Visualizing longitudinal data. Rooted cosmopolitans in the Low Countries, 1850-1914

Thomas D'haeninck, **Nico Randeraad and Christophe Verbruggen**
Affiliations: Maastricht University, Ghent University
Adresses: Maastricht, Grote Gracht 90-92; Ghent, Sint-Pietersnieuwstraat 35
E-mail adresses: n.randeraad@maastrichtuniversity.nl, christophe.verbruggen@ugent.be, thomas.dhaeninck@ugent.be

Abstract

In this article we describe the overall structure of our Virtual Research Environment 'TIC Collaborative', designed for the study of nineteenth and early twentieth century international congresses and organizations related to social and cultural reform issues. The VRE consists of a digital text platform, which includes documents related to the congresses, organizations and their members, and a relational database in which we gather biographical information about people who played a role in this transnational arena. Both elements are geared towards interactive use and 'scholarly crowdsourcing'. In order to demonstrate the potential of our VRE we present one case study in which we use longitudinal data visualisations to develop targeted research questions about congress participation and transfer of knowledge.

Keywords: Virtual Research Environments, social reform movements, network visualisation.

1. Introduction

During the nineteenth century, ideas, policies and practices circulated in numerous ways (visits, journals, correspondence, international gatherings, learned associations, and the like), and thus gave rise to a transnational social and discursive field related to social reform (Leonards & Randeraad 2010, 2015; Topalov, 1999; Saunier, 2012). Issues engendered networks, also across national borders, which have been compared to a 'nébuleuse réformatrice' (Topalov, 1999). While historians of social reform movements and collective action are increasingly adopting network concepts and perspectives in their work (Lemercier, 2011; Diani, 2011), their use of formal network analytical tools is still quite limited, mainly due to heuristic barriers. It is no mean achievement to identify valid historical data in order to chart the evolution of networks and issues, and to find, process and visualize relevant biographical data. In this paper, we illustrate the potential of a Virtual Research Environment (VRE) to operationalize the theoretical notions developed in network analysis with the help of (relatively) large amounts of historical data. The VRE functions as an infrastructure to manage, structure and optimize a research driven dataset, described in greater detail below. Apart from describing the general structure of the VRE we present a case study that highlights the potential of *longitudinal data visualizations* to examine social and cultural reform movements in the nineteenth century. We are primarily looking at participation in international reform congresses, which we assume to be a strong indicator of transnational social engagement. International congresses can be seen as laboratories of new expert knowledge (Randeraad, 2011). They were – par excellence – the sites where scientists, administrators, politicians, artists, and others reform-minded elites of different countries met and exchanged ideas. They were places, in other words, where the 'rooted cosmopolitans' of the nineteenth century connected the local, the national and the global (Tarrow, 2005).

2. A Virtual Research Environment for the study of social reform movements

TIC-Collaborative (www.tic.ugent.be) is a digital humanities project for the study of the interconnectedness of ideas and practices in matters of social reform as well as cross-border contacts between reformers. Most digital humanities projects tend to start out from existing, often random collections, thus limiting the degree to which specific questions that interest historians can be answered. Instead, in our approach the corpus selection is based on close connection with the world of social and cultural reform in the nineteenth century, and is intrinsically research driven. Our overall methodological approach is characterized by a combined use of both 'traditional' qualitative approaches as well as digital methods. Many of the advanced mining techniques—such as topic modeling, text clustering, Named-entity recognition—are currently still in an experimental stage, but recent findings show the exciting potential of multilingual text-mining in large historical datasets and text corpora (Huijnen et al., 2014). Distant reading, until now mainly used for the confirmation of existing historical knowledge, can provoke new questions and research objectives (Gibbs & Owens, 2012). We aim to avoid common pitfalls by using a hybrid approach, combining data visualization techniques (Börner & Polley, 2014) and 'traditional' close reading.

