

How to Orient Oneself in Thinking in the Digital Era

A mixed-methods ethnography of researchers' navigational practices on Gallica

Simon Dumas Primbault

Laboratory for the History of Science and Technology · EPFL · Lausanne, Switzerland

simon.dumasprimbault@epfl.ch · orcid.org/0000-0002-0012-0550 · cv.archives-ouvertes.fr/simon-dumas-primbault

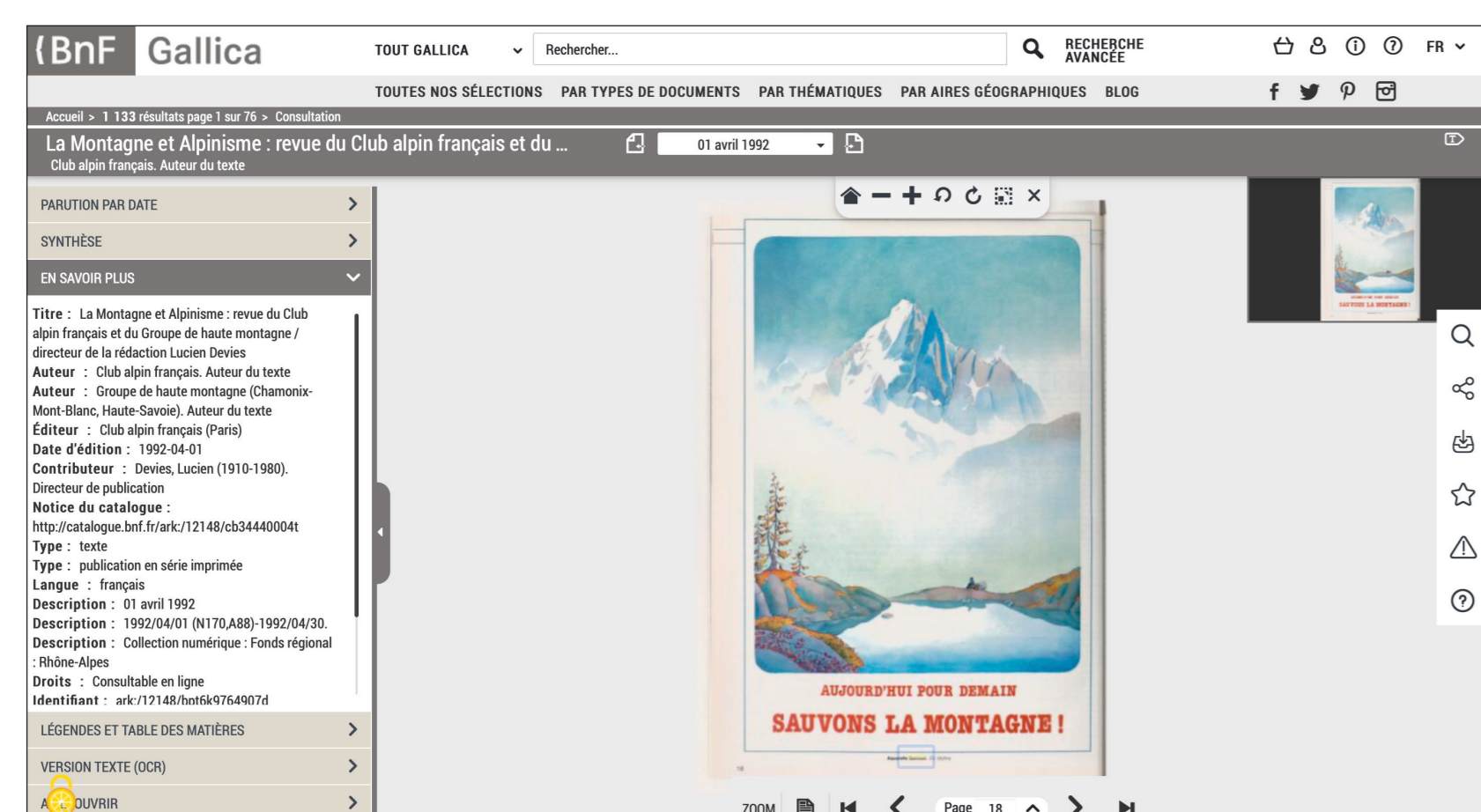


Part of this research project was led in collaboration with **Bayrem Kaabachi**, MA student in Communication Systems at EPFL.

THE BIRTH of digital libraries has radically transformed research activities and online navigation has become a fundamental information practice. By documenting researchers' reading paths, this study proposes to consider the exploratory aspect of the scientific research activity as a journey. Taking Gallica as a case study, this research situated at the crossroads of ethnography, STS, and digital humanities strives to shed light on how researchers "orient themselves" within a digital corpus understood as a space to be surveyed. By insisting on the path—as *methodos* in the Aristotelian sense as opposed to informational extractivism—the mass of data becomes a condition of possibility for navigation, for orientation—rather than being presented as a fantasy of exhaustiveness or considered as a source of information overload.

