WORKING PAPERS
EUROPEAN STUDIES
AMSTERDAM

11

Joep Leerssen
with Pim van Bree, Geert Kessels and Maarten Witberg

SpInTime: Dynamically visualizing how diffusion patterns evolve over space and time

Europese Studies, Universiteit van Amsterdam
2012

Joep Leerssen is Academy Professor and holds the chair of Modern European Literature at the University of Amsterdam.
Pim van Bree and Geert Kessels are graduate students at the University of Amsterdam and software engineers at LAB1100.
Maarten Witberg (www.maartenwitberg.nl) is a database developer

MaartenWitberg developed portals and search routines in the FileMaker database; Pim van Bree and Geert Kessels developed the javascript rendering; Joep Leerssen developed the overall concept and basic database structure.
The project was developed for, and funded by, SPIN, the Study Platform on Interlocking Nationalisms; www.spinnet.eu.

ISSN 1871-1693

Working Papers European Studies Amsterdam is a series of incidental publications by staff members, associates and collaborators of the Department of European Studies at the University of Amsterdam.
Printed texts in brochure format can be obtained (free of charge for individuals) from the Secretariat, European Studies, Universiteit van Amsterdam, Spuistraat 134, 1012 VB Amsterdam, Netherlands; e-mail: secr.es-fgw@uva.nl
The texts can be downloaded in PDF format
from www.hum.uva.nl/europesestudies > Research > Working Papers

SpInTime: Dynamically visualizing how diffusion patterns evolve over space and time

Joep Leerssen
with Pim van Bree, Geert Kessels and Maarten Witberg

Making cultural nationalism visible and datable

Funded by the NWO Spinoza Prize, the Study Platform on Interlocking Nationalisms (SPIN) was established at the end of 2008 to study the transnational connections between culture-nationalist activities in 19th-century Europe. The challenge was threefold: [a] to show that cultural nationalism was in itself a historical agency, with its own inner dynamics (and not just a passive by-product of societal developments or of political nationalism); [b] that, counter to what one might expect (given the intra-national, inwardly focused preoccupations and characteristic exceptionalism of cultural nationalism), the ideology itself was transnational, manifesting itself in parallel forms in countries that were often very dissimilar politically, socio-economically, and in terms of modernization processes; [c] that the remarkable simultaneity and parallellisms between culture-national activities at various points of the European map requires a historical explanation.

Various initiatives (still ongoing, see www.spinnet.eu) have been taken to tackle these questions, revolve largely around the historical-evenemential visibility of cultural nationalism and its cross-national manifestations. This also involved the necessity making cultural nationalism *datable*, since all too often political and social historians have dealt with “culture” as a general ambience without agency, and hence without concrete chronological fix-points, identified only by vague, unspecified indicators such as "the spreading influence of Romanticism" or "the rise of medievalist taste". In short, the programme is to establish the specific *when* and *where* of cultural nationalism. Once that precondition is met, research can proceed to test explanatory models or working hypotheses, one of these being a

viral, network-based diffusionism

Going by well-established research results on the development of literary Romanticism in Europe,[1] it seems plausible to assume, as a working hypothesis, that cultural nationalism which arose concurrently and in close association with Romanticism (and which we may therefore be justified in calling "Romantic nationalism") spread, not just because similar socio-economic or political conditions in different countries led to similar cultural responses, but largely also as part of a transnational communicative process ("cultural transfer"). The tenets of Romantic nationalism spread, not only within a given society, but also from one society to another, as a result of textual mediation and transmission. That intercultural communication existed is an obvious fact: how the examples of the Grimm Brothers and Walter Scott sparked folktale collections and historical novels in different countries is a matter of record. Solidarity movements among various national revivalisms, across national borders, (such as philhellenism and pan-Slavism) are likewise well-documented. These communicative conduits may account for the previously mentioned, remarkable simultaneity of Romantic nationalism in countries as far apart as Iceland and Bulgaria, and as dissimilar as Flanders and Estonia.

Network theory and network visualization

A pilot project to address these research perspectives was conceived: mapping multi-party communicative networks, as exemplified by the correspondence of philologists (scholars working on languages, text editions and oral literature; a cohort of intellectuals with a measurable importance in Romantic nationalism; Leerssen, *op. cit.*). The example of Jacob Grimm, who was simultaneously in touch with Finnish, Breton and Balkan scholars, suggested that through such nodal interconnections, philologists from the most distant and dissimilar European regions might be linked into a very "small world", involving only two, at most three degrees of separation.[2]

---

[1] Cf Joep Leerssen, "Literary historicism: Romanticism, philologists, and the presence of the past", *Modern language quarterly* 65.2 (2004): 221-43, and sources marshalled there.