The VRE applies the principle of scholarly crowdsourcing and offers access to dispersed sources in a

central observatory. The platform includes (1) a text corpus with annotation tools and export functionalities powered by *Islandora*. *Islandora* is an open source digital repository system based on Fedora Commons and Drupal. The text platform currently includes digitized primary sources concerning about 800 international congresses (proceedings, reports, lectures and attendees lists), international organizations, and other relevant sources (periodicals, pamphlets, and memoirs).

Next to a text platform, TIC Collaborative also includes (2) a collaborative relational database with biographical information on congress participants and their affiliations (which can be further processed in with social network tools, prosopographical methods, etc.). The collaborative biographical database is powered by *Nodegoat*. Nodegoat is a 'web-based database management, analysis and visualisation platform. Using Nodegoat, scholars define, create, query, update, and manage any number of data sets by use of a graphic user interface' (http://www.nodegoat.net). Nodegoat is object-oriented, which means that in the data model people, events and other 'objects' are treated as equal (more information in Van Bree & Kessels, 2014). Nodegoat is primarily concerned with the creation and contextualization of single objects that move through time and space, but queries and selections can be made, also for analysis outside Nodegoat. It is possible to export selections in order to process data in external software, for instance the popular open-source network analysis and visualization software package *Gephi* (http://gephi.github.io/) or software for multivariate analysis in the context of a prosopography.

Finally, the VRE also includes (3) a group bibliography, managed externally in Zotero (https://www.zotero.org). A Drupal module provides one-way syncing from Zotero to the bibliography of TIC Collaborative.

Generally speaking, virtual research environments provide tools to exploit data collections. We see our VRE first and foremost as a scientific community-building project rather than as a generic infrastructure project. Without community buy-in, we believe, the VRE cannot fulfill its function (Carusi & Reimer, 2010). Researchers are therefore triggered to get involved according to the principles of information crowdsourcing. The platform is designed to allow all parties involved in the project to reflect on, experiment with and refine the joint working methods (van der Vaart, 2010). Putting scholarly crowdsourcing into practice, TIC collaborative builds on joint efforts to collect biographical data and 'disambiguate' persons. The principle of scholarly crowdsourcing is also applied for the enrichment of metadata and correction of OCR-ed (Optical Character Recognition) sources, as the quality of OCR and metadata is a crucial condition for subsequent searches, computational linguistic research (e.g. Named-entity recognition, see below) and data mining in general.

3. Case study: longitudinal data visualisations

One of the main research interests in our project is the analysis of the emergence and development of institutional ties, generated by *multiple memberships* of social reformers. Mapping the multiple memberships of activists is an accepted way for studying the evolution of networks and organizational exchanges over time (Miche, 2007; Diani, 2011). In this example, we focus on Belgian and Dutch participants in a sample of international congresses in order to determine whether the results are sufficiently interesting to do further research on 'national cohorts' in a transnational framework. It is important to note that in the second half of the nineteenth century the number of international *organizations* was quite low. International congresses were often a first step towards institutionalization, or functioned more or less as organizations. Participants were often referred to as 'membres'. We therefore feel free to follow the methodology of Rosenthal et al. who were able to create a 'genealogy of causes' in nineteenth-century New York State, focusing on the multiple memberships of women active in social reform movements (Rosenthal et al, 1985). In this pioneering study, biographical dictionaries were used to map the affiliations between women and organizations between 1840 and 1914. Not the interconnections between the women, but the interconnections between the organizations, were the primary subject of their study. The number of mutual members or joint ties allowed the authors to make clusters of women's reform organizations. In a similar way we look at memberships of and participation in international congresses to establish whether these events could have functioned as hubs of knowledge transfer.

Looking at multiple memberships in this way allows us to compare the structure of social reform networks in particular moments in time, to identify both the persons and their institutional background (governments, learned societies, universities, welfare organizations and the like). Most international gatherings we inventoried for this article were part of series of congresses (hence highlighting the idea of 'membership'): peace congresses, statistical congresses, social sciences congresses, welfare congresses, (international) Dutch language conferences, penitentiary congresses, conferences of teachers for the deaf and the blind, congresses on artistic and literary property, congresses on the reform and codification of law, and congresses on temperance and prohibition. In total we gathered data from 54 congresses with 2890 unique participants originating from the Low Countries. We identified four moments: 1849-1853 (10 congresses), 1859-1863 (13 congresses), 1869-1873 (13 congresses) and 1879-1883 (18 congresses) to allow comparisons through time. To be sure, we have not yet inventoried all international socio-cultural reform congresses in this period, due to missing attendance lists, but with the data we have it is possible to discover meaningful patterns and

present intermediary results.

