Enlightenment Inflluencers: Networks of Text Reuse in 18th-century France

Roe, Glenn
glenn.roe@sorbonne-universite.fr
ModERN Project, Sorbonne University, France

Fedchenko, Valentina
valentina.fedchenko@sorbonne-universite.fr
ModERN Project, Sorbonne University, France

Nicolosi, Dario
dario.nicolosi.92@gmail.com
ModERN Project, Sorbonne University, France

Introduction

The 18th century is an important moment in the long history of authorship and authorship practices whose legacy we are only now beginning to appreciate (Cronk and Stalnaker 2012). In contradistinction to the myth of the solitary genius (Stillinger 1991), 18th-century writers adopted a wide variety of authorial strategies in the construction of their works: anonymity and pseudonymity; collaboration and compilation; citation, non-citation and mis-citation; allusion, paraphrase and pastiche, etc. (Coleman 2010). The ERC-funded ModERN project aims to bring the complexities and contemporary relevance of these textual practices to light through the combination of data-rich computational techniques with traditional critical methods to examine the digital archive of the Enlightenment period. [1] Specifically, this paper explores the use of new large-scale text reuse detection and network analysis to identify intertextual 'influencers' in a large heterogeneous collection of 18th-century French texts.

Methodology

Overall, our aim is to trace the circulation of 18th-century passages as a network, while, at the same time, identifying network 'profiles' that challenge traditional author-centric and/or canonical literary histories. [2] To do so, we constructed a preliminary corpus of 734 works drawn from the ARTFL-Frantext corpus of French literature and published between 1686 and 1799. [3] We then used the Text-PAIR sequence alignment system to run a pair-wise comparison of each text and identify shared sequences of n-grams, in this case using trigrams (n=3) as our default lexical representation. [4] The resulting alignment database contained 4,671 pairs of likely similar passages (see Fig. 1) with their metadata, which was then extracted from Text-PAIR as a static JSON file for further analysis.

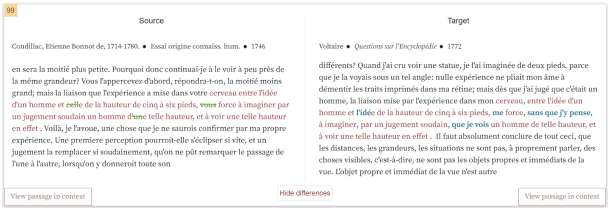

Figure 1. Identified alignment of Condillac and Voltaire via Text-PAIR.

We then constructed several semi-automatic filters in an effort to reduce noisy or unwanted passages (formulaic expressions, near-duplicates, paratexts, etc.) from our alignments, and generated both filtered and unfiltered passage networks for further analysis. We turned to network analysis as a way of understanding complex systems, drawing inspiration from 'actor-network theory' (ANT) as a method for identifying and tracing networks of influence or 'meditation' (see Latour 2007 and Latour et al. 2012). We tested several network analysis and visualisation tools, including the Python libraries Networkx, [5] Pyvis, [6] and Gravis, [7] as well as the popular platform Gephi, [8] before ultimately settling on Networkx based on its capacity for easy data manipulation, graph calculations, and visualisation. Two networks were generated: one with authors as nodes, and the other based on individual works, with edges of 1,292 and 1,507 passage-pairs respectively.

Preliminary results

Our two text reuse networks were analysed using standard force-directed graph measures: degree, in-degree, out-degree, weighted node/edge betweenness, closeness and PageRank centralities. [9] Following Text-PAIR logic (which is based on matches from chronologically previous 'sources' to later 'targets'), the first graphs were directed from nodes with quoted works or authors (sources) towards nodes containing similar passages (targets). This type of model tends to over-emphasise the importance of the target nodes in our network, i.e., those quoting rather than being quoted. We thus decided, somewhat counter-intuitively, to reverse the directionality of our graphs, rendering nodes that generate the most reuses more relevant than those that reuse others, meaning that the most quoted authors/texts should have the largest number of incoming edges. Structured in this way, the above graph centrality measures helped us define three types of author/text functions:

1. *Authorities* (high in-degree, high PageRank, low betweenness)

   Authors/texts that are highly cited/reused but that do not cite/reuse others. This is largely a function of chronology, as many late 17th- and early 18th-century authors/texts are

included in this category, e.g., Boileau (Fig. 2), Bossuet, or Fontenelle.

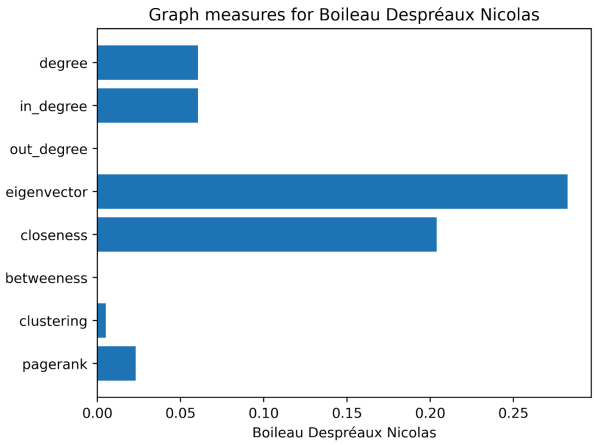

Figure 2. Graph measures for Boileau as an example of an *Authority*.