[2] *Degrees of separation* (DoS) are defined by the number of intermediary contacts through which two given individuals are connected. If A and B are in direct correspondence, their DoS is 1; if A and B do not correspond directly, but both correspond with X, their DoS is 2. A DoS of 3 means that the closest connection beween A and B runs through the fact that A's correspondent X corresponds with B's correspondent Y (which means that there is 1 degree of separation between A and X; X and Y; Y and B; and 2 degrees of separation between A and Y and X and B). Schematically:

To demonstrate and document such a network, and its high European connectedness, visualization was an obvious necessity. The obvious starting point was *Pajek* (http://pajek.imfm.si/doku.php), a world standard in network visualization software with many high-quality, expert applications. Studying the existing literature on Pajek[3] proved highly useful, not only in order to grasp the technicalities and applicabilities of the digital environment, but also to harness our thoughts on the nature of networks as such. For one thing, it allowed us to think of correspondence networks in terms of "vertices" (the individual correspondents) and "arcs" (the letters exchanged). This insight may seem blindingly obvious to network analysts but in its culture-historical application is anything but trivial, as we hope to suggest at the conclusion of this article.

Existing solutions and their limitations

However, we identified three complicating features.

[1] Pajek deals with pre-entered data, in a stripped-down plain-text tabular format. The data with which we deal are cultural and textual, at best evenemential, rather than quantitative or statistical. A text-friendly data environment was needed (also allowing Unicode characters from well beyond the range of plain-vanilla ASCII values); a database, in other words, in which we could assemble the relevant multinational information from actual correspondence archives and editions; and that text-friendly database environment would need to be open to analytical search and selection scripts (e.g., to select networks consisting of immediate contacts, or else of contacts at 2 or 3 degrees of separation) before exporting its data, or subsets of its data, for visual rendering.

[2] The main virtue of Pajek is its malleability to represent, graphically and dynamically, the uneven distribution of connection density within a network. Vertices are aligned and their mutual positions and distances are "weighted" so as

---

1 DoS: A—B
2 DoS: A—X—B
3 DoS: A—X—Y—B

It is a working hypothesis that no two philologists in 19th-century Europe were removed by more than 3 degrees of separation, and that the close-knit nature of this network accounts for the high degree of cross-national synchronicity and similarity in their activities.

[3]Vladimir Batagelj & Andrej Mrvar, "Pajek: A Program for Large Network Analysis", *Connections* 21.2 (1998): 47-57. Online at http://pajek.imfm.si/lib/exe/fetch.php?media=slides:pajek98.pdf; Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj. *Exploratory social network analysis with Pajek*. Cambridge University Press, 2005. More generally, also the Pajek wiki at http://pajek.imfm.si/doku.php.

to give spatial expression to relative intensities of connectedness; and this process can be dynamically adjusted according to various well-established and trustworthy algorithms. This feature, however was in fact less important to us than a fixed geographical representation, locating vertices (individuals) not on the variable basis of their connectedness, but on the fixed basis of their geographical locations. While Pajek allows for such a geographic fixation of vertices, it is a complicating variant rather than a basic feature of its design.

[3] For historical purposes it was vital that we could make these networks, not just spatially visible, but also chronologically datable. To show a network, that existed over a certain period of time, in collapsed form, i.e. at an ideal-typical single moment, would not suit our purposes; yet that is precisely what most network visualization does. To trace the morphing evolution of a networks as certain partners join, other drop out at successive moments, contacts intensify and dwindle, and correspondents move from one place to another, was the very essence of our ambition; and the possibilities offered by Pajek would tend towards a incremental accumulation of separate slice-moments rather than a fluid diachronic capture of evolutions in time. High-quality GIS software such as *ArcGIS* (www.esri.com/software/arcgis) or its open-source equivalent *Quantum GIS* (www.qgis.org) is primarily oriented towards visualizing geographical patterns rather than communicative relations and networks; and while animation routines are available to display changes over time, these are labour-intensive to construct and the end-result appears to be a closed-off display window with little room for user input or adjustment except for speed of animation.

We addressed these challenges as follows.