Mapping the evolution of networks over time by tracing clusters in certain sequences or snapshots in time is a good way to include and study network dynamics as exemplified by Katherine Giuffre (Giuffre, 1999). Strongly influenced by Bourdieus' notion of 'trajectory' (the series of positions successively occupied by an actor in the field), Giuffre argues that the statuses of the actors are defined by the relative positions of others actors to whom they are (indirectly) tied. Changes over time in a constantly shifting web of relationships indicate changes in status, but from our perspective also changes in personal interests and, on a different level, also organisational change. If we want to capture change, we must take time seriously. By comparing and connecting snapshots, we do. Snapshots implicitly imply that the actors and their ties have a start date and an end date. Bearing in mind that this is always somewhat arbitrary – even when we start from the actors' self-definitions of groups or networks - we add a clear temporal dimension to the development of networks and the issues that were at stake within them (Lemercier, 2014).

The use of graphic imaging of social networks has always had an important role in social network research. Network visualisation improves the communication and potential significance of relational data, and it helps to explore network properties (Brandes et al, 1999). From the beginning of SNA, drawings of networks have been used both to discover insights into network structures and to communicate those insights to others (Freeman, 2000). In our case, the latent patterns in the transnational socio-cultural reform network we want to visualize refer to Belgians and Dutchmen clustered by shared congress visits or co-memberships. We used a hierarchical clustering technique, which is a way to re-evaluate an entire network and to group actors together, who share similar positions with regard to the totality of positions in the network (Giuffre, 1999).

Our activists from the Low Countries visiting socio-cultural reform congresses were plotted in a two-step approach. First, a two-mode network (reformers and congresses) was plotted by the use of the Gephi software for each of the four periods in question (Figure 1, first snapshot). Second, the Jaroslav Kuchar's plugin was used to convert the two-mode network (persons x congresses) to a hierarchically clustered one-mode network of congresses (Figure 2). The visible, direct strength of the relations between the congresses, i.e. the number of shared participants, is indicated by the size of the line. The size of the nodes reflects the total number of visitors from the Low Countries (*indegree*). In order to exclude incidental participants, we only took persons into account who visited two or more congresses.

Finally, we combined the four snapshots. As a result, we obtained a graph, which allows us to compare network structures over time. The densely connected graph indicates a (strong) presence of Low Country reformers in the network (both synchronically as diachronically), as well as strong shared patterns of congress visits (Figure 3). It is, however, important to note that the most visited congresses were held in the Low Countries, and were expected to attract larger number of visitors from the organizing states.
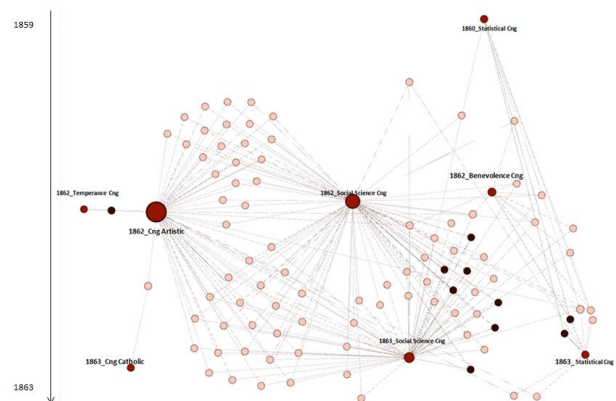

