



HAL
open science

'Double blind' graph data analysis: a pedagogical experiment to discuss the intersubjectivity of network interpretation

Mathieu Jacomy, Martin Grandjean

► **To cite this version:**

Mathieu Jacomy, Martin Grandjean. 'Double blind' graph data analysis: a pedagogical experiment to discuss the intersubjectivity of network interpretation. Digital Humanities, 2022, Tokyo, Japan. 10.5281/zenodo.6844790 . halshs-03753812

HAL Id: halshs-03753812

<https://shs.hal.science/halshs-03753812>


Submitted on 18 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

'Double blind' graph data analysis: a pedagogical experiment to discuss the intersubjectivity of network interpretation

Mathieu Jacomy, TANT Lab, Aalborg University, Denmark / [@jacomy](#)
Martin Grandjean, History Department, University of Lausanne, Switzerland / [@grandjeanmartin](#)

 Conference paper

Jacomy M. & Grandjean M. (2022). "[Double blind graph data analysis: a pedagogical experiment to discuss the intersubjectivity of network interpretation](#)". *Digital Humanities 2022*, Tokyo, Japan.

DOI [10.5281/zenodo.6844790](https://doi.org/10.5281/zenodo.6844790)

Introduction

On this page we propose an experimental process of 'double blind' comparison which consists of giving the same dataset to two network analysts and to confront their methodological choices, results and interpretations. This 'double blind' approach (Jacomy and Grandjean, 2021) makes it possible to make the intersubjectivity of the process visible and provide empirical material to discuss how the eyes of specialists from different disciplines influence the interpretation of the data. We believe that producing and making explicit such interpretive pathways can help researchers, designers and students understand the diversity of possible approaches.

The subjectivity of network interpretation in the humanities

If network analysis is a method now very widely diffused in the field of digital humanities (Ahnert et al., 2020), often because the metaphor and lexical field of 'network' are very effective in describing the objects of study of the humanities, there is no definitive interpretation procedure. Indeed, describing a network entails the recontextualization of graph-theoretical elements (e.g., the "betweenness centrality" metric) into the language of the discipline (e.g., "knowledge brokers" or simply "bridges") through a methodological "translation" (Grandjean and Jacomy, 2019). This hermeneutic process is subjective by nature, especially in the humanities where visual network analysis is often privileged, which may seem paradoxical considering that the method is supposed to objectify historical or literary sources (this is more generally true for all digital methods). Yet as Popper argued, descriptive statements necessarily draw their validity from "intersubjective agreement" (Freeman, 1973). How much agreement does network analysis offer?

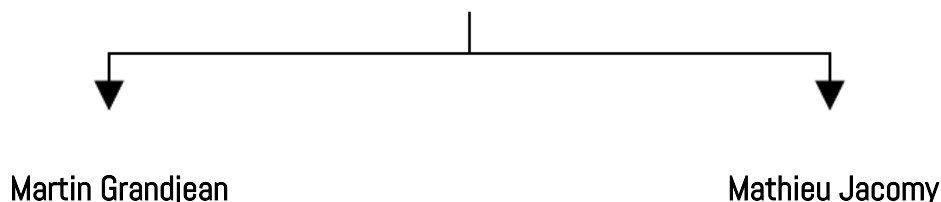
Experimental process

To ensure comparability, the phase of individual experimentation is framed by strict constraints in terms of time and tools. A network dataset is prepared upstream and duplicated: the two participants get a time budget of 2 hours to explore the data with Gephi (Bastian et al. 2009) and produce an annotated network visualization of their choice, possibly including other tables and visualizations, as well as a list of hypotheses on the structure of the graph. This phase is filmed and screencasted. Then the two people discover each other's work and compare their interpretations.

The result of these parallel interpretations is presented as two paths (below) and can be compared by means of an online document integrating the data set, interactive and/or composite visualizations, descriptive text and video sequences making it possible to follow the reasoning. In the end, the data set itself is only a pretext for an educational experiment, the aim of which is as much to confront us with the subjectivity of the interpretation as to show the interest of such an exercise of replication.

Data

The data set is a network of Wikipedia pages that link to each other. It has been harvested from the "[European cuisine](#)" category on the English Wikipedia. All its pages have been extracted as well as the pages listed in its subcategories using notebooks developed by Mathieu and available on his "[Mapping Controversies](#)" website. All the links between the pages of this list have been retrieved and all the pages of the main giant component that are linked at least twice have been kept. The network contains 5.800 nodes and 300.000 edges.



First visualization

First visualization

Network Analysis 1/4 - Introduction and Fir...

Euro cuisine network 1/4 - First visualization



Network exploration

Network exploration

Network Analysis 2/4 - Exploration

Euro cuisine network 2/4 - Network explora...