A database of correspondence metadata

In order to collect and pre-format the data about letter correspondences, we opted for FileMaker Pro as database software, which recommended itself because of its relational and scripting possibilities and its adaptability for user-defined, text-friendly layouts and interfaces.

We identified the following parameters as "vital metadata", being indispensibly meaningful for the eventual visual rendering:

[a] identity of sender;
[b] identity of receiver;
[c] location of sender
[d] location of receiver
[e] date of letter.

Sender and receiver, [a] and [b], were entered as given name, family name, and

initials, with additional "calculation fields" to parse these into various alternative name variants. We also defined calculation fields to give alternative spellings of these names, stripping diacriticals with >128 values in ASCII to their <128 form (*å* as *a*), as well as rendering >128 values as HTML codes (e.g. *å* as *å*) for proper in-browser display.

The names of locations ([c] and [d]) were entered in a "repeating field", so as to allow for name variants (e.g. Istanbul, Constantinople, and Tsarigrad), with, again, calculation fields rendering the principal name variant in ASCII-stripped and HTML-coded formats. In addition, the sender/receiver locations were also identified in terms of their latitude/longitude coordinates, entered in decimal form.

The database structure took an obvious relational form:
[1] a table of individuals, which could be drawn upon both as [a] "senders" and [b] "receivers" (**ill. a**);
[2] a table of locations (which could be drawn upon both for [c] "sender's location" and [d] "recipient's location", specifying latitude/longitude coordinates (**ill. b**);
[3] a table of letters, specifying each letter in terms of its [e] date and other metadata ([a] through [d], as preformatted in tables [1] and [2])[4] (**ill. c**).

Scripts and "portals" were devised in FileMaker to perform certain analytically important search-and-select functions, e.g.
– to select the letters involved the exchange network of a specified set of named correspondents, at various degrees of separation;
– a “route planner” to trace the connection(s) between any two given individuals through intermediaries at a specified Degree of Separation;
– a routine to export selected data or subsets as a JSON javascript file.

Rendering in a browser: cross-platform ease of use and portability

The javascript formatting of the letters’ metadata was chosen so as to make it

[4]This table was also given additional fields allowing for the closer characterization of letters (hostile, collaborative, banal etc.), correspondents, and intrinsic features (style of address; references to, or relayed communication from, third persons; mention of publications) etc. A placeholder field was also created to allow for inclusion of the letter's complete text, although this information was not as yet relevant for the initial proof-of-concept network-analytical setup. In addition, calculation fields drawing on the data entered in tables [1] and [2] can highlight "weak ties", e.g. to identify letters between correspondents with different native languages, or over a user-specified distance threshold. For the concept of "weak ties", see the classic articles by Mark S. Granovetter, "The strength of weak ties", *American journal of sociology* 6 (1973): 1360-1380, and "The strength of weak ties: A network theory revisited", in *Social structure and network analysis*, ed. P.V. Marsden & N. Lin (Beverly Hills, CA: Sage, 1982), 105-130.

possible to render exported data visually in an internet browser. Initially, the arcs and vertices of the correspondence networks were mapped onto a geographical overlay downloaded from Google Maps, but at a later stage we opted instead for locally stored geographical image files. Local data storage makes it possible to perform the entire rendering process offline, with the browser used purely as a javascript processing and visualization environment, not as a link with the internet. This avoids the need to rely on platform-specific and costly dedicated rendering software, and allows for a "lean" file format with low memory demands. The metadata of a set of 1600 letters exported as JSON file is only 450 kB in size; the entire file set to drive the browser's rendering process is 10 MB, of which 7.3 MB is accounted for by the geographical image files for the map overlay. A zip file of the CSS and javascript files can be taken to any platform and can be rendered in the local PC's browser, without the need to install or to apply dedicated software.

Displaying both space and time: *slide rule, dateline and period lens*

The spatial distribution of correspondence networks over the European landmass worked problem-free, the browser environment allowing for zooming and click-dragging geographical regions. Multiple letters between a pair of correspondence could be mapped in slightly offset near-overlaps, so as to render a "thicker" arc, the thickness indicating a more intense exchange of letters (cf. **ill. d**). This side-by-side offsetting of arcs also prevents the masking of one set of relations (between a given pair of locations) by another.

