Mapping connections

Hands-on workshop on Network Analsyis

Carmen Cabrera-Arnau
University of Liverpool

1 Introduction

This document will be your guide for the hands-on part of the Workshop "Mapping Connections", where the aim is to bring you closer to the world of Network Theory and its applications to geography and spatial analysis.

To do this, we will focus on producing some nice network visualisations as well as computing some of the network properties or metrics that we introduced in the first part of the Workshop during the mini-lecture. To do this, we will use Gephi, which is a powerful network analysis software tool, free to download and quite easy to use as it does not require users to write computer code.

This part of the tutorial follows online materials created by Martin Grandjean [1] and accessible on the site https://www.martingrandjean.ch/gephi-introduction/.

2 Set up

2.1 Downloading and installing Gephi

Gephi has already been installed in the University of Liverpool computers, so you do not have to worry about that step. However, if you want to keep playing with Gephi at home, you can do so by downloading the software from https://gephi.org/users/download/ and by following the steps for installation. Just make sure you download the version of Gephi corresponding to your operating system (e.g. Windows, MacOs).

Since the focus is on the applications of network analysis in geography, you will need a domain-specific plug-in, which you can think of as software expansion that allows you to go beyond the basic functionalities of Gephi. This plug-in is called GeoLayout and may have already been installed in the University of Liverpool computers. However, if it hasn't or you want to test it elsewhere, you can go to Tools > Plugins > Available Plugins. From the list of Available Plugins, select GeoLayout and click install.

2.2 Downloading and installing Inkscape

Inkscape has also been installed in the University of Liverpool computers, so again, you do not have to worry about this step during the Workshop. If you want to continue your Network Analysis exploration elsewhere, you can download Inkscape by visiting https://inkscape.org/. Then navigate to Downloads, and download the latest version of this software for your operating system.

Inkscape is a free and open-source vector graphics editor, which offers a rich set of features and is widely used for both artistic and technical illustrations such as cartoons, clip art, logos, typography, diagramming and flowcharting. It uses vector graphics to allow for sharp printouts and renderings at unlimited resolution. Inkscape uses the .svg file format as its main format, which is supported by many other applications including web browsers. In the context of this workshop, you will use

Inkscape to include a basemap to the network visualisation generated with Gephi.

2.3 Datasets

The aim of the hands-on workshop is to explore mail exchanges over Europe, like in [1]. We will conceptualise these communications via a network, where the nodes represent different European cities. Two cities will be connected via an edge if there are letters sent between one and the other. Each edge will include a weight representing the number of letters sent between a certain pair of cities.

If using the University of Liverpool computers, you should be able to find the necessary datasets in M: Geography Workshop, so there is no need to download them.

Following the link https://github.com/CrmnCA/nets4Y12, you can download the data for the nodes and edges of the network. All the data needed to generate the network is in the files Nodes1.csv and Edges1.csv, but on top of that, you should also download the file Mapbase.svg, which is a background map that you will use to take your visualisation to the next level. **Please, save these three files on your Desktop** or another folder of choice, but make sure you remember where you put the files!

3 Importing data into Gephi

Start by opening Gephi on your computer. In the start window, select "New project".

3.1 Nodes

Go to "Data Laboratory" and select "Import spreadsheet". Navigate to the folder where you stored the network data, i.e. the files Nodes1.csv and Edges1.csv.

Start by opening Nodes1.csv. In the next window, make sure you select "Separator: Semicolon", "Import as: Nodes table" and "Charaset: UTF-8", like in Figure 1.

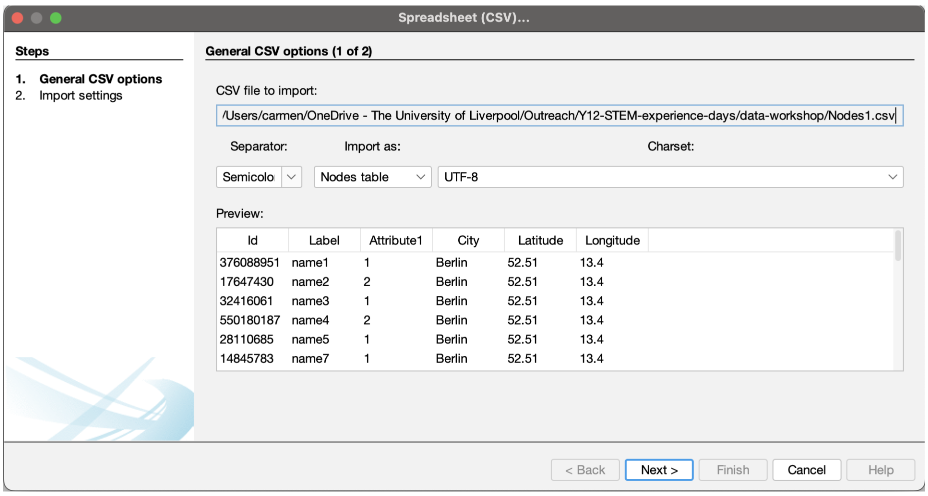

Figure 1: Importing Nodes1.csv.

