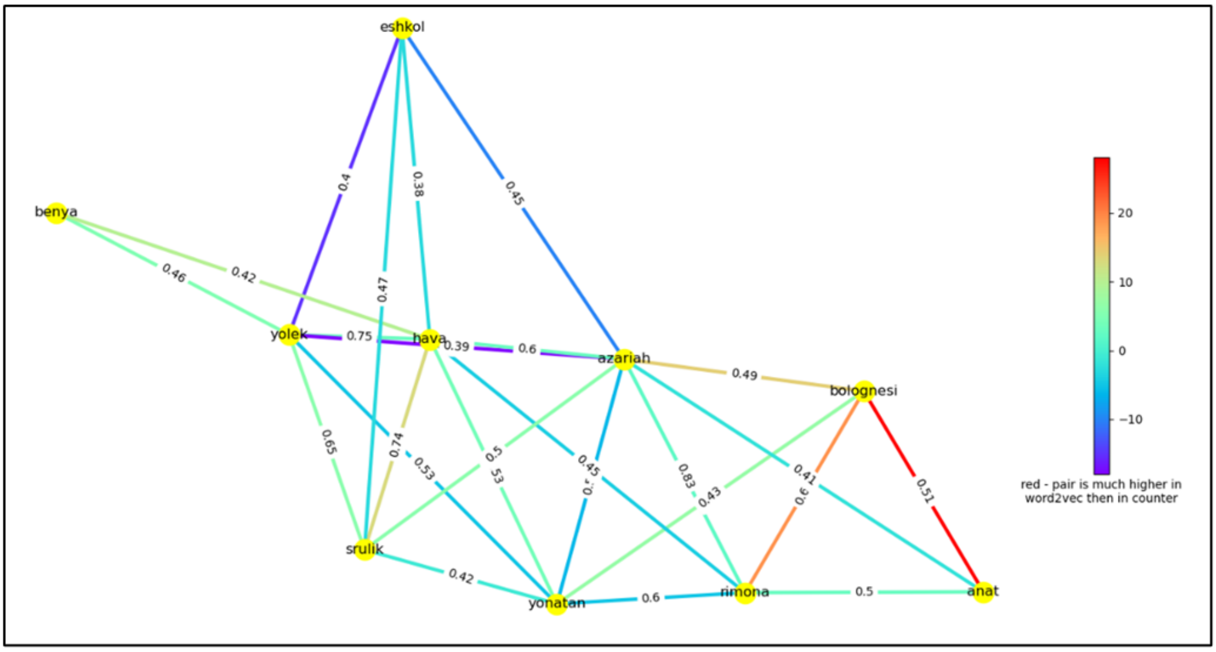
References

Elson, David et al., 'Extracting Social Networks from Literary Fiction', *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (2010), pp. 138–147.

Flüh, Marie et al. 'Introduction: Undogmatic Reading – from Narratology to Digital Humanities and Back'. In: Flüh, Marie, Jan Horstmann, Janina Jacke, Mareike Schumacher (eds.). *Toward Undogmatic Reading: Narratology, Digital Humanities and Beyond*, Hamburg: Hamburg University Press, 2021, pp.11-29.

Meister, Jan Christoph: 'Computerphilologie vs. Digital Text Studies. Von der pragmatischen zur methodologischen Perspektive auf die Digitalisierung der Literaturwissenschaften'. In: Christine Grond-Rigler und Wolfgang Straub (eds.). *Literatur und Digitalisierung*. Berlin and Boston 2013, pp. 267–296.

Moretti, Franco, 'Network Theory, Plot Analysis', Stanford Literary Lab Pamphlet no. 2 (May 2011).



Moretti, Franco: ‘Operationalizing: or, the Function of Measurement in Modern Literary Theory’, *Stanford Literary Lab Pamphlets*, No. 6 (2013).

Figure 1. An overlayed graph of two different network models of ‘A Perfect Peace’ by Amos Oz. Nodes are character names. Edge colors encode discrepancies between semantic proximity and “appearance on stage” proximity. The hotter the color the larger the gap in favor of semantic proximity.