In order to display diachronic distribution (i.e. mapping the evolving and morphing shapes and shifting densities of networks over time), we opted for a "slide rule", consisting of a period lens sliding along a dateline. A fixed chronological *dateline* is displayed as a horizontal bar at the bottom of the screen and covers, from left to right, the duration of the correspondence as defined by the earliest and latest date in the dataset. A width-adjustable "window" or *period lens* can slide along this dateline, the effect being not unlike the sliding cursor on a classical slide rule. (cf. **ill. e** and **f**)

Only those letters are displayed in the browser whose dates fall within the *period lens's* temporal focus. Widening the period lens will bring letters over a longer period into view, narrowing the period lens will tighten the selection of visible letters. Sliding the period lens across the dateline will bring letters into view as their date falls within the temporal focus, and will eliminate letters from view as the focus moves away from their date. Sliding a narrowly-focused period lend will give a greater sense of dynamic change, sliding a wider-focused period lense will give a greater sense of structural patterns. Sliding only the right-hand

(end-date) edge of the period lens across the dateline while leaving the left-hand (start-date) edge in place will progressively broaden the temporal focus and give a sense of the network's gradual accumulation.

In all these cases, the effect is a fluid, dynamically morphing and user-driven visualization of how networks evolve both in space (across the European map) and in time (over the years).

A proof of concept involving a variety of datasets was placed online at www.spinnet.eu on 29 February 2012, together with more specific instructions on how to manipulate and use the datasets, selection(s) of correspondents and the slide rule. Some illustrative screenshots are given at the end of this Working Paper

Future prospects

We feel that the "slide rule" concept and the use of javascript-driven browser rendering present a viable perspective for the diachronically dynamic visualization of evolving spatial diffusion patterns. These need not necessarily be in the nature of a correspondence network. Sample datasets in the online "Proof of Concept" display 19th-century translations of Walter Scott's novel *Ivanhoe* (cf. **ill. g**),[5] the 19th-century proliferation of male choral societies, or a "route finder" tracing the shortest connection(s) between two given individuals (cf. **ill. h**). The choice of Europe as a geographical space is just that: a choice, not a necessity.

At the moment, SpInTime is designed to work on a stand-alone basis, without the need for internet access, but requiring the underlying use of a FileMaker-driven database. Alternative possibilities are conceivable; an online version making use of MySQL is at present being developed. Conversely, it may be possible to display the visualization wholly within the FileMaker environment by means of its WebViewer functionality, which would make "kiosk"-style runtime applications (standalone, offline or online) thinkable, e.g. for interactive museum displays.[6]

At present, the database with its highly-involved script and portal structure provides the primary dataset selection tool, with additional selections and "tweaks" being possible within the browser environment. We feel that this is a workable balance. To rely wholly on the browser environment for subset selection is unworkable: it would mean having to load overly large datasets, especially if

[5]On the basis of data in the online *Bibliography of Scottish Literature in Translation* maintained by the National Library of Scotland at http://www.nls.uk/catalogues/boslit.

[6]The limited javascript support of Internet Explorer 8 may present an obstacle in this case: on many Windows PCs, IE8 is the default browser technology used by FileMaker's Web Viewer.

databases will be populated to up to 70.000 records so as to become truly representative of large-scale, long-term processes. (The Walter Scott correspondence contains 14.000 letters, the Grimm Brothers' correspondence ca. 20.000[7]). Providing preselections of exportable subsets within a robust database software environment seems indispensable. Still, the transition from database to browser presents a bottleneck, and should ideally be performed in a single, integrated, user-input dialogue process. This requires further thought, also because users may approach the database with as yet unforeseen search/selection requirements (e.g. filtering correspondences by country, correspondents' age, professional status or gender, etc. etc.).

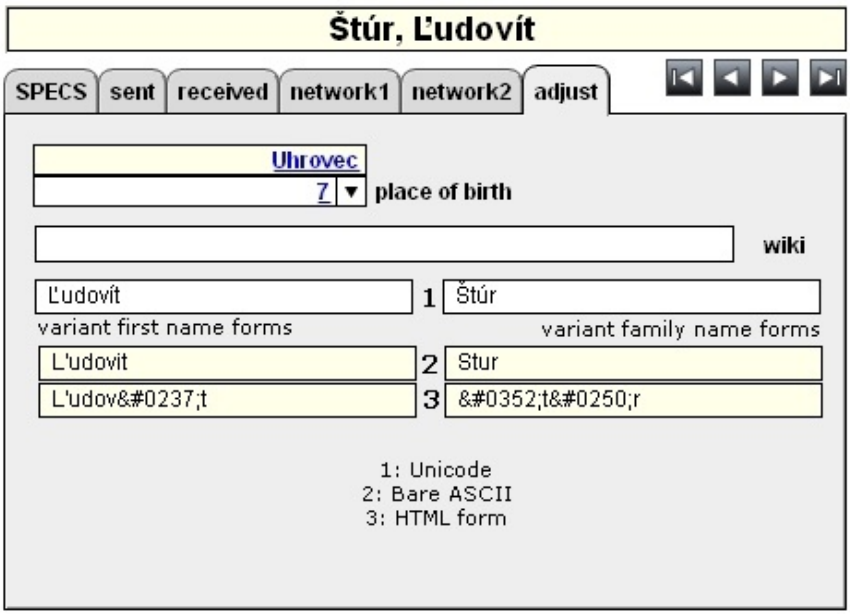
SPIN will welcome feedback and ideas for future applications and further development of this concept.

