

Mapping the Geography of Science: Distribution Patterns and Networks of Relations among Cities and Institutes

Journal of the American Society for Information Science & Technology
61(8) (2010), 1622-1634

Loet Leydesdorff¹ & Olle Persson²

Abstract

Using Google Earth, Google Maps, and/or network visualization programs such as Pajek, one can overlay the network of relations among addresses in scientific publications onto the geographic map. We discuss the pros and cons of various options, and provide software (freeware) for bridging existing gaps between the *Science Citation Indices* and *Scopus*, on the one hand, and these various visualization tools on the other. At the level of city names, the global map can be drawn reliably on the basis of the available address information. At the level of the names of organizations and institutes, there are problems of unification both in the ISI-databases and with *Scopus*. Pajek enables us to combine the visualization with statistical analysis, whereas the Google Maps and its derivatives provide superior tools at the Internet.

Keywords: map, science, city, co-authorship, international, globalization, network

¹ Amsterdam School of Communications Research (ASCoR), University of Amsterdam, Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands; loet@leydesdorff.net.

² Department of Sociology, Umeå University, SE 901 87 Umeå, Sweden; Olle.Persson@soc.umu.se.

Mapping the Geography of Science: Distribution Patterns and Networks of Relations among Cities and Institutes

1. Introduction

In this communication we report on newly available methodologies to map the sciences both statically (at each moment of time) and dynamically (over time). These techniques enable us, among other things, to visualize patterns of international collaboration using a projection on the world map (e.g., Glänzel, 2001; Hicks & Katz, 1996; Persson *et al.*, 2004; Wagner, 2008; Zitt *et al.*, 1999). We compare the different possibilities in Google Earth, Google Maps, and Pajek, and report on dedicated software (freeware) available for making these projections using data from bibliographic databases such as the *Science Citation Index* and *Scopus*.

The geographic mapping of science can be distinguished from its cognitive mapping (Frenken *et al.*, 2009; Jones *et al.*, 2008; Small & Garfield, 1985). The sciences can be mapped cognitively, for example, in terms of journal maps (e.g., Leydesdorff, 1986; Tijssen *et al.*, 1987), co-citations (Small & Griffith, 1974; Small, 1999), or co-words (Callon *et al.*, 1983). Using techniques such as multi-dimensional scaling (e.g., Kruskal & Wish, 1978; Borgatti, 2001; Leydesdorff & Schank, 2008) or spring-embedded algorithms (e.g., Kamada & Kawai, 1989; Fruchterman & Reingold, 1999), information scientists have made considerable advances during the last decade in terms of agreeing on similarity criteria (Ahlgren *et al.*, 2003; Van Eck & Waltman, 2009), possible projections

(Boyack *et al.*, 2005 and 2007; De Moya-Anegón *et al.*, 2007; Klavans & Boyack, 2009a; Rafols & Leydesdorff, 2009), and even on standard colors for distinguishing among disciplinary affiliations (Klavans & Boyack, 2009b). The latter authors suggest that a “consensus” has emerged on the mapping. Rafols *et al.* (in preparation) concludes that therefore one could project developments in science against a statistical baseline.

Since in a socio-cognitive process such as the development of the sciences, change can take place at different levels at the same time, Studer & Chubin (1982, at p. 269) have noted that “(r)elationships among journals, individuals, references, and citations can be analyzed in terms of their structural properties. But can one be used as a baseline to calibrate our understanding of another? Does it make sense to attempt to “control” for one relationship while studying others?” Narin (1976) was the first to distinguish between nations and disciplines as two analytically independent baselines for evaluation (cf. Narin *et al.*, 1972; Narin & Carpenter, 1975). Small & Garfield (1985) proposed using these two dimensions as different bases for mapping.

Intellectual developments at the global level have also to be retained locally. National (or regional) governments develop science and technology policies for this retainment (e.g., Skolnikoff, 1993). Does investment in science pay off in terms of prominence and reputation, economic returns, or the emergence of transnational linkages like those envisaged by the European Commission? (Leydesdorff & Wagner, 2008, 2009; NSB, 2010, pp. 5-33 ff.). Are national governments able to formulate policies which provide them with a possible hold on “emerging technologies”? Has a knowledge infrastructure

developed that is sufficient to play a role in the case of “generic technologies”? These and similar questions require a geographic baseline for their assessment in addition to the cognitive map.

The geographic map, of course, provides us with a natural baseline for studying spatial dynamics. In recent years, software developments have made this map increasingly available for projections at different scales and with appropriate zooming techniques, such as in Google Maps. How can one make such techniques profitable for the enterprise of science and technology studies? Having both been deeply involved in developing software for using the information contained in databases such as the *Science Citation Index* and *Scopus* for purposes of mapping, we thought it timely to provide a state-of-the-art review of the current possibilities and limitations of geographic maps. Where necessary, we have further developed our software for bridging gaps and made these tools available from our respective websites. The interested reader can find instructional materials and manuals at these sites (<http://www.leydesdorff.net/maps> and <http://www8.umu.se/inforsk/bibexcel/>, respectively).

2. Methods and materials

For didactic purposes we shall use a standard set for the various visualizations. We chose to use the footprint of the field of information science (IS) in 2009 as available in the address information in the bylines of the publications. How did we delimit this field? First, Library & Information Science (LIS) is categorized as a separate subject in the

Social Science Citation Index, but this category covers 61 journals. These lists, however, are composed for the purpose of information retrieval and are therefore not sufficiently restricted for mapping a specific field (Leydesdorff & Probst, 2009). More restricted lists of IS have been proposed in the literature. Following White & McCain (1998, at p. 300), Zhao & Strotman (2008, at p. 2072) recently provided an updated list of eight journals representing the core of IS, in their opinion. Using aggregated journal-journal citation data from the *Journal Citation Report* 2008, we found 13 journals that contribute more than 1% to the citations of *JASIST* and 11 journals that contribute more than 1% to the citations of *Scientometrics*. These two sets overlap in eight journals, of which six are also included in the list of Zhao & Strotman (2008).

Since we wished to include also the newly added *Journal of Informetrics*, we gave priority to our citation-based definition of the field and included these eight journals in the analysis (Table 1). Using a search string based on these eight journals, 621 *articles* could be retrieved from the Web-of-Science published in the year 2009.¹ Because we limited the set to articles, however, no records from the *Annual Review of Information Science and Technology*—including 10 reviews and one editorial in 2009—were retrieved. We limited the analysis to this set (available at <http://www.leydesdorff.net/maps/data.zip>). The file includes 1,479 authors at 1,107 institutional addresses.