Then press “Next” and ensure that the imported columns are Id, Label, Attribute1, City, Latitude and Longitude. Attribute1, City, Latitude and Longitude must be imported as Integer, String, Double and Double respectively. Press “Finish”.

You will be prompted with another window. Select “Graph Type: Directed”. And select the “Append to existing workspace” option, like in Figure 2. Then, “OK”.

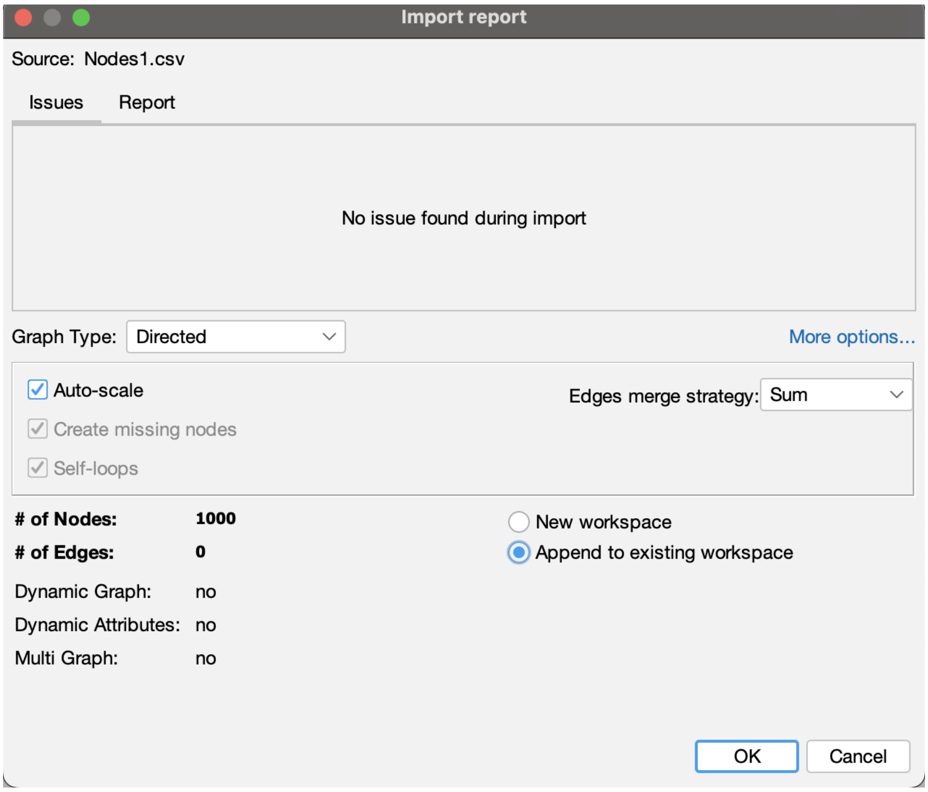

Figure 2: Importing Nodes1.csv. It is important to select “Append to existing workspace”.

3.2 Edges

To import the edges, once again go to “Data Laboratory” and select “Import spreadsheet”. Navigate to the folder where you stored the network data and open Edges1.csv.

Select “Separator: Semicolon”, “Import as: Edges table” and “Charaset: UTF-8”. Ensure time representation is “Intervals”. And the imported columns are Type and Weight, with Weight imported as Double.

In the next window, select “Graph Type: Directed” and select the “Append to existing workspace” option, then, “OK”.

4 Network visualisation

To work on the visualisation, let’s switch to the “Overview” panel. Gephi produces an overview of the network, which initially, it is randomly spatialised and as you will probably tell, completely un-

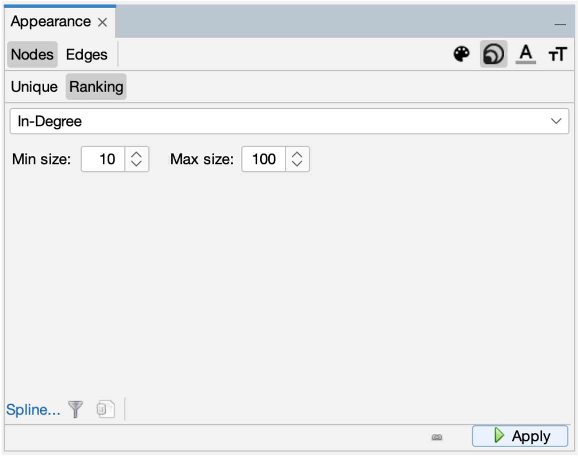

Figure 3: Caption

readable. Your task is to make it into something nice, that conveys as much information as possible.