Analytical insights

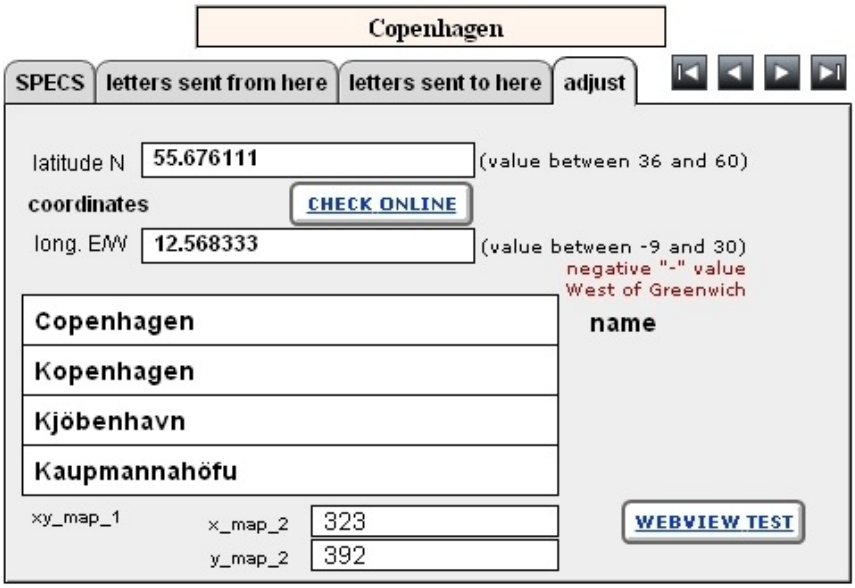
The proof of concept developed so far works on limited datasets and displays the properties, not of the historical past but of a researcher-imposed selection from that past. We must therefore be extremely cautious in drawing analytical conclusions from the materials presented here. Even so, the experience of conceptualizing a network visualization like SpInTime has offered some valuable insights. To begin with, SPIN's working hypothesis that no two philologists in 19th-century Europe were further removed than 3 degrees of separation (note 3, above) has been given increased weight. Moreover, to conceive of a correspondence network as a set of arcs and vertices is in itself a useful thought experiment. It shows that the communication act and the communicating individual, while being fundamentally different elements in a historical process, are each others' full equals in the cognitive hierarchy of how we conceptualize that historical process. The arc emanates from the vertex, the arc's end-points pinpoint the vertices. Each is meaningful only inasmuch as it defines the other. This insight is an important caution against our tendency to read texts purely as documentation of their author's mindset or situation. Texts should not automatically be subordinated to their authorial genesis, but should also be read functionally, as events and historically operative presences in their own right.

[7] See www.nls.uk/catalogues/resources/scott/index.cfm for the *Millgate Union Catalogue of Walter Scott Correspondence*, and www.grimmbriefwechsel.de/arbeitsstelle/arbeitsstelle.html for the *Briefverzeichnis* of the correspondence of Jacob and Wilhelm Grimm.