Figure 1: Socio-cultural reform congress visited by Dutchmen (dark) and Belgians (light).
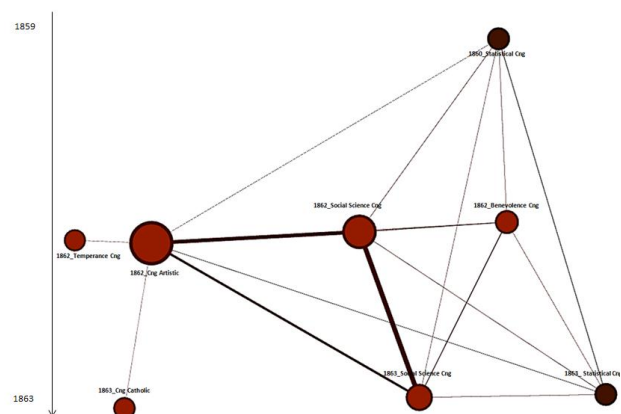

Figure 2: Socio-cultural reform congress linked by shared visits.

In Figure 2 and 3 we calculated the modularity, which is indicated by the different colours of the nodes. The modularity is a measure to calculate the strength of division of a network into clusters, or sets of nodes. In our network, congresses are connected by shared visitors. In other words, the modularity measure groups congresses together visited by the same Belgian and Dutch reformers. Four clusters can be discerned. The red-coloured group on the left (Figure 3), mostly statistical and hygienic congress, and the pale-coloured group in the centre of the graph, for the larger part artistic and linguistic conferences, are thematically and homogeneously grouped. The centre right congresses, the darker nodes, contain a wider range of themes. What stands out is the strong centrality of the multidisciplinary social science

congresses. This pattern follows the inquiry of Van Praet and Verbruggen who indicated the significance of the international social science congresses for establishing personal connections and information flows on public education and many other fields (Van Praet & Verbruggen, 2015). In the right margin, the congresses are grouped that were (almost) not visited.
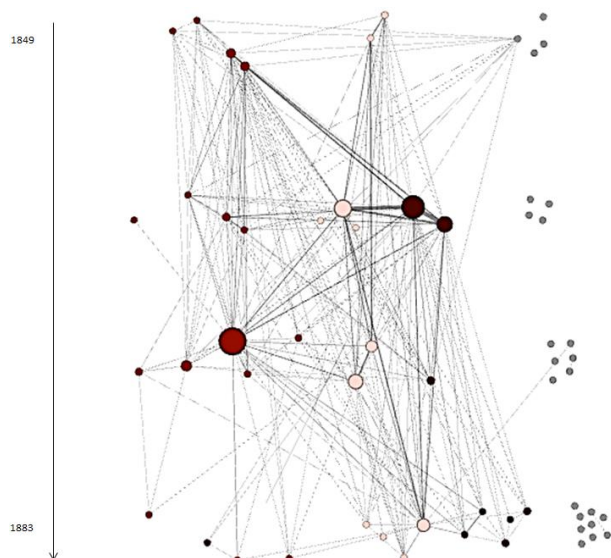

Figure 3: Socio-cultural reform congress linked by shared visitors. The color of the nodes cluster congresses with relatively more shared visitors.

On the basis of a time-bound macro perspective we can address larger questions related to core-periphery relations and cohesive subgroups on the level of the interconnectedness of local reformer milieus as well as the formation of expert groups and dilettantes. From a micro-perspective embeddedness in networks matters for recruitment in campaigns (especially in advocacy networks), knowledge circulation etc. From a combined macro- and micro-perspective, multiple and shared memberships constitute movement milieus, communication and interacting settings (Mische, 2008).

At this stage we did not take into account common biographical attributes, except nationality. A micro analytical or prosopographical approach is virtually impossible for this sample. It is hardly feasible to include social attributes such as age, gender, education, profession, ideology and religion into the analysis for each and every conference participant, even if we limit ourselves to the Low Countries. By conducting SNA, we have been able to pinpoint relatively small groups and brokers and to identify cohesive subgroups or cohorts that call for further research, as been argued before (Verbruggen, 2006). In doing so and in addition to the identification of the local, regional and national organizational 'roots', we can trace and compare the professionalization of welfare in the Low Countries and for instance the emergence of the 'social engineer' and the above mentioned expert formation.