	Zhao & Strotman	Citation	Citation	Journals included
--	-----------------	----------	----------	-------------------

¹ The records were downloaded on January 14, 2010. The search string was: so=(annual review of information science “and” technology or journal of information science or journal of information science or journal of the american society for information science “and” technology or scientometrics or journal of informetrics or information processing management or information research an international electronic journal or journal of documentation) and document type=(article) timespan=2009

	(2008)	environment <i>JASIST</i> (2008)	environment <i>Scientometrics</i> (2008)	in this analysis
<i>ACM Transactions of Information Systems</i>		+		
<i>Annual Review of Information Science and Technology</i>	+	+	+	(+)
<i>Computation and Human Behavior</i>		+		
<i>Decision Support Systems</i>		+		
<i>Information Processing & Management</i>	+	+	+	+
<i>Information Research</i>		+	+	+
<i>Journal of the American Society for Information Science and Technology</i>	+	+	+	+
<i>Journal of Documentation</i>	+	+	+	+
<i>Journal of Informetrics</i>		+	+	+
<i>Journal of Information Science</i>	+	+	+	+
<i>Library & Information Science Research</i>	+	+		
<i>Knowledge Organization</i>		+		
<i>Online Information Review</i>			+	
<i>Proceedings of the ASIST Research Evaluation</i>	+		+	
<i>Research Policy</i>			+	
<i>Scientometrics</i>	+	+	+	+

Table 1: Core journals of Information Science according to Zhao & Strotman (2008); on the basis of the citation impact of *JASIST* and *Scientometrics* in 2008; and our selection of eight journals in this study.

In a later section, we compare this set with a similar set downloaded from the *Scopus* database. This set contained 551 articles published in 2009 for the same seven journals. The difference (of 70 articles) occurs because the two databases are organized differently. *Scopus* uses publication dates and not tape years: these articles were downloaded on

January 23, 2010.² However, the institutional addresses are organized differently in *Scopus*. Since some of our programs carefully parse the address information, we elaborated a previously existing routine (Scop2ISI.Exe, available at <http://www.leydesdorff.net/software/scop2isi>) in order to make the address information in the *Scopus* data as comparable with ISI-data as possible. The current version correctly displays most of the nodes and links on the maps, but the labels may still be incomplete.

The data can be processed using BibExcel (available at www.umu.se/inforsk/bibexcel/) or ISI.Exe (at <http://www.leydesdorff.net/software/isi/index.htm>). The latter routine was further refined for the purpose of this project into Cities1.Exe (at <http://www.leydesdorff.net/maps/index.htm>). We discuss these dedicated extensions below as they become relevant to the argument.

The 1,107 addresses contain 385 unique city names and 593 unique institutional addresses.³ These contained 591 and 697 postal addresses⁴ which could be provided with geo-coordinates, respectively, at <http://www.gpsvisualizer.com/geocoder/>. Yahoo! was

² The search string in *Scopus* was: (PUBYEAR IS 2009 AND SRCTITLE(scientometrics)) OR (PUBYEAR IS 2009 AND SRCTITLE(journal of informetrics)) OR (PUBYEAR IS 2009 AND SRCTITLE(journal of documentation)) OR (PUBYEAR IS 2009 AND SRCTITLE(journal of information science)) OR (PUBYEAR IS 2009 AND SRCTITLE(journal of the american society for information sc*)) OR (PUBYEAR IS 2009 AND SRCTITLE(information processing AND man*)) OR (PUBYEAR IS 2009 AND SRCTITLE(information research)) AND (LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "ar") OR LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO(EXACTSRCTITLE, "Journal of the American Society for Information Science and Technology") OR LIMIT-TO(EXACTSRCTITLE, "Scientometrics") OR LIMIT-TO(EXACTSRCTITLE, "Information Processing and Management") OR LIMIT-TO(EXACTSRCTITLE, "Journal of Information Science") OR LIMIT-TO(EXACTSRCTITLE, "Journal of Documentation") OR LIMIT-TO(EXACTSRCTITLE, "Journal of Informetrics") OR LIMIT-TO(EXACTSRCTITLE, "Information Research")).

³ The 551 records from *Scopus* contain 801 institutional addresses of 1014 authors. At the level of the set, 457 city names are unique and 687 institutional names in this case.

⁴ This number is larger because of different postcodes provided for the same city.

used for obtaining the coordinates.⁵ With the exception of one institutional address (“Isle Man Int Business Sch, Douglas 1M2 1QB, UK”), all coordinates could be retrieved automatically. Incomplete address information in the original data of course leads to error.

The asymmetrical matrices of documents as cases versus cities or institutions as variables were transformed into symmetrical co-occurrence matrices of links among these cities and institutions, respectively. These matrices were the input to the further analysis and mapping. The city or institutional nodes can be scaled with the respective number of occurrences (or the logarithm thereof), as will be indicated where appropriate in the text. The width of the links is proportionate to the number of co-occurrence relations. We first develop the argument with the city names because the institutional addresses generate some further complications (which will be discussed in a later section).

3. Google Earth and Chaomei Chen’s *CiteSpace*

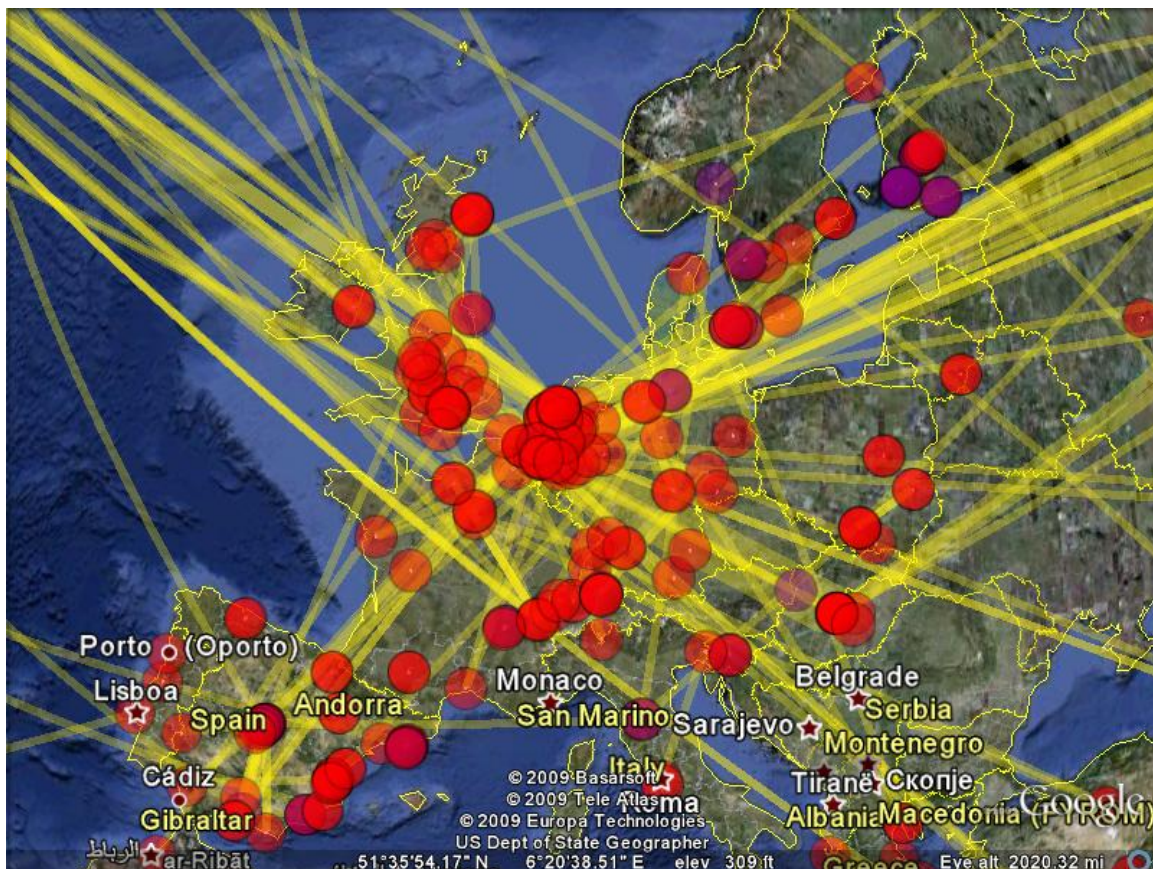
The first application that made it possible to generate geographic maps of science in the Google format was Chaomei Chen’s program *CiteSpace*. Chen and colleagues have reported on themes such as “(v)isualizing and tracking the growth of competing paradigms” (Chen *et al.*, 2002; cf. Chen, 2003) since the early 2000s. The program has been elaborated ever since and is publicly available as *CiteSpace II* at <http://cluster.ischool.drexel.edu/~cchen/citespace/>.⁶ This program requires as input a download of the data in the standard (tagged) format at the Web-of-Science interface of

⁵ We found the geo-encoder of Yahoo! currently more successful in retrieving Asian addresses than the one at Google.

