Refining Character Relationships in a Narrative using Embeddings of Interactions

Guillaume Guex
Departement of Language and Information Sciences
University of Lausanne
gguex@unil.ch

Abstract

The construction of character networks from narratives can be decomposed into three steps: (1) the identification of characters; (2) the detection of interactions; (3) the extraction of the graph (for an extensive review, see (Labatut and Bost, 2019)). When the studied narrative consists in textual resources (e.g. novels, plays, movie scripts), the detection of interactions (step 2) between two characters generally consists in identifying *segments of text* where protagonists are involved in some type of interaction. These segments of text can be, among others, narrative units containing character co-occurrences (Elsner, 2012; Rochat and Kaplan, 2014; Min and Park, 2019), conversations (Nalisnick and Baird, 2013; Kwon and Shim, 2017), or direct actions (Sudhahar and Cristianini, 2013; Srivastava et al., 2016). In step 3, these interactions are usually counted to obtain an undirected/directed, static/dynamic weighted character network where edges represent the number of interactions. However, apart from a few examples where edges are signed according to a sentiment score (Nalisnick and Baird, 2013; Trovati and Brady, 2014; Min and Park, 2019), almost no information is extracted from the *corpus formed by the segments of text defining interactions*. Interpreting relationships in a resulting network can be therefore difficult, as edges aggregate blindly various types of interactions.

The current submission proposes to refine the study of relationships using a *bag-of-words* approach and *embeddings of interactions* to construct various indices on interactions. This approach allows the construction of multiple character networks representing different types of relationships. More specifically, we extract from the studied text a contingency table $\mathbf{N}$, with n rows and v columns, where rows represent the n detected interactions, columns the v words in vocabulary, and each cell contains n_{ij}, the number of occurrences of word j in interaction i. Moreover, each interaction i is endowed with a set $\mathcal{C}_i$ containing all characters taking part in this interaction (this set can also be structured, in the case of oriented interactions). These character sets permit the reconstruction of relationships from interactions, and thus the graph, by aggregating over all interactions where a pair of character are involved, either over the whole work (static graph) or by, e.g., chapters (dynamic graph). This contingency table $\mathbf{N}$ and the character sets $\mathcal{C}_i$ form the initial building block for our analyses. From there, different approaches can be made.

A first approach consists in a *contrastive* analysis of interactions in the narrative, using no external resources. It allows to study of interactions as intended by the author, regardless of his writing style or the type of the work. A natural tool to perform such an analysis on a contingency table is *Correspondence Analysis* (CA) (Lebart et al., 2019), which allows to embedded interactions and words (rows and columns in $\mathbf{N}$) in a factorial map. The interaction scores (positive or negative) along a chosen factor can reveal a particular type of contrast, which can be interpreted by word positions regarding the same axis, as seen in Figure 1.

The second studied approach consists in embedding interactions using *pretrained word-embedding* such as Word2Vec (Mikolov et al., 2013), Glove (Pennington et al., 2014), or FastText (Bojanowski et al., 2017) trained on very large corpora. Mapping interactions in those spaces consists in finding a weighting scheme (frequency, TF-IDF) for words and representing interactions as centroids of the word vectors composing them (Agarwal et al., 2019). This approach contrasts from the first as a particular narrative is mapped in an *absolute referential*, thus enabling comparisons between various works and authors. The interactions and relationships can then be examined along different direc-

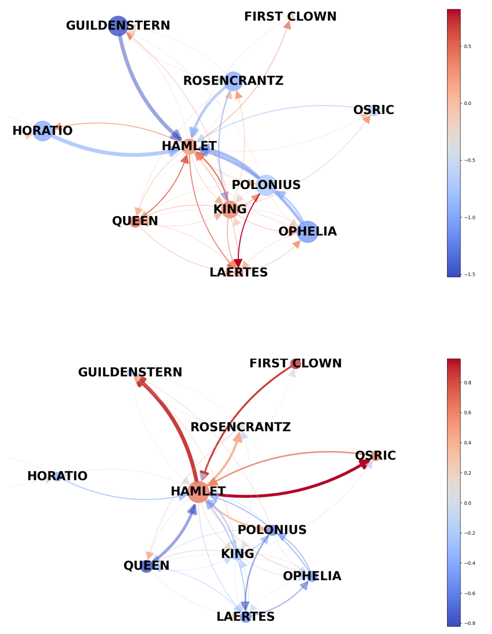

Figure 1: Dialog networks between main protagonists in *Hamlet* using CA. The top figure represents mean relationship scores (colors) along the first factor, where most positive words are {"thy", "thou"} and most negative {"lord", "good"}, which thus shows social hierarchy between characters. The bottom figure represents mean relationship scores along the second factor, where most positive words are {"welcome", "sir"} and most negative {"hast", "fear"}, which can be interpreted as the confidence-fear spectrum.

