

Mapping as a Contemporary Instrument for Orientation in Conferences

Chloe Ye Eun Moon
Columbia University
ym2761@columbia.edu

Dario Rodighiero
Massachusetts Institute of Technology
rodighie@mit.edu

Abstract

English. This article presents a case study analyzing submissions from the Digital Humanities 2019 conference by visualizing a network of authors situated according to their shared lexicon. This new form of summarizing a conference is an effective way to grasp the whole conference at once. The hope is that this method of visualization will not be employed merely as a retroactive way to reflect on past events, but rather as an instrument to prepare the visit and orientate the attendees during the conference.

Italiano. Questo articolo presenta una mappa lessicale creata per studiare la distanza linguistica tra gli autori che hanno partecipato alla conferenza Digital Humanities 2019. I testi degli articoli sono raggruppati per autore e analizzati secondo la loro frequenza, per poi essere tradotti in connessioni. Il risultato è una mappa topologica della conferenza, composta degli stessi autori situati in uno spazio di linguistico. Questa nuova forma di visualizzazione è un efficace strumento per riassumere una comunità scientifica in un'immagine. L'augurio è che metodi come questo non siano soltanto degli strumenti usati a posteriori, ma piuttosto messi a disposizione in anticipo per preparare la propria visita e permettere ai partecipanti di orientarsi durante una conferenza.

1 Introduction

Maps are formidable instruments for abstracting territory and travels. Centuries of cartographic mapping marked the evolution of world history, and today's technological innovation has opened a new paradigm of mapmaking (Dodge et al., 2011; Lévy, 2016). Maps are no longer static; instead, they can change dynamically given external inputs. For instance, users can zoom into the map to obtain detailed information, and they can choose to filter selected information. Due to the big data technology, maps are now able to represent larger and larger datasets (Kitchin, 2014). Moreover, maps now have a new level of abstraction not only for the territory but also for individuals; unlike before, visualizing social relationships became a common practice since Jacob Moreno's work in social relationships (1934).

This article presents a case study of human mapping, specifically of scholars and their scientific productions, focusing on the Digital Humanities Conference 2019 that took place in Utrecht, the Netherlands in July 2019. Such a case study develops the previous work of Rodighiero (Rodighiero, 2015, 2018; Rodighiero et al., 2018), demonstrating how a map can be a powerful instrument of reduction not only for territory but also for individuals. The cartography of DH2019¹ is intended to be a generic tool for mapping scientific communities, in the form of an open-source project². The article will present the result of such cartography by discussing the following four sections: 1) *Documentality* as a way human activity is regulated through textual inscriptions 2) *Lexical Analysis* as the way documents can be automatically analyzed under human supervision 3) *Graphic Design* as a visual translation that makes a conference

¹The cartography of DH2019 is accessible at <https://rodighiero.github.io/DH2019/>

²The open-source project is available at <https://github.com/rodighiero/DH2019/>

attended by a thousand authors wholly graspable 4) *Reading* as a form of user interaction through which the map reader acquires information from the visual media.

2 Documentality

A series of social rules govern humans behaviors (Kaplan, 2012), and scholars are no exception. Their activities follow specific rules when they attend a conference; submitting articles, waiting for reviews, attending the conference, and presenting their work are the steps they are required to complete in order to be a part of an academic field. Such behaviors are auto-regulated by the scientific community itself, which gives a structure to the research domain.

Articles play a significant role in this process as they are an extension of their authors. They convey different types of information, such as collaboration, affiliation, scientific interests, and the writing proficiency of the authors. Therefore, articles are valuable as they portray the authors' values and social relevance. We hypothesize that documents embody textual information, which can be used to measure the proximity between scholars most effectively. As individuals express themselves through their language, authors can be described through their writing. Terms are not a barrier, nor they are private. Everyone can choose and use preferred words and speech styles according to their taste, and the choice is profoundly affected by social and cultural environments, such as their education level and location. Nonetheless, there is complete freedom in the selection of language, and this freedom goes beyond any collaboration or citation. Just like how Pierre Bourdieu used personal interviews to classify individuals (Blasius and Schmitz, 2014; Romele and Rodighiero, 2019), the goal here is to map scholars using scientific articles.

3 Lexical Analysis

Natural Language Processing (NLP) is a branch of artificial intelligence that aims to process and analyze large amounts of text (Manning and Schuetze, 1999). Since humans' natural language has no structured rules, computers can understand, NLP is especially challenging; therefore, the techniques in NLP are significant as they derive meaningful insights from texts written in such a language. In order to profile the authors who attended the Digital Humanities 2019 conference, we performed a lexical analysis to map the distance from one author to another.