⁶ *CiteSpace* assumes the presence of the Java VM at the local computer.

the *Science Citation Indices*, and then allows the user to make a geographic map of the institutional addresses and their relations—in addition to the many other facilities for citation analysis that this program offers (e.g., Zhu *et al.*, 2008).⁷

When one clicks within *CiteSpace* on the tab “Geospatial Maps” in the main menu, one can select the option “Google Earth (KML)” independently of performing a citation analysis of the data. In the resulting file, links can be included optionally in addition to the nodes. The program then generates a so-called .kmz file which is the standard input for Google Earth (Figure 1).



⁷ One can use *Scopus* data in *CiteSpace* after parsing them with *Scop2Isi.Exe*. The network and nodes will be displayed correctly, but the labels may sometimes contain insufficiently standardized information.

Figure 1: European centers and their network in the IS set 2009 as output of *CiteSpace* as represented in Google Earth. Shades of red and purple for the nodes indicate differences in numbers of publications. (In order to enhance visibility when printing in black and white, the color of the network links was changed from red to yellow.)⁸

Figure 1 provides the result for Europe using our data set. Within Google Earth, one can zoom in or out and click on links and nodes to obtain precise address information. In the output file of *CiteSpace*, the nodes vary in size as stacked bars which can be seen by tilting the image horizontally. However, in Google Earth the background is not adjustable from the satellite image to a street map as with Google Maps, and because of the satellite's position in the projection one is not able to draw a global map.

One may therefore wish to bring this information under Google Maps. Google Maps reads .kmz files when uploaded to a website. It is also possible to unzip .kmz-files to the .kml format which one can read and edit as a text file.⁹ (KML is a markup language like HTML.) The resulting kml-file (available at <http://www.leydesdorff.net/maps/master-medium.kml>) contains all the information in the map, but this file cannot easily be parsed and changed, for example, in order to modify the node-sizes in accordance with the volume of publications. However, one can read this file using the web address of the upload within Google Maps. Alternatively, there are sites on the Internet where one can interactively visualize one's kml-files, such as at <http://display-kml.appspot.com/>. Using Google Maps, the problem of different backgrounds can be solved, and the global map can also be drawn.

⁸ The line colors in the output of *CiteSpace* indicate different years of publications. In this case, however, the publications were all in 2009.

⁹ Google Maps and Google Earth are able to read both .kmz and .kml files.

4. Google Maps and Google Earth

The facility to read .kml files into Google Maps provides us with many options to generate maps from the data by parsing and reformatting them into this rich markup language. However, the kml-language was primarily developed for Google Earth. (A subset of kml can also be read by Google Maps for Mobile.) Thus, the functionality in Google Maps is restricted to only a subset of tags. For example, one cannot scale the node sizes in Google Maps, but one can do so by using the same file in Google Earth. Google Earth, however, does not allow us to show the global map at a single glance because of the globe format of the visualization, and has the noted disadvantage of only a single “satellite view” for the mapping. However, this image can be overlaid with street names and one can tilt the image.

The various possibilities for using a kml-file make this option a potentially attractive alternative for a number of applications. The zoom-facilities in Google Maps and Google Earth are superior. Thus, we decided to develop this interface further. For this purpose, the existing routine ISI.Exe¹⁰ was further elaborated into Cities1.Exe, which can be retrieved from <http://www.leydesdorff.net/maps/index.htm>. This program is called Cities1.Exe because after an intermediate step one needs Cities2.Exe to complete the routine. Cities1.Exe reads the same data as *CiteSpace*, but allows users to set relevant

¹⁰ Available at <http://www.leydesdorff.net/software/isi/index.htm>.

thresholds (either in absolute values or as percentages) on the fly, and to include the generation of a cosine normalized matrix in addition to a co-occurrence matrix.¹¹

The next (intermediate) step is to read the file cities.txt—which is one of the outputs of Cities1.Exe—at a geo-coding website which adds the geographical coordinates to the city names and postcodes. Geo-coding this information can be done, for example, at <http://www.gpsvisualizer.com/geocoder/>.¹² The program Cities2.Exe reads the output of this geocoder and generates, among other things, the file cities.kml (available at <http://www.leydesdorff.net/maps/cities.kml>) which can be uploaded and read by Google Maps or directly into Google Earth. (The various processing steps are summarized in an Appendix. Instructions are provided at <http://www.leydesdorff.net/maps/index.htm>.)

¹¹ In the case of large matrices, the generation of a cosine-normalized matrix may be time-consuming. The generation of a co-occurrence matrix can be speeded up by using the file matrix.txt in Pajek for the generation of an affiliations matrix. (This option is further explained at <http://www.leydesdorff.net/maps/index.htm>.)

¹² Unlike the geo-coder at Google, the one at Yahoo! led in our data to the retrieval of nearly 100% of the addresses.

and links can be visualized, and at variable scales (e.g., globally, nationally, or regionally).

Figure 3: Global map of information science with the network of coauthorship relations using Google Maps (with <http://www.leydesdorff.net/maps/cities2.kml>).

In summary, the advantages and disadvantages of using Google Earth and/or Google Maps are a bit complex, but the kml-file offers a set of options. Google Maps is

particularly useful for the global perspective and for showing the network dynamics. If the size of the vertices matters, and the perspective is not global but local (or regional), Google Earth provides an alternative to Google Maps, since this program allows for the visualization of the sizes of the nodes using kml.

5. The GPS Visualizer

As noted, the kml-language is not central to Google Maps since it was developed for Google Earth. The focus of developers is nowadays on feeding Google Maps with Javascripts using an API (that is, an application programming interface; e.g., Zoss *et al.*, 2010). However, this is not easy for the unskilled programmer. Fortunately, a number of websites come to the rescue of the user. One of them is the GPS Visualizer at http://www.gpsvisualizer.com/map_input?form=data. This site allows the user to input data either interactively or to read a file containing the required input information directly from one's disk. Cities2.Exe makes this file available as "inp_gps.txt." (See for an example, at http://www.leydesdorff.net/maps/inp_gps.txt.)

