Shape of Stories Parameters
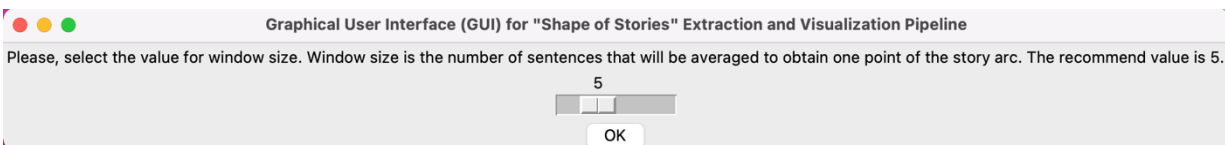
Parameters of the Shape of Stories Tool

Once the selected Shape of Story algorithm starts running, it will ask you for parameter preferences in order to have maximized results depending on your corpus characteristics (Size & Type). Each one of the 3 parameters will give you recommended values computed by the NLP Suite that best suits your corpus. However, you are free to adjust parameters to your convenience if you wish to experiment.
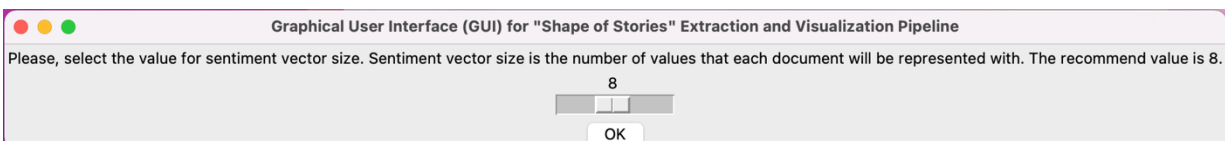
Window Size

First, it will ask you for the "Window Size" which is the number of sentences the algorithm will use to output a Sentiment Score of a point in the story arc. This has a huge impact on precision but at the same time it can drastically increase running time when too large. A lower Window Size means more points generated for each text but also less sentences incorporated in a single point, meaning that each point will be highly sensitive to the sentiment score of a single sentence.

Graphical User Interface (GUI) for "Shape of Stories" Extraction and Visualization Pipeline

Please, select the value for window size. Window size is the number of sentences that will be averaged to obtain one point of the story arc. The recommend value is 5.

5

OK

Sentiment Vector Size

Second, it will ask you for the "Sentiment Vector Size" which is the number points or "length" of the vector that is computed for each text. A large vector size can lead to overfitting and bad results.

Graphical User Interface (GUI) for "Shape of Stories" Extraction and Visualization Pipeline

Please, select the value for sentiment vector size. Sentiment vector size is the number of values that each document will be represented with. The recommend value is 8.
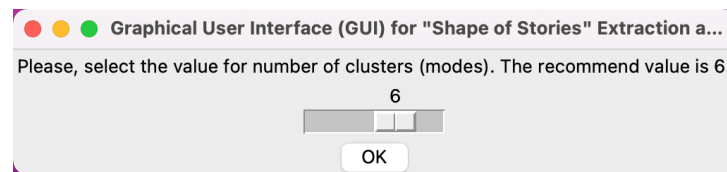
8

OK

Number of Clusters

Finally, it will ask for the number of clusters created by the algorithm. The more clusters you get the more precise your clusters. Too much precision will prevent you from highlighting common

trends in your corpus while not enough will only give you superficial results. If you would like to experiment with the number of clusters k, please use the Best Topic Estimation Function.

To learn more about the Best Topic Estimation function, read the TIPS_NLP_Best Topic Estimation.

References

Burrows, J.F. 1987. "Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style." *Literary and Linguistic Computing*, Vol. 2, No. 2, pp. 61-70.

Franzosi, Roberto. 2004. *From Words to Numbers*. Cambridge, UK: Cambridge University Press.

Franzosi, Roberto. 2010. *Quantitative Narrative Analysis*. Thousand Oaks, CA: Sage.

Reagan, Andrew J., Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. 2016. "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes". *EPJ Data Science*, Vol. 5, No. 31, pp. 1-12.

Vonnegut, Kurt. https://www.youtube.com/watch?v=oP3c1h8v2ZQ

Vonnegut, Kurt. https://www.youtube.com/watch?v=GOGru_4z1Vc

Vonnegut, Kurt. 2005. "Here is a Lesson in Creative Writing." In: pp. 23-28, Kurt Vonnegut, *A Man Without a Country. Edited by Daniel Simon.* New York: Seven Stories Press.