Final visualization

Final visualization

Network Analysis 3/4 - Final Visualization

Euro cuisine network 3/4 - Final visualization



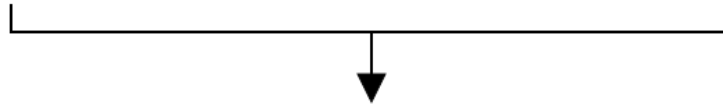
Summary

Summary

Network Analysis 4/4 - Summary

Euro cuisine network 4/4 - Summary





Conclusion

As the videos show, the two analysts used the same building blocks: layout algorithm, centrality metrics, filtering, coloring, interaction with the network in Gephi... But they assembled those building blocks in different ways: in a different order, and sometimes for different purposes. Martin's committed to the "national cuisines" angle by focusing on those he identified early on as the most represented or interesting (Turkish, English, and Italian), and using metrics to compare them. Mathieu's approach was more descriptive, focusing on delineating the clusters and comparing their qualitative content. Basically, both questioned the coherence of the clusters but from a different point of view, one by isolating the national parameter and observing how it correlates with the visible groups, the other by trying to understand the content of the groups. They also had different findings, but only because they looked at different things.

The two analyzes sometimes concluded the same thing, notably those points:

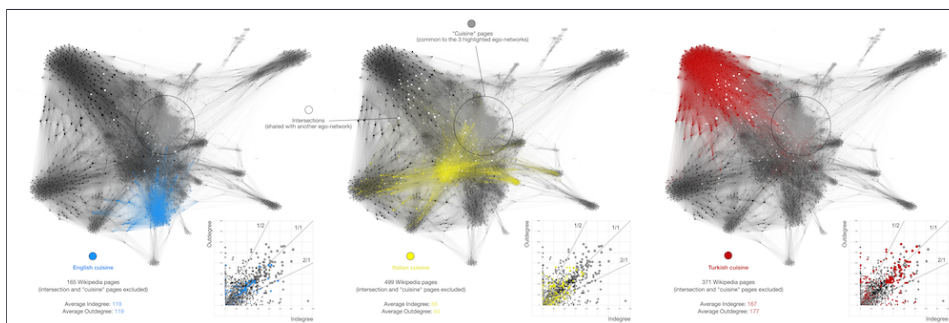
- A minority of national cuisines (Italian, Turkish and English, but also Greek, French, Portuguese and Spanish, whose pages are very connected but whose clusters are less visible) are much more represented than the vast majority of others, especially Eastern Europe and less populated countries.
- There is a cluster specifically about the national cuisines of various countries, suggesting that the national dimension is the main structuring factor and bridge between ingredients and dishes.
- The Turkish cuisine cluster is peripheral and very dense while the Italian cluster is topographically central while not very dense.

It is also worth highlighting that the analyzes had no contradictions. They did not differ by having a disagreement, but by covering partially different features of the network.

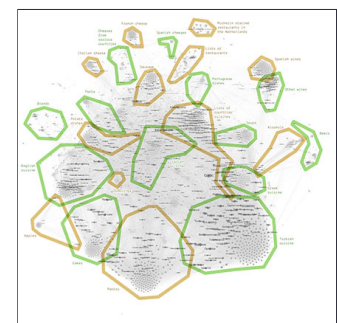
Let us summarize this experiment in three short takeaways about intersubjectivity of network analysis:

- First, the subjectivity of the analyst determined which aspect of the network were interpreted. This was a direct consequence of having different approaches. In this case, one sought to test a research hypothesis and compare it to the structure of the graph while the other sought to give the best possible account of what the network was showing.
- Second, the subjectivity of the analyst shaped how their findings were framed and phrased. Indeed, each approach comes with its own methodological commitments.
- Third, and most importantly, the subjectivity of the analyst did not shape the content of the findings. We can see this to the fact that there was a significant overlap in the findings, and no contradictions. The overlap is remarkable given the very short time given for the analysis (2h max). However, the very short time constraint is also a factor that limits the possibility of producing very fine analyzes that are more likely to contradict each other. Subjectivity shaped the direction of the inquiry, but not the content of the observations.

We do not have any claim to generality, as we only reflect on this particular case. A different and more ambitious research design would be necessary, combining very different analysts, networks and software. This experiment is a first step, and it offers a direction for further works. It suggests that the main way subjectivity may lead different analysts to produce different analyzes is through their curation of the salient features of the network. It also suggests that when the same features get analyzed, subjectivity does not cause contradictions. In other words, it suggests that network analysis is robust enough to support intersubjective statements.



Martin Grandjean's three facets of the same network (English, Italian and Turkish cuisines).



Mathieu Jacomy's final annotated network visualization.

References

Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. *The Network Turn: Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press, 2020.

Bastian Mathieu, Sébastien Heymann and Mathieu Jacomy. 'Gephi: an open source software for exploring and manipulating networks'. International AAAI Conference on Weblogs and Social Media, 2009.

Jacomy, Mathieu, and Martin Grandjean. 'Double Dating Data: Intellectual Cooperation in the League of Nations'. Aarhus, 2021.

Freeman, Eugene. 'Objectivity as "Intersubjective Agreement"'. *The Monist*, 1973, vol. 57, no 2, p. 168-175.

Grandjean, Martin, and Mathieu Jacomy. 'Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics'. In *Digital Humanities 2019*. Utrecht, 2019.