One can interactively change the various parameters of the data points on the Google Map to be drawn to the screen.¹⁵ Furthermore, the color of the nodes can be chosen in the input file (e.g., inp_gps.txt). Cities2.Exe colors connected nodes red and unconnected ones orange as the default, but one can edit the file. (Of the 392 nodes used in this study, 97 were not connected in the network.) Alternatively, BibExcel.Exe now contains a

¹⁵ Our programs work optimally with the input type set to "default" (instead of "waypoints" which is default); "Colorize using this field" set to "custom field"; "Resize using this field" to "custom field", and "Custom resizing field" to "n".

module for generating this webpage on the basis of ISI data at

<http://www8.umu.se/inforsk/geography/BibExcelGPSexercise.xls>.

The Google Map which is generated at this interface can be saved both as a picture and in terms of the generating source code (containing Javascripts). One can adapt this source code within the html. For example, at <http://www.leydesdorff.net/maps/IS2009.html>, the zoom was reset at “2” instead of “1” for esthetic reasons. The resulting files work promptly at one’s local computer. Before the upload, however, one has to add a “Google Map API key” at the place which specifies “var google_api_key = ' ' ;” (that is, line 62) within the code. These API keys are freely and instantaneously available for each web address at <http://code.google.com/apis/maps/signup.html>.

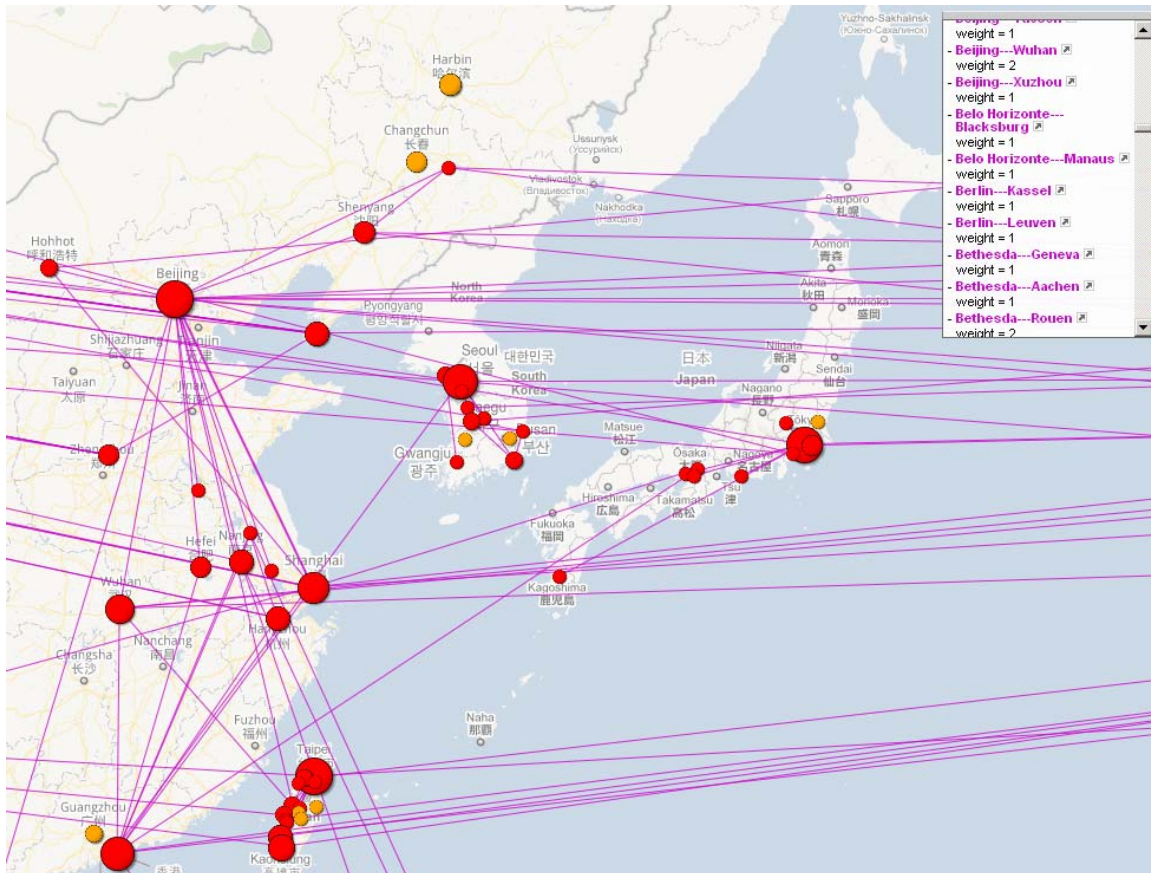


Figure 4: Visualization of IS in 2009 in East Asia using Google Maps via the GPS Visualizer (at <http://www.leydesdorff.net/maps/is2009.html>).

In summary, the use of the GPS Visualizer has advantages over feeding kml files into Google Maps. One can vary the sizes and colors of the nodes and the weights of links. Furthermore, one can make an animation at the web using a so-called redirect statement in the html (e.g., `<meta http-equiv="refresh" content="5;url=page2.html">`).¹⁶

For example, Figure 4 provides a zoom of the file for East Asia. The top-right pane provides users with direct access and zooming to each link in the network. Not surprisingly, the link between Beijing and Shanghai is strongest in this network, with a value of six. As noted, we scale the nodes in numbers and the weights of the nodes default with the logarithm in order to avoid too heavy links and nodes in the visuals.

6. Pajek

In addition to the kml files and the input for the GPS Visualizer, Cities2.Exe also generates a file “cities.paj” (available at <http://www.leydesdorff.net/cities.paj>) which can be read into Pajek¹⁷ as a project file (by using <F1>). Drawing this file provides a visualization with sizable arcs and vertices. The vertices are proportionate to the logarithm of the occurrences plus one (since the $\log(1) = 0$); and the links are proportionate to the co-occurrences. All statistics available in Pajek can be applied (De

¹⁶ See <http://www.basictips.com/html-slideshow-5-easy-steps.shtml>.

¹⁷ Pajek is available for non-commercial use at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.

Nooy *et al.*, 2005; Hanneman & Riddle, 2005). The cities are drawn at their coordinates, and one can directly compare the geographic map with layouts generated, for example, by using the algorithm of Kamada & Kawai (1989).

A layout in Pajek can be exported as a transparent overlay using the .eps format. Thus, one is able to overlay these results on any equirectangular projection of the world map. Additionally, we generated a world map in terms of coast lines which can be imported into Pajek and then merged with the overlay map.¹⁸ This file is available in Pajek format at <http://www.leydesdorff.net/maps/coast.zip>. If one reads this file into Pajek in addition to cities.paj, one obtains two networks which can both be selected (in two different Network windows) and then merged within Pajek using *Nets > Union of vertices*. One can color and size the network and the coastlines independently because the latter are defined in Pajek as edges and the former as arcs. The arcs can also be used asymmetrically as arrows, and thus one can also visualize flows using arrows (Börner *et al.*, 2006; Phan *et al.*, 2005; Tobler, 1987).

One can zoom into Pajek figures by marking a piece of the drawing with a right-clicked mouse. Using the *k*-core algorithm in Pajek teaches us, for example, that the core centers of the coauthorship network in Europe are mostly in Belgium: Antwerp, Louvain,¹⁹

¹⁸ The coast lines are based on the geographical coordinates of the Coast Line extractor available at the website of the National Geophysical Data Center (NGDC) at <http://rimmer.ngdc.noaa.gov/mgg/coast/getcoast.html>. We used the World Coast Line data designed to a scale of 1:5,000,000 for this purpose. In order to match the coordinates of Pajek's Draw window, which may vary between 0 and 1, we linearly transformed the latitudes and longitudes of coastlines and cities.