4.1 Choosing node size

You can start by giving the nodes a size proportional to their degree or number of connections. Since you are working with a directed network, this will actually be the in-degree or out-degree. In the “Appearance” panel, select “Nodes” and the “Size” icon on the top right (the icon with several circles of different radii). Then, select “Ranking: In-Degree”. Select 10 and 100 as the minimum and maximum sizes. This option will scale the size of the nodes so that the ones with the minimum and maximum in-degree have radius 10 and 100 respectively, and all the rest will be somewhere in between. Press “Apply”. See Figure 3.

4.2 Spatialisation

As it stands, our network is a messy blob of nodes, randomly spatialised. However, this can be easily resolved with a few easy steps.

You can start by trying a spatialisation that gives some more room to the network, but keeping it in a bounded area. This will be achieved via spatialisation algorithm called Fruchterman Reingold. To apply this algorithm, go to the “Layout” panel, and select the layout with this name. Give the value 20,000.0 to area and 10 to gravity and speed. Press “Run”. Let the algorithm run for a bit until the visualisation stabilises. This algorithm is arranging the nodes so that the more densely connected parts of the network (i.e. clusters of nodes with lots of edges between them) are close together. Clusters that are connected by just a few edges, are further apart. Use the little blue magnifying glass (bottom left of the graph panel) to re-center the zoom.

Following Furchterman Reingold, you can use the Force Atlas 2 algorithm, which will have the effect of dispersing groups and giving space around larger nodes. The parameters you enter before running the algorithm significantly alter the final appearance. You are advised to select 50 for scaling and tick the box for “prevent overlap”. Let the function run until the network is mostly stabilised. Once again, use the blue magnifying glass to re-center the zoom if you need to. In principle, you could apply Force Atlas 2 directly without having applied Fruchterman Reingold, but as the layout from the initial network is random, it is generally better to disentangle the network beforehand.

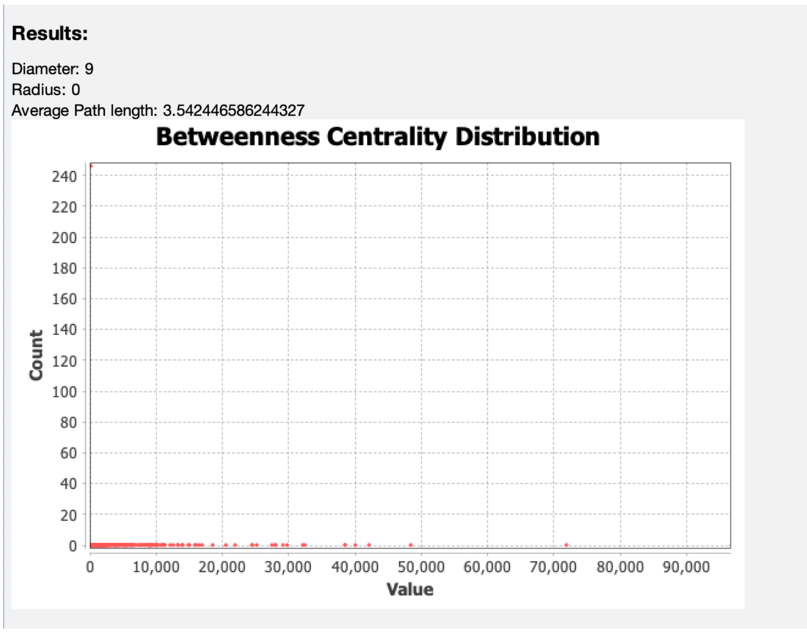

Figure 4: Betweenness centrality distribution.

4.3 Visualisation 1 – by betweenness centrality

You can add even more information to your network visualisation by computing some additional network properties and representing this, for example, through the color of the nodes. For example, you can color the nodes according to their betweenness centrality. Remember that the betweenness centrality of a node measures all the shortest paths between every pair of nodes of the network and then counts how many times a node is on a shortest path between two others. Therefore, the betweenness centrality gives an indication of the importance of a node in ensuring that the network stays connected (if a node with high betweenness centrality was removed from the network, the network could break into isolated components!).