2. *Mediators* (high betweenness, closeness, and local clustering coefficient)

   Authors/texts that comment, compare, or critique other sources, e.g., literary or cultural 'critics' such as Jean-François Marmontel, Louis-Sébastien Mercier (Fig. 3) and Philippe-Louis Gérard.

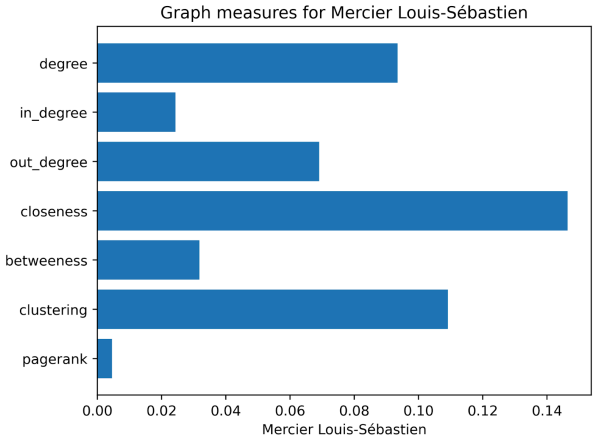

Figure 3. Graph measures for Mercier as an example of a *Mediator*.

3. *Observers* (low graph measures with possible average out-degree/degree)

   Author/texts that are rarely cited/reused and that tend to belong to the periphery of the network, but that nonetheless cite/reuse others extensively. Here again chronology plays a role, with later authors such as Augustin Barruel and the Marquis de Sade belonging to this category along with more classic 'observers' such as the Comte de Caylus (Fig. 4).

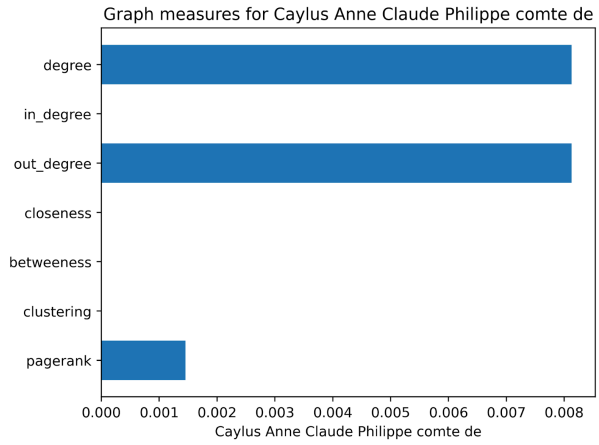

Figure 4. Graph measures for Caylus as an example of an *Observer*.

Outliers

Along with these more general network profiles, we identified several outliers in the network that warranted further investigation:

1. *Influencers*: i.e., authors/texts with high values for all centrality and clustering measures. These nodes share the qualities and functions of both an *authority* and a *mediator*, serving as a reference for the most part of the network members and as a bridge between different literary communities. Voltaire is the most obvious *influencer* in our network (Fig. 5).

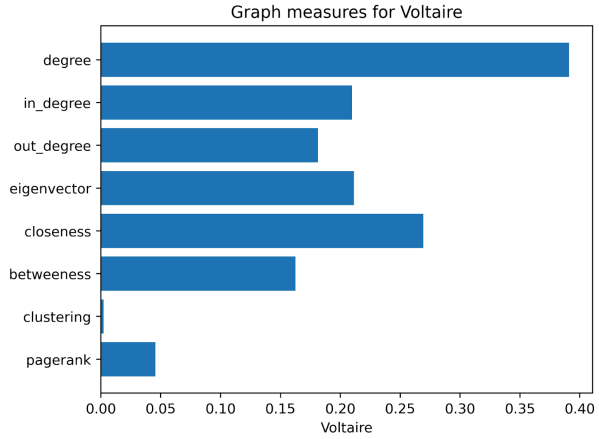

Figure 5. Graph measures for Voltaire as an example of an *Influencer*.

2. *Noise:* the combination of unproportionally high and low values can help detect problems in network and identify 'noise': i.e., near-duplicate texts and clusters of false-positive matches such as formulaic text that occurs across many documents but that bears no genetic relationship (paratextual elements, publishing boilerplates, *privilege du roi*, etc.). Statistical methods for outlier detection can thus be combined with qualitative contextual historical analysis to identify and eliminate

this ‘noise’ from our text reuse networks. For example, high measures for Antoine Galland led us to examine our corpus and identify a near-duplicate of his 1704 *Mille et une nuits* published anonymously in 1716 under the title *Vieilles nouvelles rajeunies*, among other paratextual noise (Figs. 6 and 7).