¹⁹ All 18 publications in the database with "Louvain, Belgium" as address are from the Katholieke Universiteit in Leuven, which in 2009 published only a single time with its Flemish city name.

Heverlee—a suburb of Louvain²⁰—Oostende, and Diepenbeek. (An additional relation with Budapest is generated by Wolfgang Glänzel, who routinely adds his affiliation in Budapest to his institutional address in Louvain.) Brussels moreover provides a secondary center with Hamburg, Geneva, Rouen, Paris, and Nantes. This prominent position of Belgium is an artifact of the common practice of authors in Flanders to publish papers individually at more than a single city address. We did not correct for this specific effect of the networking, which is induced by policies of the regional government of Flanders (Debackere & Glänzel, 2004).²¹

²⁰ One publication of the Catholic University of Leuven has exclusively an address in Heverlee, a suburb of Louvain which hosts one of the university's institutes (the Department of Mathematics).

²¹ The Flemish government uses a model ("BOF") for the funding of basic research in academia which is based on whole-number counting for each institutional address in a (coauthored) publication.

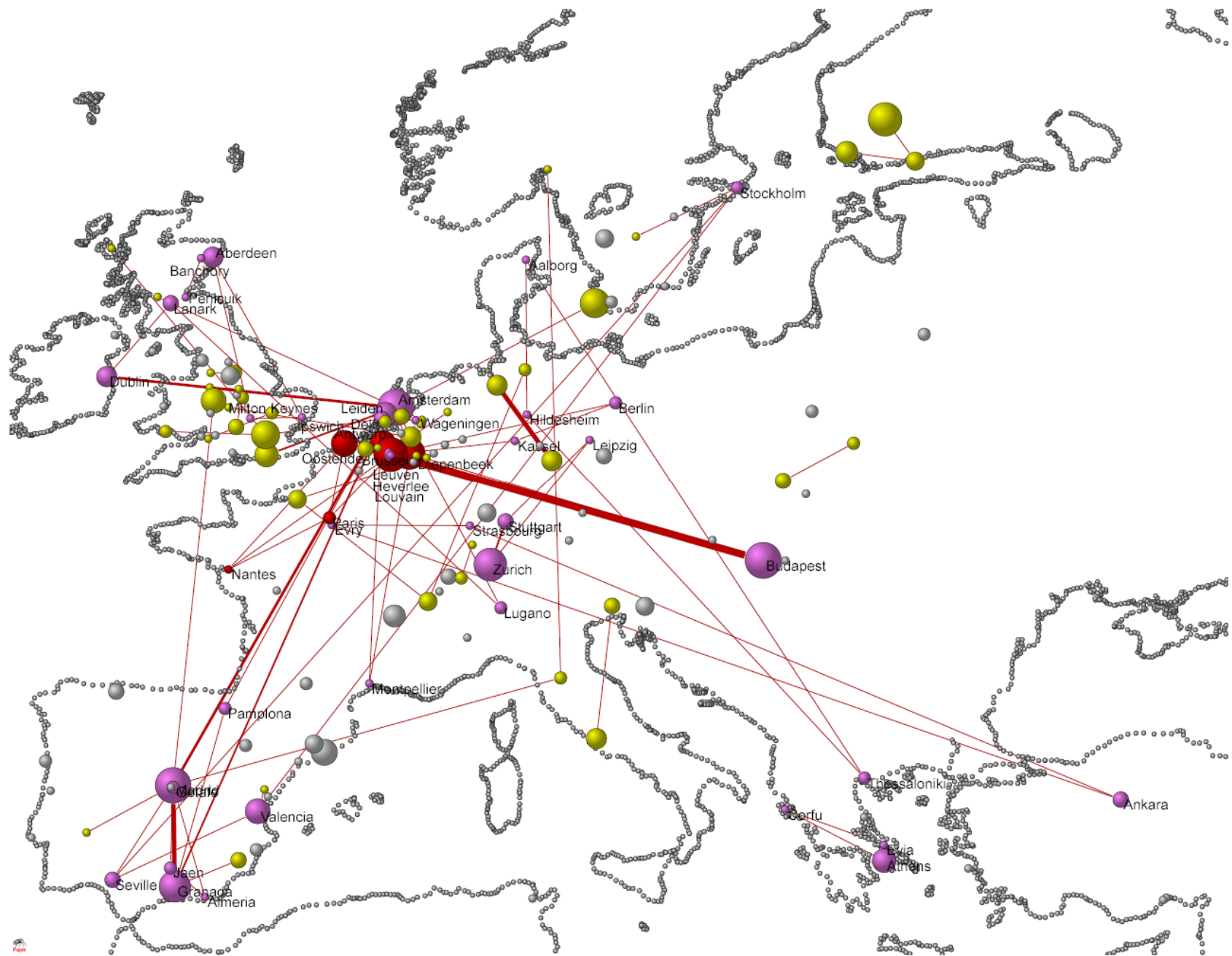


Figure 5: Primary and secondary centers—indicated in red and violet, respectively—in the European network of cities. (Colors were attributed using the k -core algorithm in Pajek.)

In a Pajek drawing of this network, many arcs cross the EU indicating relations between American and Asian cities. This can be prevented by choosing another projection of the earth or by refining the set (Figure 5).²⁴ In the exclusively European network, most core centers have lost one connection ($k = 3$ instead of $k = 4$).

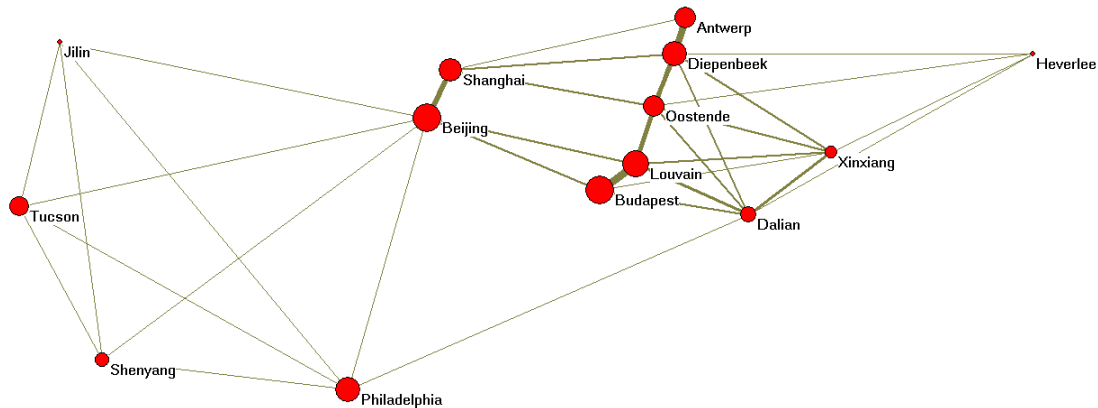


Figure 6: The network among 14 core cities in the network of IS 2009 (Kamada & Kawai, 1989).

Figure 6 shows the structure of the ($k \geq 4$) core network of the field. The Belgian groups do not collaborate internationally other than with cities in China (and, because of the noted affiliation, with Budapest, Hungary). As noted, these relations among the Belgian (and Hungarian) cities are largely spurious. The Chinese partners also have American collaborators.

