Representing and visualizing the narrative of texts as complex networks

H. F. de Arruda[1,2], D. R. Amancio[2], L. da F. Costa[3], F. N. Silva[4]

[1] University of Zaragoza, Spain
[2] Institute of Mathematics and Computer Sciences, University of São Paulo, Brazil
[3] São Carlos Institute of Physics, University of São Paulo, Brazil
[4] Indiana University Network Science Institute, US

Not many studies explored the representation of narratives in the literature.

We devised a method to construct a narrative recurrent network (NRN) from books.

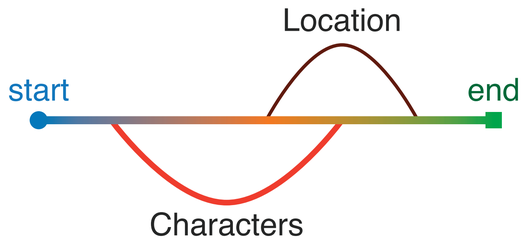
Paragraphs of the text are connected in a chain and according to a textual similarity.

Network visualization is employed to create a graphical representation of the NRNs by bending the chains of nodes according to text similarity.

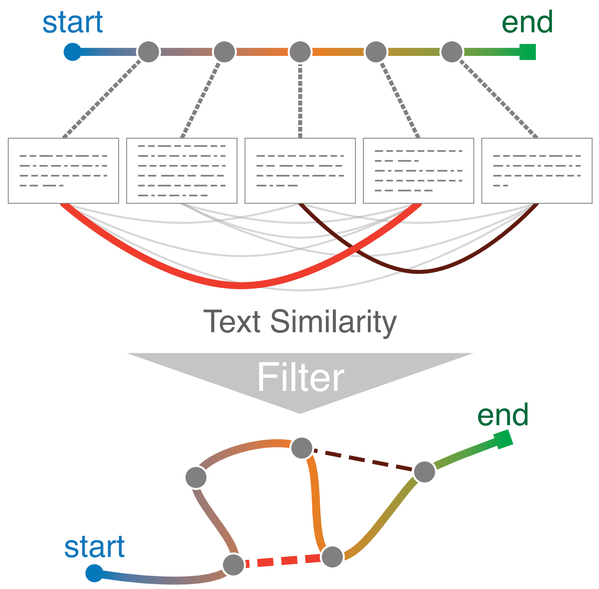
We found that the proposed representation was able to capture structure that is similar across different languages.

Narrative as a network

A narrative can be understood as a linear chain of events that may share characters, locations, and other entities as it unfolds.



We represent a narrative of a book in terms of a recurrence network (NRN) [1]. Paragraphs are nodes connected according to adjacency.



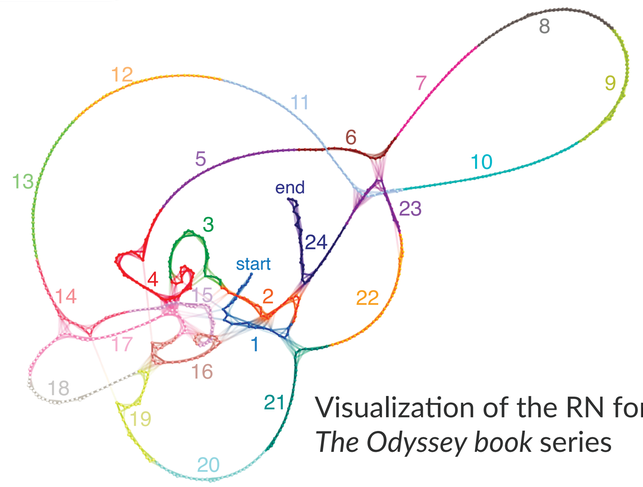
The remaining links are drawn from the text similarity based on:

- The cosine similarity of the TF-IDF vectors derived from the words of each paragraph;
- or the number of shared nouns and related syntactical dependencies obtained by a language model [2].

Visualizing the narrative network

We employed a force-directed layout algorithm to construct graphical representations for the narrative recurrence networks.

On the right side, we show the network obtained for The Odyssey book series (attributed to Homer). Nodes colors and respective numbers indicate the books they belong to.

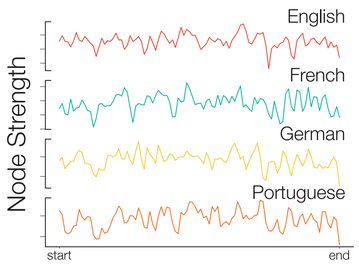


Visualization of the RN for *The Odyssey book* series

Comparison among different languages

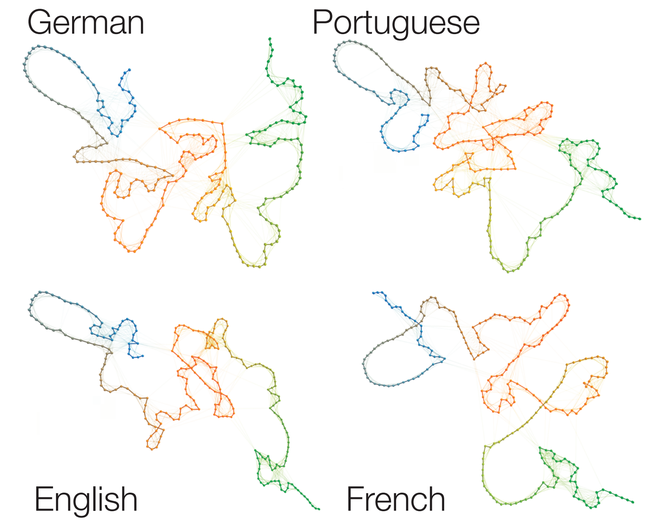
To evaluate the potential of the NRN. We applied the method on the book Alice's Adventures in Wonderland (by Lewis Carroll) and considered translations for 4 distinct languages.

The overall structure of the NRNs appears to be language agnostic according to the visualizations.



Correlations of node strength

	EN	FR	DE	PT
EN		0.43	0.59	0.59
FR	0.43		0.46	0.56
DE	0.59	0.46		0.61
PT	0.59	0.56	0.61	



Comparison of RNs for 4 translations of the same book

This aspect is also confirmed by looking at the curves of node strength. We also found a significant correlation across the considered translations of the book.

Future developments of this study should focus on using the NRNs for classification tasks, e.g., detection of genre, publication period, or authorship.

[1] de Arruda, H. F.; Silva, F. N.; Marinho, V. Q.; Amancio, D. R.; & Costa, L. da F. (2018). Representation of texts as complex networks: a mesoscopic approach. Journal of Complex Networks, 6(1), 125-144.

[2] Chen, D., & Manning, C. D. (2014, October). A fast and accurate dependency parser using neural networks. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 740-750).