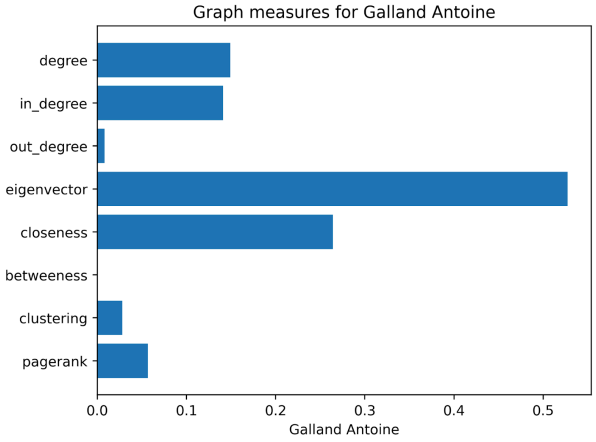

Figure 6. High PageRank and very low betweenness centrality for Galland in an unfiltered graph.

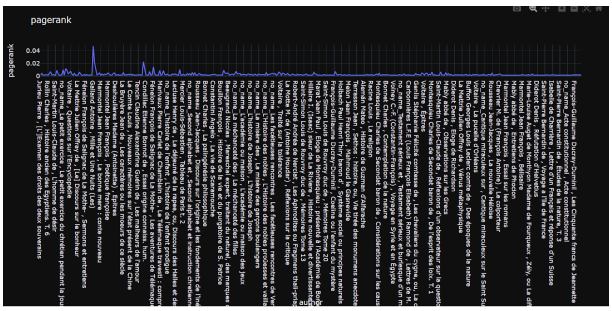

Figure 7. Galland’s work “Mille et une nuits” (duplicate) as an outlier in an unfiltered graph (per work).

Future work

As our project moves forward, the above observations will need to be projected on larger-scale text reuse networks, based on data that contains OCR errors, differing metadata schemes, and other types of ‘noise’. As we have seen, network analysis enables us to identify and reduce this noise and to classify an increased number of network members based on network profiles. Scalable networks will also be introduced using clustering techniques to extract smaller networks. This will allow us to model specific literary communities, and, by focussing on the relationships of cluster centres, to study the mechanics of information transfer within our networks.

Notes

1. See https://cordis.europa.eu/project/id/101043369.
2. Our notion of ‘network profiles’ is borrowed from (Ahnert and Ahnert 2015).
3. Our thanks to the University of Chicago’s ARTFL Project for providing us access to this dataset. See https://artfl-project.uchicago.edu/content/artfl-frantext. For the purposes of DH2023, we removed all works of theatre and correspondences, which tended to generate an inordinate amount of false-positive alignments.
4. On Text-PAIR, see https://github.com/ARTFL-Project/text-pair, and, more specifically, (Olsen et al. 2011; Roe 2012). Our choice of matching parameters was based on past experience using Text-PAIR, with flexible trigrams representing the best trade-off between signal and noise.
5. https://networkx.org/.
6. https://pyvis.readthedocs.io/en/latest/.
7. https://robert-haas.github.io/gravis-docs/.
8. https://gephi.org/.
9. For an excellent synthesis of network analysis techniques for humanities research, see Ahnert et al. 2020.

Bibliography

Ahnert, Ruth / **Ahnert, Sebastian E.** (2015). “Protestant Letter Networks in the Reign of Mary I: A Quantitative Approach”, in *ELH* 82, 1: 1–34. DOI:10.1353/elh.2015.0000.

Ahnert, Ruth / **Ahnert, Sebastian** / **Coleman, Catherine N.** / **Weingart, Scott** (2020). *The Network Turn: Changing perspectives in the Humanities*. Cambridge: Cambridge University Press.

Coleman, Patrick (2010). *Anger, Gratitude, and the Enlightenment Writer*. Oxford: Oxford University Press.

Cronk, Nicholas / **Stalnaker, Joanna** (2012). “Eighteenth Century Authorship”, special issue of *Romanic Review* 103, 3-4.

Latour, Bruno (2007). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, Bruno / **Jensen, Pablo** / **Venturini, Tommaso** / **Grauwin, Sébastien** / **Boullier, Dominique** (2012), “The whole is always smaller than its parts”, in *The British Journal of Sociology* 63: 590-615.

Olsen, Mark / **Horton, Russell** / **Roe, Glenn** (2011). “Something Borrowed: Sequence Alignment and the Identification of Similar Passages in Large Text Collections”, in *Digital Studies / Le Champ numérique* 2, 1.

Roe, Glenn (2012). “Intertextuality and Influence in the Age of Enlightenment: Sequence Alignment Applications for Humanities Research”, in *Digital Humanities 2012*: 345-47.

Stillinger, Jack (1991). *Multiple Authorship and the Myth of Solitary Genius*. Oxford: Oxford University Press.