²⁴ Coast lines for Europe and Korea in Pajek-format are available at <http://www.leydesdorff.net/maps/eurcoast.net> and <http://www.leydesdorff.net/maps/korcoast.net>.

Since one can package Pajek configurations using the project file format (.paj), the information can be communicated comprehensively. (One can find the results of these analyses as examples at <http://www.leydesdorff.net/maps/world.zip> and <http://www.leydesdorff.net/maps/europe.zip>, respectively.) However, unlike Google Maps, the resulting figures cannot be made interactive at the web. Animations, however, can be made using PajekToSvgAnim,²⁵ SoNIA²⁶ or the dynamic version of Visone²⁷ (Leydesdorff *et al.*, 2008). Furthermore, one can use the partitions generated in Pajek to edit precisely the color indications in the input files (inp_gps.txt) of the GPS Visualizer.

7. Institutional collaboration

Strictly analogous to the programs cities1.exe and cities2.exe, we also developed inst1.exe and inst2.exe. These latter programs include the first subfields of the institutional addresses in the ISI data in addition to the city, postcode, and country information. Using the GPS Visualizer, it thus becomes possible to map relations even at the street level (Figure 7).

²⁵ Available at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/SVGanim/>.

²⁶ <http://www.stanford.edu/group/sonia/documentation/install.html>.

²⁷ <http://www.leydesdorff.net/visone/index.htm>.

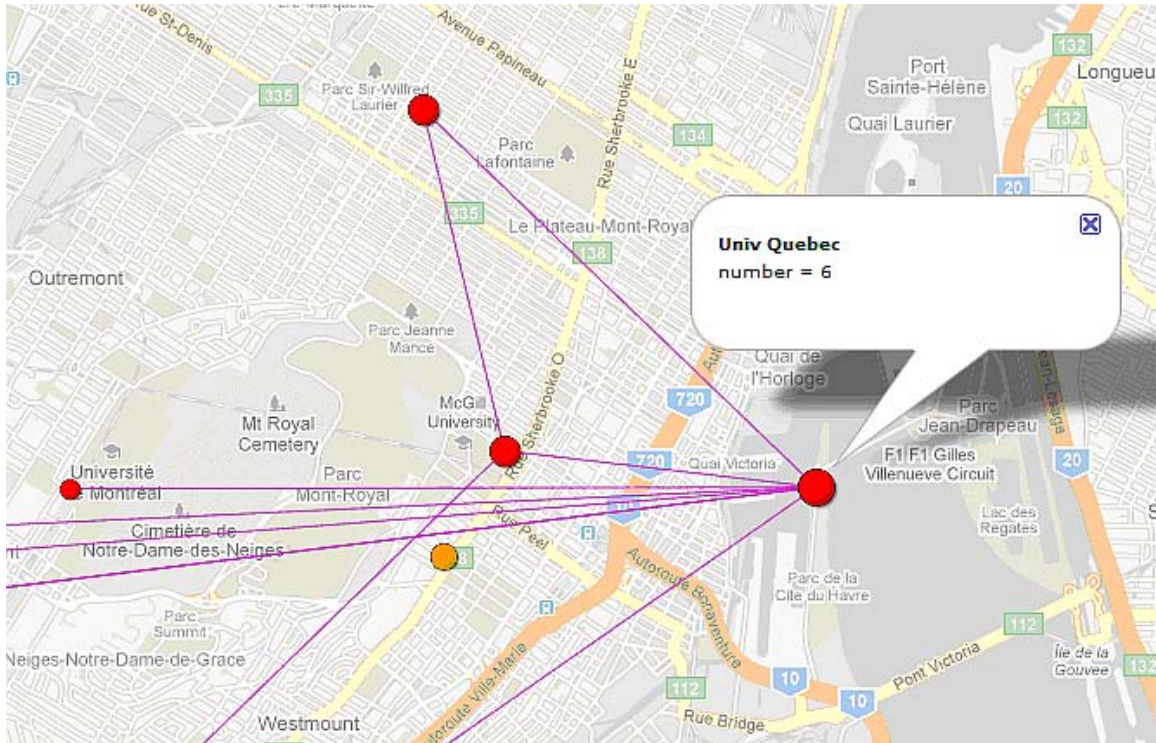


Figure 7: Relations among different institutions in Montreal. (The orange one is Concordia University and the top one *Sci Metrix*.)

The global map of institutions in the IS 2009 set can be retrieved at <http://www.leydesdorff.net/maps/institutions.html>. Of the 593 institutions, 128 centers were not connected to another one and are therefore colored orange in this map; 557 institution names are unique if one disregards the different street addresses. At <http://www.leydesdorff.net/maps/inst.kml> one can find the file which can be read into Google Maps and Google Earth in order to show the network relations.

There are a number of problems, because the same institution may publish with different addresses and addresses are often incomplete. Costas & Irribaren-Maestro (2007) noted that valuable address information can also be found in the address of the corresponding

author when it otherwise is missing from the record. We include this information in the analysis although sometimes it contains only the postal address and not the institute's name.

Institutional addresses are organized hierarchically in the ISI databases, with first the organization and then after a comma the sub-organization (department or faculty) as a second subfield. If the name or the organization is missing, however, the sub-organization moves to the first subfield. However, a computer program cannot evaluate these differences. Thus, we used the first subfield, but always in combination with the city and country names.

Some organizations are dispersed over various address (such as the Catholic University of Louvain mentioned above), but in other cases these different addresses host relatively independent organizations. For example, the *Consejo Superior de Investigaciones Científicas* (CSIC) is housed at various locations in Madrid, but also elsewhere in Spain. In our data, we found 19 records with addresses in Valencia, Sevilla, and Burgassot. In Valencia, however, this same abbreviation (CSIC) is subsumed under the Universita Polytechnica of Valencia.

In summary, the different addresses can be meaningful or not, and this cannot be decided automatically, but depends on the research question. The program Inst2.Exe therefore offers the option not to aggregate into a single institutional name. For most purposes, however, the results of these fine-grained analyses may contain considerable error.

Furthermore, inst2.exe currently cannot distinguish between different locations of the same institutional name in terms of the network links. This can be further improved in the future by incorporating also into inst1.exe the option to disaggregate single institutional names in terms of different street addresses.