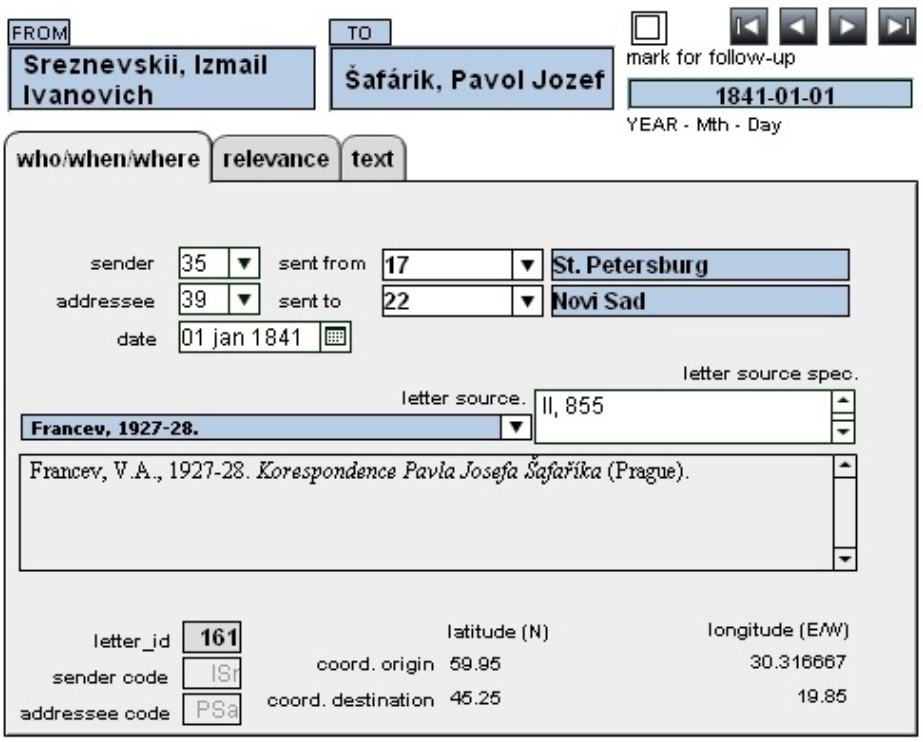
Illustrative screen shots



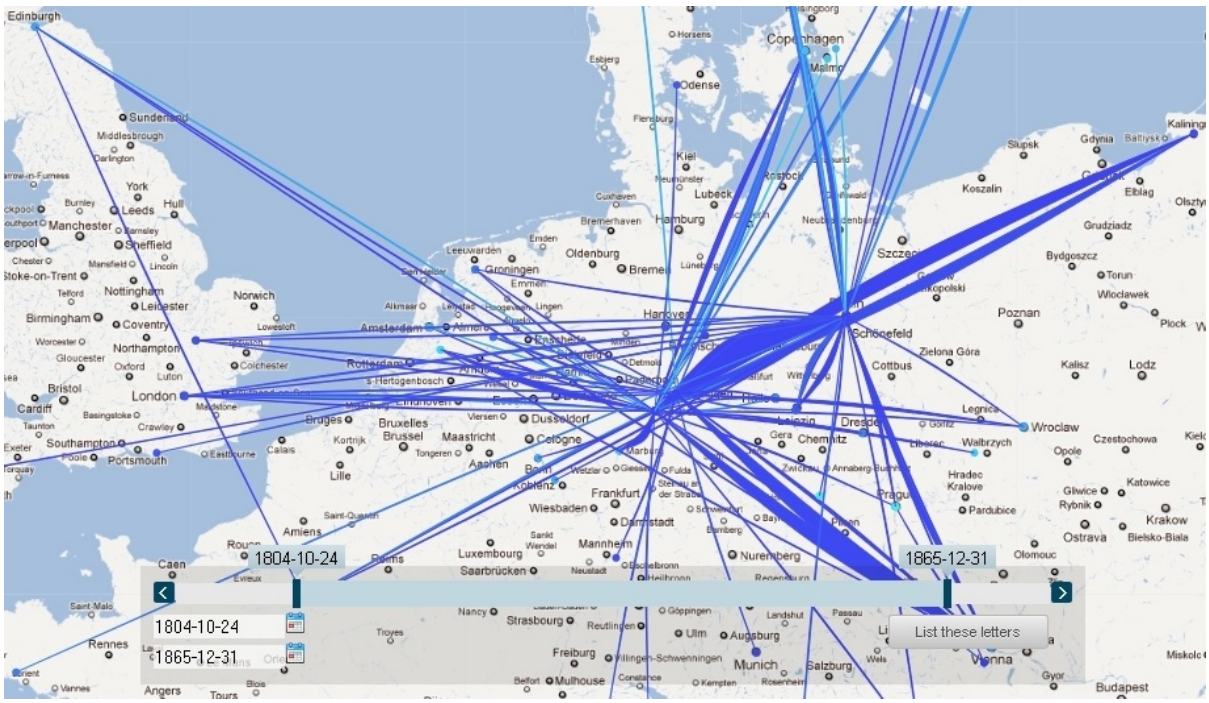
ill. a: database record (person)