Case Study

Founded in 1997, Gallica is the online platform of the Bibliothèque nationale de France (BnF)—the second largest library in the world—dedicated to providing free access to digitized historical sources and documents preserved at BnF. Thought from the beginning as a "virtual library" it was endowed with a public patrimonial mission, tailored to make available documents to the widest audience.



- Over 9 million documents in open access
- 8 languages
- Books, manuscripts, newspapers, engravings, photographs, musical scores...
- No built-in recommendation system
- Dewey Decimal Classification
- Online reader
- Widely used by genealogists, amateur and professional historians, and more generally in the humanities and social sciences

Data and Method

Methodology for an ethnography of the digital

Semi-directed interviews
1-2 hours
7 interviews

1. Basic sociological profile
2. Within one platform
3. Between platforms
4. On the computer
5. Outside the computer
6. Observing three tasks

Qualitative approach
7 participants, frequent users of Gallica with connection times longer than 10 min, research as their principal or professional activity. Interviews were recorded and transcribed.

Pipeline for a digital ethnography

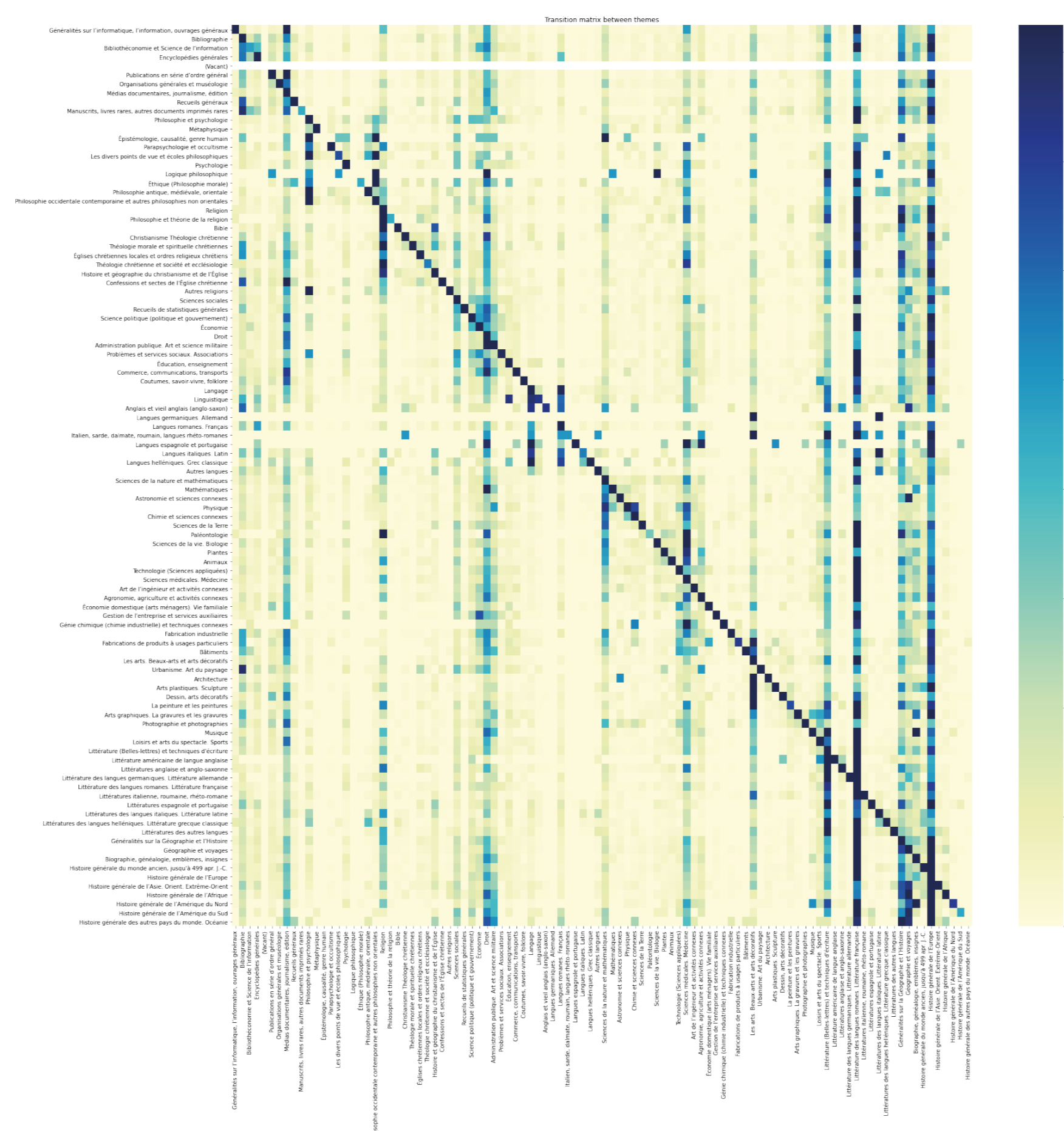
1. Harvest server logs
2. Model and extract navigational paths
3. Enrich navigational paths by querying BnF Archival Resources Keys (ARKs)
4. Model paths as Markov chains of Dewey classes and generate transition matrix from one class to another
5. Generate a topological space of disciplines
6. Cluster paths by morphological features