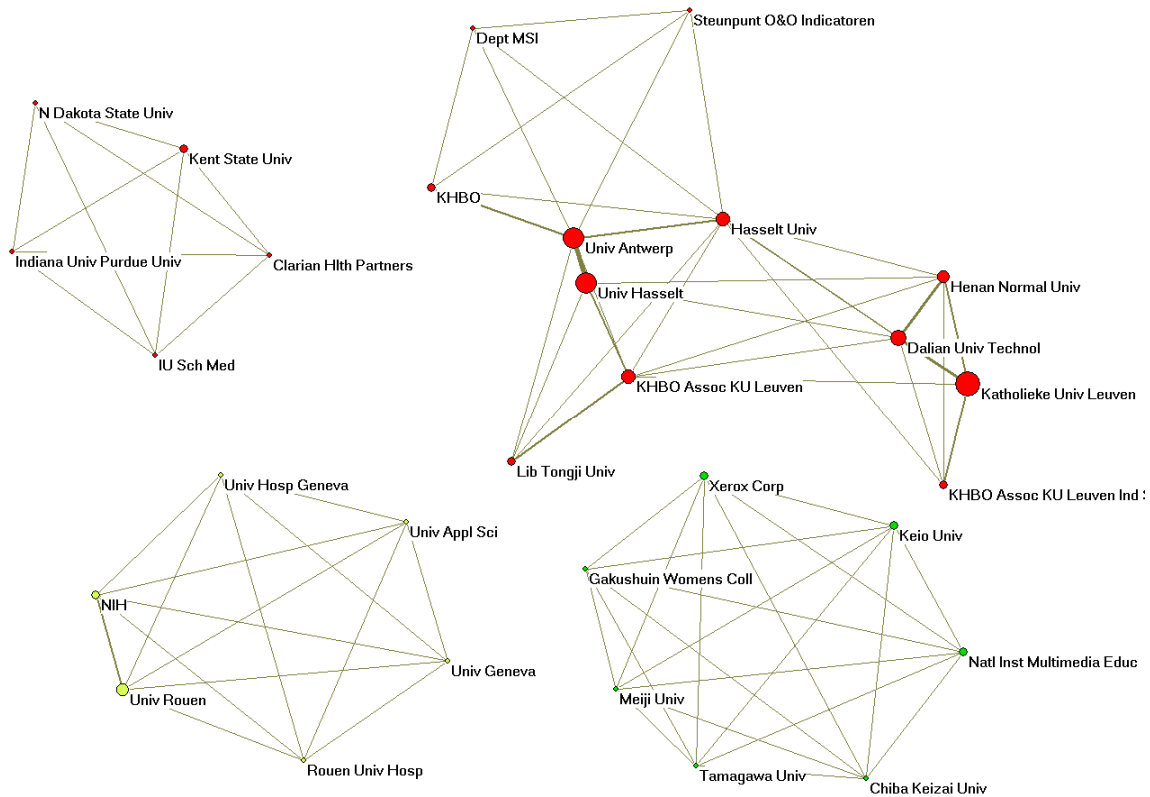


Figure 8: $k > 4$ networks of collaboration between leading institutes in the field of information science in 2009.

The (in this case, aggregated) institutional names provide us with a different view of the core network among these centers than was achieved above in terms of city names.

Figure 8 shows a highly connected network ($k = 6$ among 7 partners) of Japanese centers and the Xerox corporation. The figure illustrates the problem of the various institutional

names in the Belgian/Chinese network at the top ($k = 4$). The size of the nodes is again proportionate to the logarithm of the number of papers plus one (in order to prevent a zero as the evaluation of $\log(1)$). Figure 8 also demonstrates the effects of the noted policies of the Flemish government and the lack of standardization in the naming of institutions.

8. *Scopus* data

The problems with the institutional identification made us turn to *Scopus* for a comparison. Unlike the ISI databases, *Scopus* is based on index keys, and one might hope that this would make a difference for the standardization. However, in this database institutional names are even less standardized than in the ISI data: city names and even (technically necessary) delimiters are sometimes missing.²⁸ Using the 551 articles which could be retrieved with the equivalent search string, we found, for example, the three name variants “KU Leuven,” “KULeuven,” and “Katholieke Universiteit Leuven” among twenty records.²⁹ More seriously, two nodes in the Belgian network (Dalian and Xinxiang) were attributed to addresses in Taiwan according to this database (Liang & Rousseau, 2009). The geo-coder, however, recognized this as a mistake and was able to make the correction automatically.

²⁸ A referee noted that if one has access to raw *Scopus* data in XML, affiliation IDs are available which seem to merge the institutional name variations to a large degree. Unfortunately, the affiliation IDs are not provided in the online data.

²⁹ As city names, these records in *Scopus* contained “Leuven,” “Leuven (Heverlee),” and “B-3000”—that is, the postcode without mentioning the city. Thus, making the present routines fit for *Scopus* data would require another round of careful parsing of this data.

Nevertheless, the city networks using *Scopus* data are highly comparable with those based on the ISI set. As can be expected, the nodes are larger because of the larger coverage of the *Scopus* database. The respective files are available at <http://www.leydesdorff.net/maps/scopus.kml> for Google Earth, <http://www.leydesdorff.net/maps/scopus2.kml> for Google Maps, <http://www.leydesdorff.net/maps/scopus.html> using the GPS Visualizer, and <http://www.leydesdorff.net/maps/scopus.paj> for Pajek. The institutional networks suffer from the same problems with inconsistent naming by authors, which hitherto has been beyond the control of the database providers, and therefore *a fortiori* for users without building extensive thesauri.

9. Conclusions and discussion

We have wished to show the current possibilities available to the bibliometric researcher for the visualization of geographic data, and hopefully have provided some help by developing dedicated software to bridge existing gaps between using on the one hand databases like the *Science Citation Index* and *Scopus*, and on the other hand the geographical projections in Google Earth, Google Map, and Pajek. (The various processing steps are summarized in an Appendix.) It seems to us that for scholarly purposes, the options in Pajek are very rich and sufficiently beautiful for illustrations. Furthermore, the data in the Pajek format can be read into a large number of available software programs; for example, at the Network Bench (NWB) of Indiana University (at

<http://nwb.slis.indiana.edu/>).³⁰ Interfaces with animation programs—for time-series—are also available. However, the interface of the GPS Visualizer is superior for the presentation. As noted, one can also use results in Pajek for coloring the nodes at the Internet. By editing the html of the GPS Visualizer, one can also generate animations at the Internet.

At the Google interfaces, one can import the complete dataset (as kml- or kmz-files) into Google Earth, but the limitations are inherent to the satellite projection. Thus, one cannot draw a global map and one has no access to a street map. The same files can be read into Google Map. In that case, one has the full scale of projections and the network, but the nodes cannot be scaled. Using GPS Visualizer, one can scale the nodes and weigh the links. The view options are different between Google Earth and the GPS Visualizer.

Which one of these options one wishes to use will depend on one's research question. This contribution was primarily methodological. In addition to network analysis, one can think, for example, of studies about diffusion and about correlations between distances and relations (Andersson & Persson, 1993; Katz, 1994; Wuchty *et al.*, 2007). Geo-coordinates can be translated into distances using, for example, the calculator available at www.gpswaypoints.co.za/downloads/distcalc.xls.

The case in this study was selected so that the results would be recognizable in terms of flaws by this community. For example, further standardization of address information in

³⁰ The group at Indiana University developed also the Sci2 Tool (available at <http://sci.slis.indiana.edu/sci2>) that includes geographic visualization capabilities (e.g., Zoss *et al.*, 2010).

bylines seems highly desirable, particularly at the institutional level. City names are currently sufficiently standardized (because of postcodes) for research purposes.

The results further clarify that co-authorship, co-location, collaboration, etc., are all different dimensions in the scientific enterprise that may or may not overlap (Katz & Martin, 1997; Wagner, 2008). The relatively new tendency to add more than a single university address to each author (Persson *et al.*, 2004) further complicates the issue, as we showed for the Belgian case. By making these tools available, we hope to encourage other information scientists to use them in a further proliferation of research questions.

Acknowledgement

We are grateful to Wouter de Nooy and Chaomei Chen for advice and suggestions.