Finally we also intend to include a discussion of the content of the relationships, and the dynamics of the transferred ideas and practices (which are often absent from most formal network approaches into social and intellectual movements). In this we follow Frickel and Gross (2005) who suggested to combine insights from the sociology of ideas, the history of science and the literature on social movements, in order to explain the dynamics of scientific and intellectual movements such as social reform movements. Interrelated objectives, movements dynamics and mobilization structures can indeed be related to the so called "framing" of meanings and issues in different settings (Benford & Snow, 2000).

4. Concluding remarks: named entities and linked data

The case study also tempts us to look beyond the current possibilities, in particular at Named-entity recognition and options for linking data. Our collective database of reformers can also be used for semi-automatic generation of networks in our text corpus. Originally developed by computational linguists as an information extraction subtask, Named-entity recognition has subsequently attracted the attention of researchers in various fields but in historical research it still is in an experimental stage. The original concept of a "named entity" (NE), proposed by Grishman and Sundheim (1996), covered names of people, organizations, and geographic locations as well as time, currency, and percentage expressions. The NER task is strongly dependent on the knowledge bases used to train the NE extraction algorithm. Leveraging on the use of for example DBpedia, Freebase, and YAGO, recent methods have been introduced to map entities to relational facts exploiting these fine-grained ontologies (van Hooland et al, 2013). In addition to the detection of a NE and its type, efforts have been made to develop methods for disambiguating information units with a Uniform Resource Identifier (URI). But even better results are expected from the use of a predefined and thematically coherent list of people and organizations to perform NER tasks.

In the next phase of our project we intend to make full advantage of the possibilities of NER and Linked Data in general. The documentary value is obvious. Linking our biographical data to external sources is also a step in the consortium's internal workflow for the disambiguation of persons and dealing with the transliterations of names in different languages. Authority control and the common use of authority lists is also the first step towards the integration of similar or complementary data collections related to international organisations and transnational elites. In accordance with the International Council on Archives we hope to use authority records as authorized names that identify and describe the named entity (i.e., as a starting point, persons). A few options have been considered. Biographical portals such as the *Biographie-Portal* oriented towards German speaking countries (http://www.biographie-portal.eu/en/) and the

Dutch Biography Portal (http://www.biografischportaal.nl/en/) only include biographical information published in national biographical dictionaries or other similar entries. At this stage it is impossible to add new biographical descriptions. VIAF in its turn is library-oriented and not every social activist or social reformer included in our data collection meets the requirements for inclusion in Wikipedia/DBpedia. Finally we are considering using ODIS to link with (http://www.odis.be). ODIS is a contextual database with data on intermediary structures (civil society), persons and families, events and buildings related to their history and periodicals and archives. ODIS currently is in a transition from a mainly Flemish project into an international contextual database.

None of these options completely fulfills the project's needs. Instead of the exclusive use of one of these endpoints, we should rather link our data collection to a wide range of collections containing (mainly) biographical information, including VIAF, Wikipedia and ODIS. A new SPARQL module is to be included in TIC Collaborative's Nodegoat-installation in order to query and link external information from these endpoints. Further steps, however, are needed, if we want to link data from complementary projects, not least to agree upon an exclusive use of a single endpoint.

5. Acknowledgements

6. References

Brandes, U. et al. (1999). Explorations into the Visualization of Policy Networks. *Journal of Theoretical Politics*, 11(1), pp. 75--106.

Benford, R.D., Snow, D.A. (2000). Framing Processes and Social Movements: an Overview and Assessement. *Annual Review of Sociology*, 26, pp. 611--639.

Carusi, A., Reimer, T. (2010) *VRE Collaborative Landscape Study.* Bristol: JICS, pp 106.

Diani, M. (2011). Social movements and collective action. In J. Scott and P. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis*. London: Sage Publications, pp. 223--235.

Freeman, L. (2000).Visualizing Social Networks. *Journal of social structure*, 1.