Quantitative approach
Gallica server logs from April to June 2016. All code was written in Python. Word2vec embedding was used at step 5 to compute pairwise distances between disciplines. Giotto-tda library was used at step 6.

Results

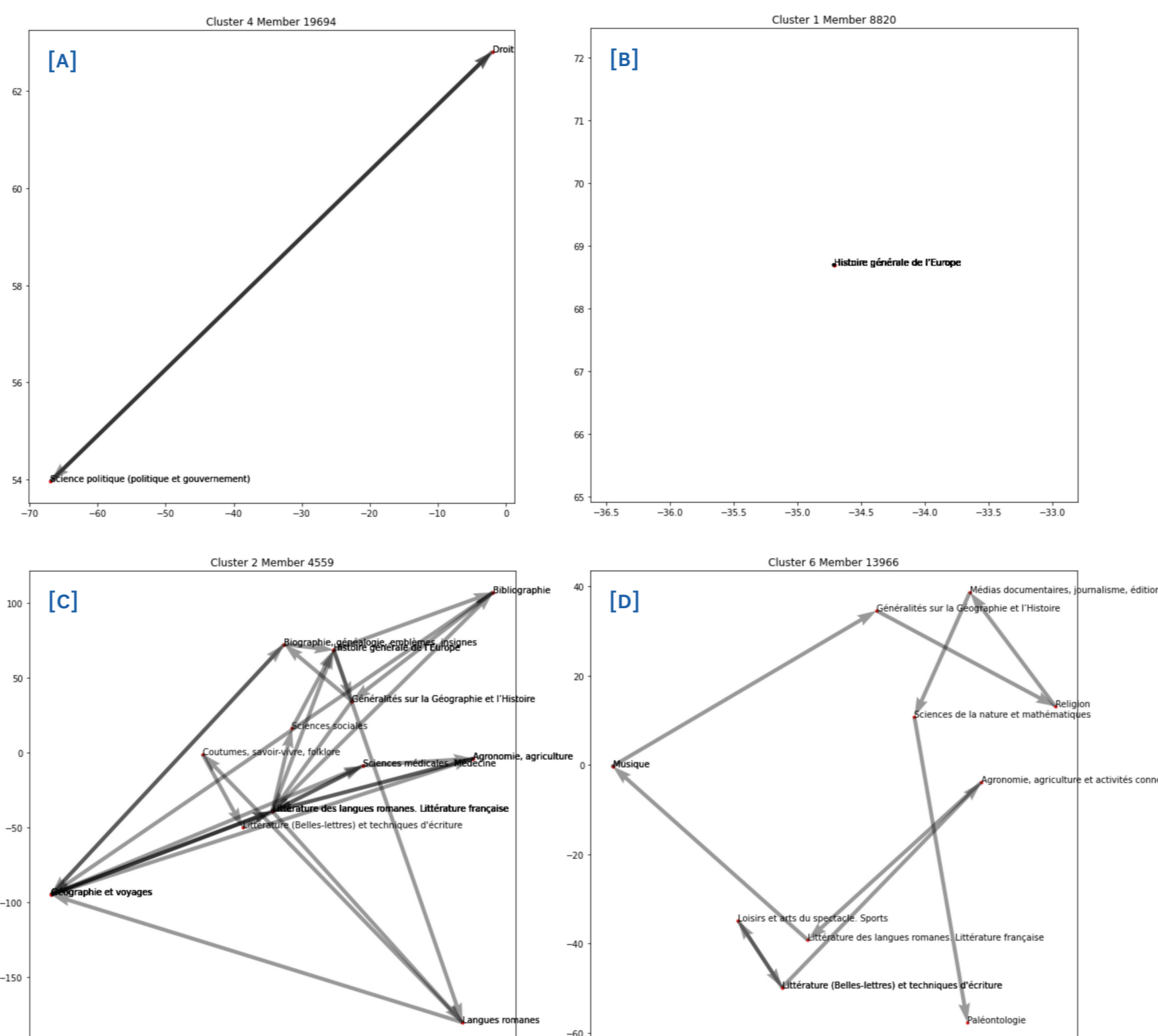
- All participants always work "in parallel" (i.e., multiple tabs and platforms, notetaking on paper...)
- None had been specifically trained to digital tools
- All participants had difficulties objectifying their practices:
 - Tension between idiosyncrasy and generality
 - Tension between organized and disorganized practices
 - Contradictions between "this is exactly how I do it" and "I could not tell"
 - Contradictions between discourse and practice
- Most readers tend to stay within one single Dewey class or few subclasses [see transition matrix below]
- Some classes act as "pivotal literature," facilitating the transition from one field to another (e.g., "General History," collections, encyclopedias...) [see transition matrix below]

