Word2Vec-Based Literary Networks - Challenges and Opportunities

Marienberg-Millikowsky, Itay

itay.marienberg@gmail.com
Ben-Gurion University of the Negev, Israel

Vilenchick, Dan

vilenchi@bgu.ac.il
Ben-Gurion University of the Negev, Israel

The construction and analysis of social networks in literary texts has been one of the recent most prominent applications of computational literary studies (Elson et al. 2010; Moretti 2011). Since novels and short stories represent a fictional social world, this application seems particularly appropriate. But like other literary phenomena examined by 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; Meister 2013; Flüh et al. 2021). The common approach to model relationships between characters is by using relational graphs. In that graph, nodes represent a character, and an edge connecting two nodes signifies some relationship between them. The edges can be weighted or not. When applied to the literary text, the graph is usually derived from the co-appearance relation of characters. There are different ways to quantify the co-appearance; one common way is just to count in how many sentences/paragraphs the two characters appear together (the edge weight is that number); another way is to consider direct dialogues (this method already needs AI - or, in the case of a small corpus, manual annotation - to separate a simple co-appearance from direct dialogue).

In this work - to the best of our knowledge, for the first time - we take a completely different approach that does not rely on co-appearance, but rather on what can be called semantic context. Conceptually, we want to connect two characters by an edge if they share a semantic neighborhood. Namely, if they appear in similar meaning-based contexts, even (or especially) when these contexts are far apart in terms of their position in the linear sequence of the text. This approach is being used in the field of Natural Language Processing (NLP) to accomplish tasks like machine translation, topic modeling, hate speech detection, etc. Such a method for example may connect “milk” and “cookie” with an edge because they appear in similar contexts such as breakfast, or tea-time. It may also connect “happy” and “sad” as they are both emotions and therefore appear in similar contexts (“I was happy/sad to see you leave”). Thus the semantic contextual relationship approach goes beyond synonyms or trivial similarities, and may find more complex layers of association.

To make things concrete, we shall use the word2vec technology. Word2vec is a group of models that are used to produce word embeddings. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words. Word2vec takes as its input a large corpus of text and produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space. Word vectors are positioned in the vector space in such a way that words that share common contexts in the corpus are located close to one another (cosine similarity). In our work we are mainly interested in the embedding of the names of the characters. There are some technical obstacles to overcome (for example, how to recognize nicknames); but these are mere technicalities that are already handled pretty well by existing software packages like BookNLP.

We are going to train a word2vec model on a given book.

There are two options that one can consider: take a pre-trained network and fine-tune it on the book, or train from scratch. Each choice has its merits, and both can be tested. Once we obtain an embedding, we compute the cosine similarity between every pair of characters. For each pair, this will give some score in the range [-1,1]. A score close to 0 stands for a weak relationship. A score close to 1 (or -1, but negative scores are rarely obtained, and if so, are typically close to 0) signifies a very strong relationship. We discreteize the range into three labels: “strong”, ”medium”, ”weak”. Weak edges will not appear, medium will be colored blue and strong - red. Now that we have this weighted graph, we can ask the same questions as before, just noting that the interpretation may be different than in the co-appearance model. For example, the character with the highest degree (or highest red-degree) will be a central character in the plot, but in the sense that it plays a similar semantic contextual role with many other characters; in that capacity, it may be a prominent driving force of events. A red connection triangle, for example, built on the basis of a semantic model, may differ greatly from connection triangles built on the basis of co-appearance models: even if the characters within it rarely communicate with each other, they may perform similar metaphorical, symbolic, or plot-wise functions (think of two lovers of the same character; three characters representing diverse views in a certain ideological system; and the like).

One computational caveat that needs to be addressed is that co-appearance naturally affects the embedding. If we take the extreme case of two characters that always appear together, then naturally they will also have the same context. There are some measures to control this effect if called for. One way is to change the window of the word2vec; smaller windows will rarely include two characters. Another way is to ignore windows that contain more than one character. From a literary point of view, the potential of such an approach is promising, as it enables the examination of hidden parallel relations, which are among the most important organizing principles of prose. This approach, however, is also full of challenges, including the fact that semantic linguistic similarity does not always correlate with literary similarity.

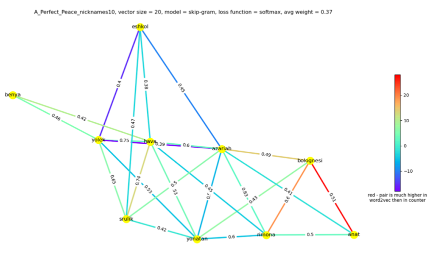

Figure no.1. A word2vec network model of ‘A Perfect Peace’ by Amos Oz. Cosine similarity is marked on edges. Color codes were normalized by co-appearance to control for that effect.

Bibliography

Elson, David (2010): “Extracting Social Networks from Literary Fiction”, in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* , 138–147.

Flüh, Marie (2021): “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, 11-29.

Meister, Jan Christoph (2013): “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, 267–296.

Moretti, Franco (2011): *Network Theory, Plot Analysis* , Stanford Literary Lab Pamphlet 2.

Moretti, Franco (2013): *Operationalizing: or, the Function of Measurement in Modern Literary Theory* . Stanford Literary Lab Pamphlets 6.