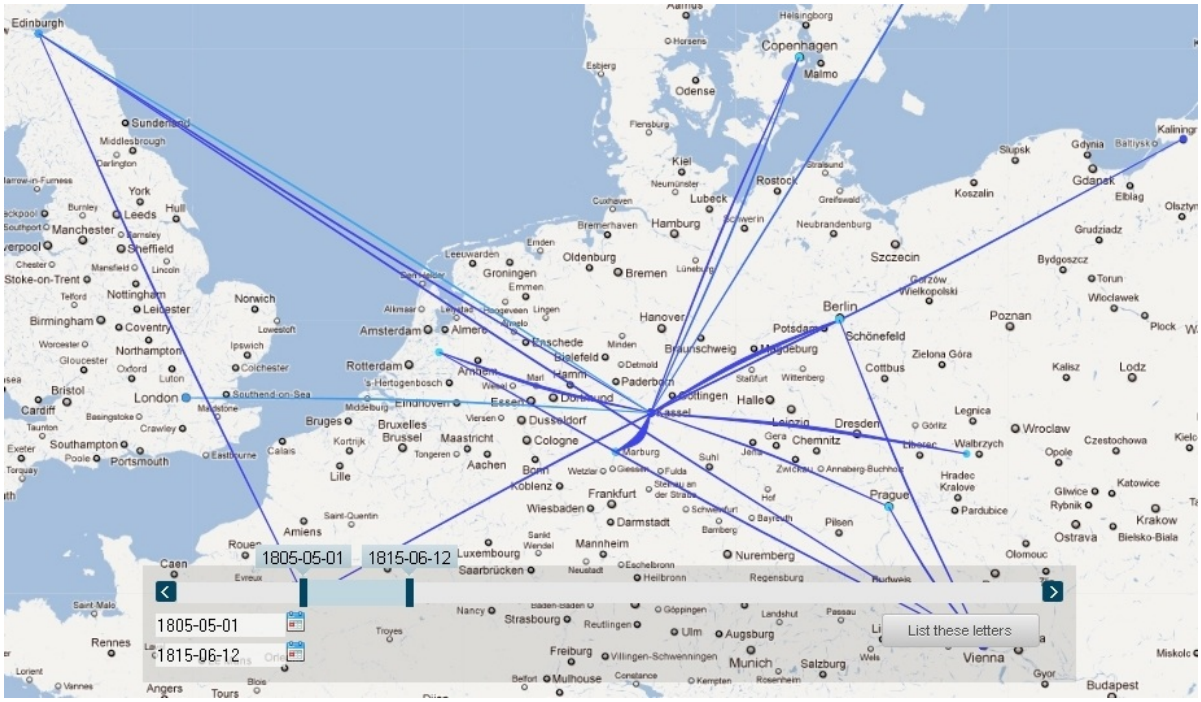
ill. b: database record (place)