In Gephi, it is very easy to compute the betweenness centrality. In the "Statistics" panel, go to "Network Diameter" and click run. A window will appear on the screen, with a few options for the type of network. Click "Directed", leave the other boxes unticked, then press "OK". A new window titlted "Graph Distance Report" will be shown, with several plots of different centrality measures. You can focus on the betweenness centrality distribution, like in Figure 4. This plot represents on the x-axis the different values of betweenness centrality and on the y-axis, the count or frequency of nodes that have a certain value of betweenness centrality. You can see from the plot that many nodes (more than 240!) have a betweenness centrality of 0. This means the nodes are isolated. The rest of nodes have varying values of betweenness centrality, and the maximum value is somewhere between 70,000 and 80,000.

Now that theses values are calculated, the betweenness centrality becomes available in the appearance panel to tweak the color of the nodes accordingly. Select the "color" icon (the palette), and chose "betweenness centrality" to color nodes according to this property. As a visual tip, you might get better results by clicking on the icon that is right of the color bar and selecting invert. Then hover over the color bar and bring the middle pointer towards the left, do the same with the right-most pointer. You can check Figure 5 for guidance. This way, nodes with low betweenness centrality will be darker and the highly connected nodes will be lighter and more visible.
You can now find way to highlight nodes that have a high betweenness centrality by tweaking the node colour. It quickly appear that nodes with a high degree/weighted degree does not always have

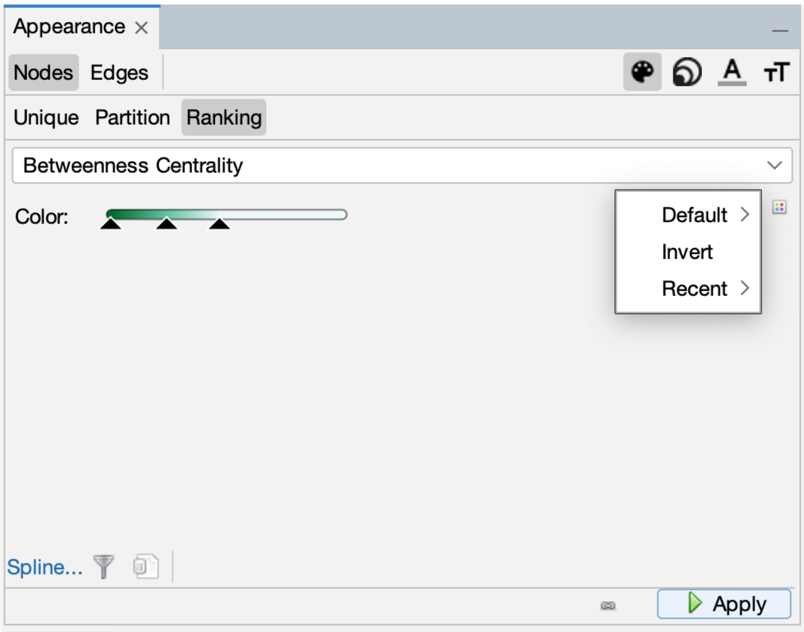

Figure 5: Colouring nodes according to their betweenness centrality.

a high betweenness.
To finalise the visualisation, go to the "Preview" panel. Unlike during previous stages, changing settings in this menu is reversible, and do not affect the structure of the graph. For good visual results, you are recommended to set the edges opacity to 70%. Be aware that due to its large size, the netowrk may take a few seconds to update after each change. You will need to click on "refresh" to apply the changes. Tick the box for curved edges. As a convention, curved edges are used to show the direction of the edge, always turned clockwise. Non-curved edges are generally undirected graphs.

And there you have, a pretty graph that should look like that in Figure **??**. You can save the image by exporting it. To export, press the "Export" button at the bottom left. Choose the format. Note that exporting in .png produces figure with worse resolution. You may want to opt for .svg or .pdf, which have the advantage of being modifiable by your own drawing software.

4.4 Visualisation 2 – by modularity and with geographical layout

Go back to the preview menu. You will try another visualisation, this time taking the role of geography a lot more seriously!

A network contains internal groupings called communities. There are algorithms that are able to detect these communities, based on comparing the density of edges between a sub-group of nodes compared to the the overall density of the network. You can compute modularity in Gephi by going to the "Statistics" panel and running "Modularity". When asked, choose a resolution of 0.5 and click "OK".

Next, go to the "Appearance" panel, choose the node-colouring icon and in the "Partition" menu choose "Modularity class". Then press apply. You will be then able to modify the colors attributed to the detected communities by clicking on them if you wish to.