Transition matrix from one Dewey class to another in navigational paths modeled as Markov chains



- The concept of "navigational paths" is relevant to describe research practices:
 - Rich oral testimonies and use of spatial, temporal, physical, entertainment, or discovery metaphors to try and objectify practices that do not have a priori common representations
 - Paths obtained by quantitative modeling were tested with participants and deemed realistic
- Paths can be clustered into "regimes of navigation" corresponding to different phases of research and different temporalities identified by the participants:
 - "Directed" or "punctual" research to extract a specific piece of information, "very precise" and "short sessions" [A]
 - "Constitution" or "consultation of a corpus" within one specific class [B]
 - Crossed, compared, or interdisciplinary readings centred around one or multiple classes [C]
 - Exploratory research, "wandering" or "fluttering about" on their "freetime" [D] etc.

2D projections of selected navigational paths drawn within the inferred topological space



Conclusion

Although traditionally associated with hermeneutic disciplines like the humanities, the ever more piecemeal access to information brought about by digital media has gradually made navigation a transversal practice. Navigation does not amount to the unconstrained roaming of infinite spaces but must rather be understood as a patient work of contextual surveying working with the framework imposed by their media. As such, it is a diachronic, albeit not simply sequential, practice aimed at finding contextual relationships between documents based on their types, topics, authors, disciplines, or methods. Navigation therefore encompasses a wide variety of information practices spanning searching (or information-mining), collecting (information-farming), skimming (information-strolling), garnering (information-stumbling, serendipity, or accident), and forays into neighboring fields (information-poaching). Thus, navigation is an exercise in orientation within knowledge: moving from one document to another, researchers learn to draw for themselves an information landscape by grouping works encountered on their path into disciplines, periods, or approaches.

Selected References

Beaudouin, V., and Denis, J. (2014). "Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique" (research report), BnF / Télécom ParisTech. hal.archives-ouvertes.fr/halshs-01078530

Beaudouin, V., Garron, I., and Rollet, N. (2016). "Je pars d'un sujet, je rebondis sur un autre". Pratiques et usages des publics de Gallica" (research report) Télécom ParisTech / BnF / LabEx Obvill. hal.archives-ouvertes.fr/hal-01709238

Béquet, G. (2014). *Trois bibliothèques européennes face à Google. Aux origines de la bibliothèque numérique (1990-2010)*. Paris, École des Chartes.

Deveaud, R. (2018). "Modalités d'accès au savoir ouvert sur les plateformes d'OpenEdition", Aix-Marseille Université.

Galoin-Mélénec, B., and Zliti, S. (dirs.) (2013). *Traces numériques. De la production à l'interprétation*, Paris, CNRS Éditions.

Nouvellet, A., Beaudouin, V., D'Alché-Buc, F., Prieur, C., and Roueff, F. (2017). "Analyse des traces d'usage de Gallica. Une étude à partir des logs de connexions au site Gallica" (research report), Télécom ParisTech / BnF. hal.archives-ouvertes.fr/hal-01709264

Souchier, E., Jeanneret, Y., and Le Marc, J. (dirs.) (2003). *Lire, écrire, réécrire. Objets, signes et pratiques des médias informatisés*, Paris, Bibliothèque publique d'information. doi.org/10.4000/books.bibpompidou.394

Venturini, T., Cardon, D., and Cointet, J.-P. (2014). "Méthodes digitales. Approches quali/quantitatives des données numériques", *Réseaux*, Vol. 6, No. 188, pp. 9-21. doi.org/10.3917/res.188.0009

Data availability

In compliance with GDPR and the recommendations of EPFL ethical committee (HREC No. 056-2020), the interviews, their transcripts, and the server logs cannot be made public.

All the code produced for this project is open source and available under a GNU license on the following GitHub repository: github.com/LHST-EPFL/TDA-Gallica.

Acknowledgments

This research project, originally entitled "Digital Trail Blazing: Understanding and remaking intellectual mobility on online research platforms," was supported by a CROSS 2020 grant awarded by EPFL's College of Humanities and the University of Lausanne.

A follow-up project entitled "Naviguer le savoir à l'ère numérique. Une étude des pratiques de navigation sur Gallica" is funded by the 2022 Mark Pigott grant for the digital humanities, awarded by the Bibliothèque nationale de France.

Rights and permissions

This poster is licensed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License.