First, the XML data of DH2019 were cleaned for a more accurate analysis utilizing JavaScript programming language and the Cheerio library³. From each article are extracted authors, title, and text body. Since multiple authors can co-author a paper, the text body is grouped by authors, allowing multiple occurrences of the same publication if a paper is co-authored. Each text is then tokenized using a lexical analyzer provided by the Natural library⁴ for NLP. Successively, tokens are singularized and filtered by a list of stopwords in various languages, including Brazilian, English, French, German, Italian, and Portuguese.

The arrays of tokens associated with authors are then computed via the Term Frequency - Inverse Document Frequency algorithm, also known as TF-IDF (Luhn, 1957; Sparck Jones, 1972). TF-IDF extracts the most relevant terms for each author by counting the frequency of each term with respect to the inverse frequency of the entire collection of words. The list of terms for each other is then shortened to the fifteen most relevant terms in order to simplify the visual computation. Table 1 shows a sample concerning the scholar Frédéric Kaplan.

³Cheerio is an open-source library available at <https://github.com/cheeriojs/cheerio/>

⁴Natural is an open-source library available on GitHub at <https://github.com/NaturalNode/>

Token	Value
Parcel	156.29240966203554
Amsterdam	147.92161777735498
Street	97.3716554233416
City	78.98538009996346
Cadastre	68.97376382900369
Rue	64.52753719852456
Neighbourhood	62.45323622544596
Urban	61.733372168535126
Century	52.240930125777
Bottin	51.62202975881964
Transcription	51.58659792981266
Geometrical	44.74034261035416
Extraction	43.895897928544414
Geographical	41.91500209834842
Geometry	39.821449253041536
Wine	38.71652231911473
Activity	37.66779857289015
Dialect	37.53334200758336
Cinema	34.18896461357028
Time	33.67020614879688

Table 1: An excerpt from the JSON file that describes the profile of Prof. Frédéric Kaplan. Among the metadata are his name, the number of articles, and a list of fifteen tokens weighted through the TF-IDF algorithm. As his research is mainly focused on the European Time Machine, which is focused on the computational analysis of ancient maps, the result can be considered adequate.

4 Graphic Design

Data analysis is followed by the creation of a network, in which each author forms a node, and the shared tokens are transformed into weighted edges. The resulting visual rendering does not recall a classic network visualization, such as Gephi’s (Bastian et al., 2009), but rather a cartographic projection. It is a hybrid form that combines the characteristics of networks and maps.

Authors are placed on the map using the Simulation function⁵ from the d3.js library (Bostock et al., 2011). Then, between each pair of authors, is displayed the most relevant token whose size corresponds to the TF-IDF value; a high TF-IDF value corresponds to a high degree of relevance in the whole collection. The elevation contours (Monmonier, 1991) displayed at the end of the simulation due to computation limits make the density of documents visible; as a result, an author who authored many articles is placed at a peak. The result is an elevation map that shows the most relevant tokens, such as languages, music, newspapers, and films. When zoomed in, the map can be enlarged to display the details, and the user will notice the tokens changing. The reason is that the tokens are selected according to the zoom level; when the map is utterly visible, the user can see the most relevant tokens with high-frequency values, and by zooming he will see the most generic ones. This choice makes the view more specific and less generic, while a zoom-in allows viewing the complete gradient of tokens (See Figures 1, 2, and 3).

⁵Simulation runs to place the nodes in a proper way, more information is available at <https://github.com/d3/d3-force/>

5 Reading

In cartography, the reader is the individual who interacts with the map (Dodge et al., 2011; Lévy, 2016). Readers interpret the cartographic visualization differently depending on their knowledge and culture. In front of the DH2019 visualization, the reader identifies a configuration of scholars who attended the conference with an article. When authors share the same text because they co-authored a paper, they appear to be close on the map, and the most relevant token between them appears (Figure 1). When there is a continuity of tokens, it indicates an area of particular interest (Figure 2). When the density of individuals is visible, there is a high chance that it represents a collective work (Figure 3).

The map offers a novel point of view on the conference, which is a different approach from the proceedings or the website. It makes all the authors and their work graspable at the same time, while also allowing the readers to navigate through individual authors.

When a reader starts to explore a specific part of the map by zooming in, the interaction becomes personal and unique. Therefore, readers play an active role while putting an interpretation on the map; their pathway influences what they see. Furthermore, if a reader who participated in the conference recognizes her identity in the map, the reading validates the reader's representation by evaluating the correctness of the map and the neighborhood where the reader is placed (Rodighiero and Cellard, 2019).