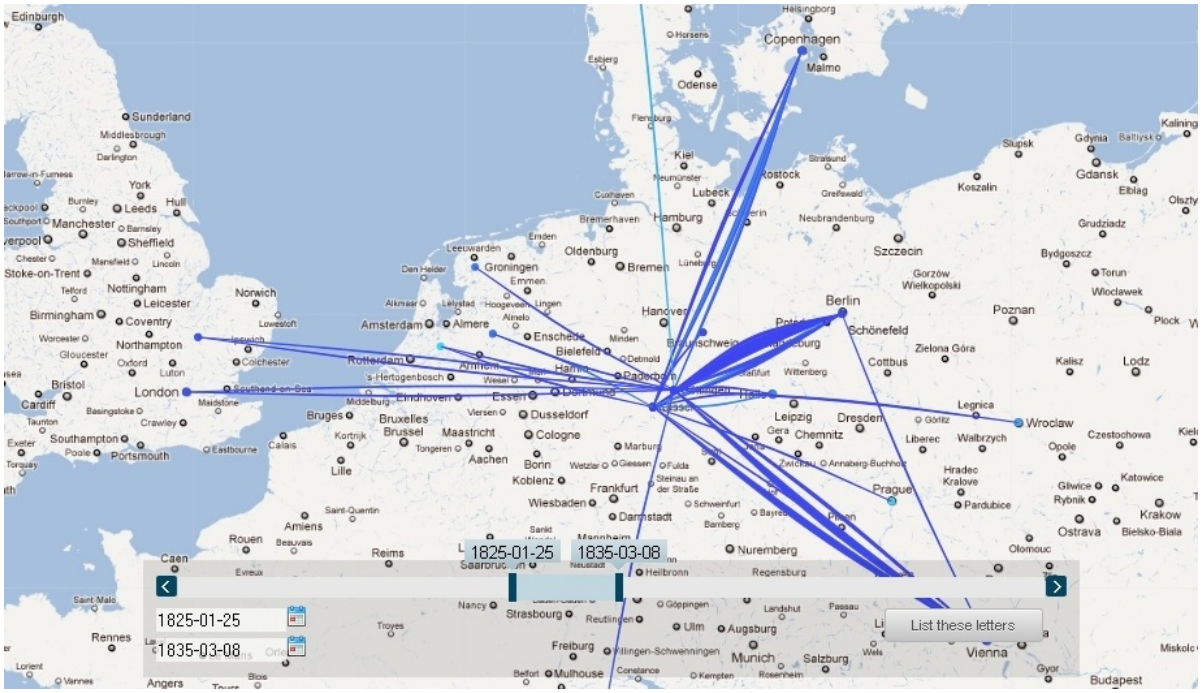
ill. c: database record (letter)



ill. d: Jacob Grimm's correspondence network (time lens fully opened)

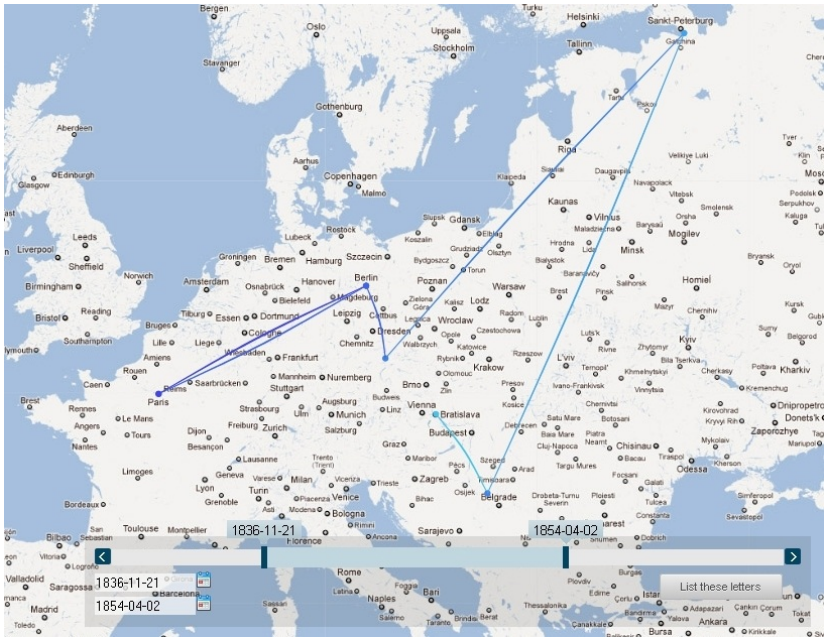


ill. e: Jacob Grimm's network (time lens set to period 1805-1815)



ill. f: Jacob Grimm's network (time lens set to period 1825-1835)

ill. g: Pre-1914 translations of *Ivanhoe*



ill. h: Route finder tracing connection between La Villemarqué and Štúr

published titles:

1. P.J.G. Kapteijn, *Over de Europese Grondwet* (2005)
2. Joep Leerssen, *The Cultivation of Culture* (2005)
3. Peter Hoppenbrouwers, *Medieval Peoples Imagined* (2005)
4. Iraklis (Hercules) Millas, *Tourkokratia* (2006)
5. Annie Jourdan, *Amsterdam en révolution, 1795-1798* (2006)
6. Annemarie van Heerikhuizen, *Een Haagse salon* (2007)
7. Max Bader, *Making Parties Fit for Democracy* (2007)
8. C.R.M. Versteegh, *Human Rights Enforceability* (2008)
9. Eva Navarro Martínez, *Images of Europe in American Films* (2009)
10. Joep Leerssen, *Irish Anthologies and Literary History* (May 2010)
11. Leerssen, van Bree, Kessels & Witberg, *SpInTime* (April 2012)