Frickel, S., Gross, N. (2005). A general theory of scientific/intellectual movements. *American Sociological Review*, 70(2), pp. 204--232.

Gibbs, F.W., Owens, T.J. (2012). The Hermeneutics of Data and Historical Writing. In J. Dougherty, K. Nawrotzki (Eds.), *Writing History in the Digital Age, web-book edition.* Michigan: University of Michigan Press.

Giuffre, K. (1999). Sandpiles of Opportunity: Success in the Art World. *Social Forces*, 77(3), pp. 815--832.

Huijnen, P. et al. (2013). A Digital Humanities Approach to the History of Science: Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining. In A.

Nadamoto et al. (Eds.), *SocInfo 2013 Workshops Proceedings*, Berlin, pp. 71--85.

Keck, M. E., Sikkink, K. (1998). *Activists beyond borders: advocacy networks in international politics*. Ithaca: Cornell University Press.

Lemercier, C. (2014). *Taking time seriously. How do we deal with change in historical networks?* (Paper presented at HNR 2014, Ghent).

Leonards, C., Randeraad, N. (2010). Transnational Experts in Social Reform, 1840-1880, *International Review of Social History* 55(2), pp. 215--239

Leonards, C., Randeraad, N. (2015). Building a transnational network of social reform in the 19th century, in D. Rodogno, B. Struck and J. Vogel (Eds.), *Shaping the Transnational Sphere. Experts, Networks and Issues from the 1840s to the 1930s*, New York. pp. 111--130

McCarty, C. et al. (2007). A Comparison of Social Network Mapping and Personal Network Visualization. *Field methods*, 19(2), pp. 145--162.

Mische, A. (200*8). Partisan Politics. Communication and Contention across Brazilian Youth Activist Networks*, Princeton, Princeton University Press.

Randeraad, N. (2011). The International Statistical Congress (1853-1876): Knowledge transfers and their limits. *European History Quarterly*, 41(1), pp. 50--65

Rosenthal, N., Fingrutd, M., Ethier, M., Karant, R. and McDonald, D. (1985). Social Movements and Network Analysis: A Case Study of Nineteenth-Century Women's Reform in New York State. *American Journal of Sociology*, 90(5), pp. 1022--1054.

Saunier, P.-Y. (2012). La secrétaire générale, l'ambassadeur et le docteur. Un conte en trois épisodes pour les historiens du «monde des causes» à l'époque contemporaine, 1800-2000. *Monde(s)*, 1, pp. 29--46.

Tarrow, S. (2005). *The new transnational activism*. Cambridge: Cambridge University Press.

Topalov, C. (Eds.) (1999). *Laboratoires du nouveau siècle. La nébuleuse réformatrice et ses réseaux en France (1880-1914)*. Paris: Editions de l'Ecole des Hautes Etudes en Sciences Sociales.

Van der Vaart, L. (2010). *Collaboratories: Connecting Researchers – How to facilitate choice, design and uptake of online research collaboratories*. Utrecht: University Press.

Van Bree P., Kessels, G. (2014), Mapping Memory Landscapes in nodegoat in Social Informatics. In L. Aiello and D. McFarland (Eds.), *Social informatics*. Paranaque City: Springer International Publishing, pp. 274--275.

Van Hooland et al., (2013). Named-entity recognition: A Gateway Drug for Cultural Heritage Collection to the Linked Data Cloud? *Literary and Linguistic Computing*, 1, pp. 1--16.

Van Hooland et al., (2012). Evaluating the success of vocabulary reconciliation for cultural heritage collections. *Journal of the American Society for Information Science*, pp. 1--21.

Van Praet, C., Verbruggen, C. (2015). 'Soldiers for a Joint Cause': a Relational Perspective on Local and International Educational Leagues and Associations in the 1860s. *Bmgn-the Low Countries Historical Review*, 130 (1), pp. 4--24.

Verbruggen, C. (2007). Combining Social Network Analysis and Prosopography. In K. Keats-Rohan,

Prosopography Approaches and Applications. A Handbook. Oxford: Occasional Publications UPR, pp. 579--601.