6 Conclusion

Language is not only a means to convey one's ideas, but also to express interest and background. If a conference forms a scientific community by attendance, the articles presented at the conference shape the specific language of such a community with a delicate balance of different voices (Von Glasersfeld, 1992).

Lexical analysis of terminologies is an effective means to study the community. Thanks to the current technology and visualization techniques, we are now able to create a dynamic, interactive map of the community, which is dense and rich in information. From this new form of data visualization, readers can interpret the lexical proximity of all the authors at a glance, both the distance and placement.

Now the question is, why don't we use the map as an instrument during the conference? It would undoubtedly be a much more contextually rich and visually intriguing way of understanding the conference, instead of merely using statistics to summarize the event.

Acknowledgement

We want to thank Kurt Fendt for his constant support and supervision, and the MIT Literature section for hosting us, in particular Shankar Raman, Diana Henderson, and Alicia Mackin. Thanks also to Jeffrey Schnapp and the laboratory members of Harvard MetaLab.

Acknowledgement also goes to Stephan Risi for developing the search function and Philippe Rivière for his priceless collaboration, which was fundamental for recent projects. A special mention to Daniele Guido, an inseparable friend and colleague whose design capacities are stimulus for doing better; the grapefruit color palette is his merit.

This article is part of the grant Early Postdoc.Mobility P2ELP1_181930 *Worldwide Map of Research* funded by the *Swiss National Science Foundation*.



Figure 2: This image illustrates a specific area of the map in which the term ‘dariah’ is recurrent. The term reassembles the people working within the Dariah community. By zooming, the terms change according to the scale; the more the reader zooms in, the more generic the terms are.



Figure 3: Co-authoring is an easily recognizable phenomenon, especially in areas with low density. At the center, it is visible a group of scholars that share the term ‘interdisciplinarity.’ Terms, like in this case, can be used to spot small communities within the conference.

References

- Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International ICWSM Conference* .
- Jörg Blasius and Andreas Schmitz. 2014. Empirical Construction of Bourdieu's Social Space. In *Visualization and Verbalization of Data*, CRC Press, pages 205–222.
- Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17(12):2301–2309. <https://doi.org/10.1109/TVCG.2011.185>
- Martin Dodge, Rob Kitchin, and Chris Perkins, editors. 2011. *The Map Reader: Theories of Mapping Practice and Cartographic Representation*. John Wiley & Sons, Ltd, Chichester, UK. <https://doi.org/10.1002/9780470979587>
- Frédéric Kaplan. 2012. How Books Will Become Machines. In *Lire demain : des manuscrits antiques à l'ère digitale*, PPUR, pages 27–44.
- Rob Kitchin. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications.
- Jacques Lévy, editor. 2016. *A Cartographic Turn*. EPFL Press, Lausanne.
- Hans Peter Luhn. 1957. A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development* 1(4):309–317. <https://doi.org/10.1147/rd.14.0309>
- Christopher D. Manning and Hinrich Schuetze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- Mark Monmonier. 1991. *How to Lie with Maps*. University of Chicago Press.
- Jacob L Moreno. 1934. *Who Shall Survive?. A New Approach to the Problem of Human Interrelations. Nervous and Mental Disease Publishing Co., Washington, DC.*
- Dario Rodighiero. 2015. Representing the Digital Humanities Community: Unveiling the Social Network Visualization of an International Conference. *Parsons Journal of Information Mapping* VII(2). <https://doi.org/10.5281/zenodo.3464433>
- Dario Rodighiero. 2018. Printing Walkable Visualizations. In *Transimage Conference 2018*. University of Edinburgh, Edinburgh, pages 58–73. <https://doi.org/10.6084/m9.figshare.6104693>
- Dario Rodighiero and Loup Cellard. 2019. Self-Recognition in Data Visualization. *EspacesTemps.net Electronic Journal of Humanities and Social Sciences* <https://doi.org/10.26151/espacestemp.net-wztp-cc46>
- Dario Rodighiero, Frédéric Kaplan, and Boris Beau. 2018. Mapping Affinities in Academic Organizations. *Frontiers in Research Metrics and Analytics* 3(4). <https://doi.org/10.3389/frma.2018.00004>
- Alberto Romele and Dario Rodighiero. 2019. Digital Habitus, or Personalization without Personality (Forthcoming).
- Karen Sparck Jones. 1972. A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation* 28(1):11–21. <https://doi.org/10.1108/eb026526>
- Ernst Von Glasersfeld. 1992. Why I Consider Myself a Cybernetician. *Cybernetics and Human Knowing* 1(1):21–25