References

- Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirement for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient. *Journal of the American Society for Information Science and Technology*, 54(6), 550-560.
- Andersson, Å., & Persson, O. (1993). Networking scientists. *The Annals of Regional Science*, 27(1), 11-21.
- Borgatti, S. P. (1998). Social Network Analysis Instructional Website, at <http://www.analytictech.com/networks/mds.htm>.
- Börner, K., Penumarthy, S., Meiss, M., & Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major US research institutions. *Scientometrics*, 68(3), 415-426.
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the Backbone of Science. *Scientometrics*, 64(3), 351-374.
- Boyack, K., Börner, K., & Klavans, R. (2007). Mapping the Structure and Evolution of Chemistry Research. *Proceedings of the 11th International Conference of Scientometrics and Informetrics*, D. Torres-Salinas & H. Moed (Eds.), Vol. 1, pp. 112-123, CSIC, Madrid, 21-25 June 2007.

- Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From Translations to Problematic Networks: An Introduction to Co-word Analysis. *Social Science Information* 22, 191-235.
- Chen, C. (2003). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. London: Springer.
- Chen, C., Cribbin, T., Macredie, R., & Morar, S. (2002). Visualizing and tracking the growth of competing paradigms: Two case studies. *Journal of the American Society for Information Science and Technology*, 53(8), 678-689.
- Costas, R., & Iribarren-Maestro, I. (2007). Variations in content and format of ISI databases in their different versions: The case of the Science Citation Index in CD-ROM and the Web of Science. *Scientometrics*, 72(2), 167-183.
- De Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Muñoz-Fernández, F. J., & Herrero-Solana, V. (2007). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology*, 58(14), 2167-2179.
- De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek*. New York: Cambridge University Press.
- Debackere, K., & Glänzel, W. (2004). Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics*, 59(2), 253-276.
- Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: towards a cumulative research program. *Journal of Informetrics*.
- Fruchterman, T., & Reingold, E. (1991). Graph drawing by force-directed replacement. *Software--Practice and Experience*, 21, 1129-1166.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69-115.
- Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside; at <http://faculty.ucr.edu/~hanneman/nettext/>.
- Hicks, D., & Katz, J. S. (1996). Science policy for a highly collaborative science system. *Science and Public Policy*, 23(1), 39-44.
- Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: shifting impact, geography, and stratification in science. *Science*, 322(5905), 1259-1262.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7-15.
- Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1-18.
- Klavans, R., & Boyack, K. W. (2009a). Identifying Distinctive Competencies in Science. *Journal of Higher Education*, under submission.
- Klavans, R., & Boyack, K. (2009b). Towards a Consensus Map of Science. *Journal of the American Society for Information Science and Technology*, 60(3), 455-476.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage Publications.
- Leydesdorff, L. (1986). The Development of Frames of References. *Scientometrics* 9, 103-125.

- Leydesdorff, L., & Schank, T. (2008). Dynamic Animations of Journal Maps: Indicators of Structural Change and Interdisciplinary Developments. *Journal of the American Society for Information Science and Technology*, 59(11), 1810-1818.
- Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4), 317-325.
- Leydesdorff, L., & Wagner, C. S. (2009). Macro-level indicators of the relations between research funding and research output. *Journal of Informetrics*, 3(4), 353-362.
- Liang, L., & Rousseau, R. (2009). Bibliometric characteristics of the journal Science: Pre-Koshland, Koshland and post-Koshland period. *Scientometrics*, 80(2), 359-372.
- Narin, F. (1976). *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*. Washington, DC: National Science Foundation.
- Narin, F., Carpenter, M., & Berlt, N. C. (1972). Interrelationships of Scientific Journals. *Journal of the American Society for Information Science*, 23, 323-331.
- Narin, F., & Carpenter, M. P. (1975). National Publication and Citation Comparisons. *Journal of the American Society of Information Science*, 26, 80-93.
- National Science Board. (2010). *Science and Engineering Indicators*. Washington DC: National Science Foundation; available at <http://www.nsf.gov/statistics/seind10/>.
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421-432.
- Phan, D., Xiao, L., Yeh, R., Hanrahan, P., & Winograd, T. (2005). *Flow map layout*.
- Rafols, I., & Leydesdorff, L. (2009). Content-based and Algorithmic Classifications of Journals: Perspectives on the Dynamics of Scientific Communication and Indexer Effects *Journal of the American Society for Information Science and Technology*, 60(9), 1823-1835.
- Rafols, I., Porter, A., & Leydesdorff, L. (in preparation). Science overlay maps: a new tool for research policy and library management.
- Skolnikoff, E. B. (1993). *The Elusive Transformation: science, technology and the evolution of international politics*. Princeton, NJ: Princeton University Press.
- Small, H. (1999). Visualizing Science by Citation Mapping. *Journal of the American Society for Information Science*, 50(9), 799-813.
- Small, H., & Garfield, E. (1985). The geography of science: disciplinary and national mappings. *Journal of Information Science*, 11, 147-159.
- Small, H., & Griffith, B. (1974). The Structure of Scientific Literature I. *Science Studies* 4, 17-40.
- Tijssen, R., de Leeuw, J., & van Raan, A. F. J. (1987). Quasi-Correspondence Analysis on Square Scientometric Transaction Matrices. *Scientometrics* 11, 347-361.
- Tobler, W. R. (1987). Experiments in migration mapping by computer. *Cartography and Geographic Information Science*, 14(2), 155-163.
- Van Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635-1651.
- Wagner, C. S. (2008). *The New Invisible College*. Washington, DC: Brookings Press.

- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036-1039.
- Zhu, X., Hou, J., & Chen, Y. (2008). *Distribution and research fronts of international energy technology*. Paper presented at the Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, Berlin, August 2008.
- Zitt, M., Barré, R., Sigogneau, A., & Laville, F. (1999). Territorial concentration and evolution of science and technology activities in the European Union: a descriptive analysis. *Research Policy*, 28(5), 545-562.
- Zoss, A. M., Conover, M., & Börner, K. Where are the Academic Jobs? Interactive Exploration of Job Advertisements in Geospatial and Topical Space. In S.-K. Chai & J. Salerno (Eds.), *2010 International Conference on Social Computing, Behavioral Modeling and Prediction (SBP10)*. Bethesda, MD: Springer.

Appendix 1: Overview of routines for the data processing.

<i>Scopus data</i> ↑ → Scop2Isi.Exe ³¹	<i>Web-of-Science data</i> (in tagged format)	Output		
		(1) kml-files	(2) html	(3) paj-files
→	→	Cities1.Exe → Cities.txt	Inp_gps.txt	Cities.paj
		Inst1.Exe → Inst.txt		
		Geo-coding ³²		
		Cities2.Exe		
		Inst2.Exe		
		Cities.kml Cities2.kml Inst.kml		Inst.paj
		↓	↓	↓
		Matrix.txt	Input to GPS Visualizer ³³	Merge with Coast.net ³⁴ within Pajek
		→		
		Possible shortcut to make co-occurrence matrix in Pajek		
		↓		
		Paj2Cooc.Exe	Edit the html (api-key)	
		1. Use with Google Earth		
		2. Upload for Google Maps		
		3. Use at http://display-kml.appspot.com/		

³¹ Available at <http://www.leydesdorff.net/software/scop2isi>

³² Available at <http://www.gpsvisualizer.com/geocoder/>

³³ Available at http://www.gpsvisualizer.com/map_input?form=data

³⁴ Available at <http://www.leydesdorff.net/maps/coast.zip>