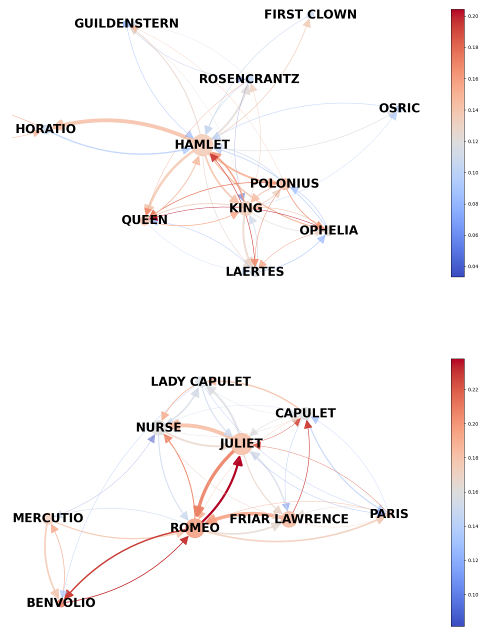

Figure 2: Dialog networks between main protagonists in *Hamlet* (top) and *Romeo and Juliet* (bottom), using Word2Vec word vectors pretrained on Wikipedia (Yamada et al., 2020). Scores (colors) express the cosine similarity between relationships and the word "love" minus the cosine similarity between relationships and the word "hatred", thus defining a score along the love-hatred spectrum. We can see that this spectrum clearly segments the types of relationships in Romeo and Juliet, but is less discriminant on Hamlet.

tions, which can be defined by the user using word vectors (Figure 2).

While promising, several methodological questions arise from these approaches. Which weighting scheme must be used? Should stopwords be removed? A cutoff on minimum or maximum occurring words should be performed? How indices must be computed in the embedding space? Moreover, the step involving aggregation of interactions, which gives final relationship scores, can be performed before or after the embedding step, which gives different results. These approaches could also be combined with a clustering of the words, or a *Topic Modeling* approach, thus permitting indices along axes representing these topics. The current submission proposes to compare approaches, explore leads, and answer these questions using multiple case studies. The long-term goal is to find a strong framework for mining textual information on character interactions that could be used as a new distant reading tool, allowing digital humanity researchers to refine their analyses of character relationships in textual narratives.

References

Lucky Agarwal, Kartik Thakral, Gaurav Bhatt, and Ankush Mittal. 2019. Authorship Clustering using TF-IDF weighted Word-Embeddings. In *Proceedings of the 11th Forum for Information Retrieval Evaluation*, pages 24–29, Kolkata India. ACM.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Micha Elsner. 2012. Character-based kernels for novelistic plot structure. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 634–644, Avignon, France. Association for Computational Linguistics.

Ho-Chang Kwon and Kwang-Hyun Shim. 2017. An

Improved Method of Character Network Analysis for Literary Criticism: A Case Study of Hamlet. *International Journal of Contents*, 13(3):43–48.

Vincent Labatut and Xavier Bost. 2019. Extraction and Analysis of Fictional Character Networks: A Survey. *ACM Computing Surveys*, 52(5):89:1–89:40.

Ludovic Lebart, Bénédicte Pincemin, and Céline Poudat. 2019. *Analyse des données textuelles*. Number 11 in Mesure et évaluation. Presses de l'Université du Québec, Québec.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]*. ArXiv: 1301.3781.

Semi Min and Juyong Park. 2019. Modeling narrative structure and dynamics with networks, sentiment analysis, and topic modeling. *PLOS ONE*, 14(12):e0226025.

Eric T. Nalisnick and Henry S. Baird. 2013. Character-to-Character Sentiment Analysis in Shakespeare's Plays. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 479–483, Sofia, Bulgaria. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Yannick Rochat and Frédéric Kaplan. 2014. Analyse des réseaux de personnages dans Les Confessions de Jean-Jacques Rousseau. *Les Cahiers du numérique*, 10(3):109–133. Place: Cachan Publisher: Lavoisier.

Shashank Srivastava, Snigdha Chaturvedi, and Tom Mitchell. 2016. Inferring Interpersonal Relations in Narrative Summaries. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1). Section: Technical Papers: NLP and Machine Learning.

Saatviga Sudhahar and Nello Cristianini. 2013. Automated analysis of narrative content for digital humanities. *International Journal of Advanced Computer Science*, 3(9):440–447. Publisher: Citeseer.

Marcello Trovati and James Brady. 2014. Towards an Automated Approach to Extract and Compare Fictional Networks: An Initial Evaluation. In *2014 25th International Workshop on Database and Expert Systems Applications*, pages 246–250, Munich, Germany. IEEE.

Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, and Yuji Matsumoto. 2020. Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 23–30, Online. Association for Computational Linguistics.