During the import, you've noticed that every node had a Latitude and a Longitude. You can now change the network layout to represent the location of the nodes according to their geographical

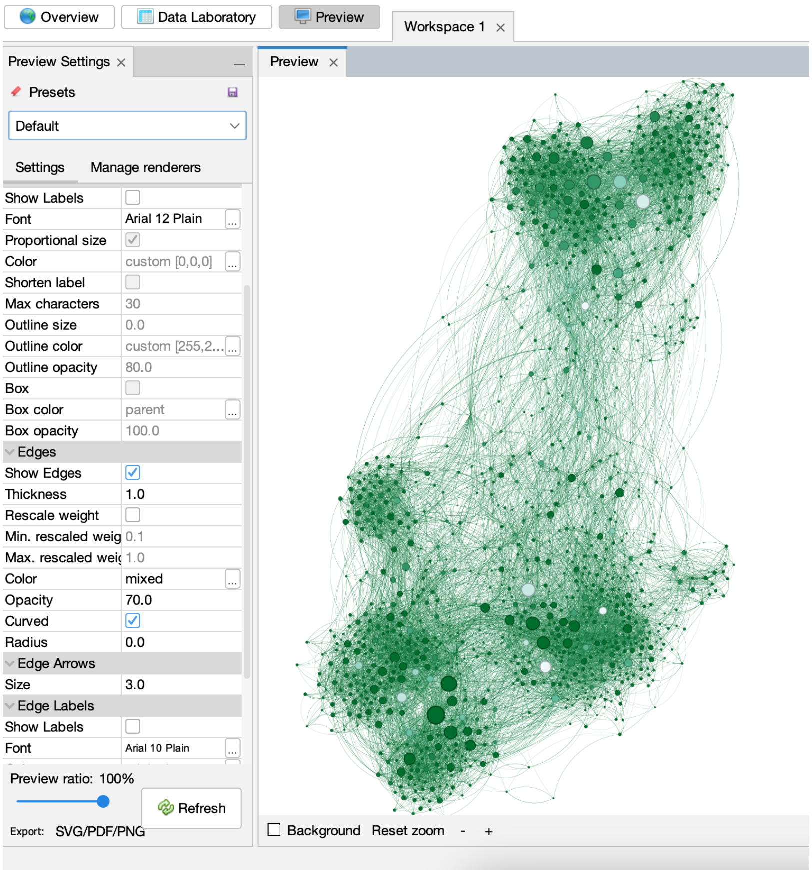

Figure 6: Result of visualisation by betweenness centrality

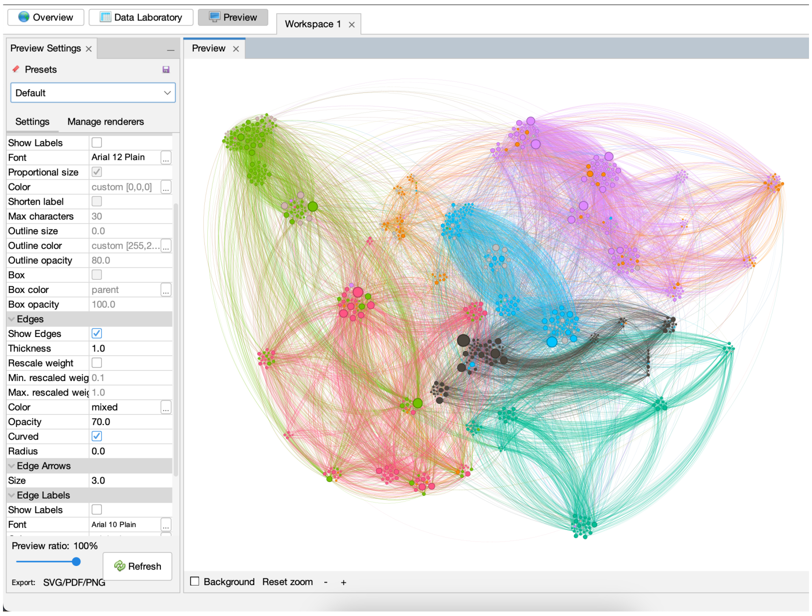

Figure 7: Result of visualisation by modularity and with geographical layout.

coordinates. In the “Layout” panel, select Geo Layout and give it a scale of 20,000. Be sure to select “latitude” and “longitude” for the “Latitude” and “Longitude” options, and set the projection to “Mercator” (this is a map projection that will make sure the scale of the node locations fits with the background map you will use later). Press “Apply”. As several nodes are now collapsed into the same a geographical coordinate, you will have to give them some extra room to make them visible. To do this, use the Noverlap layout, setting a margin of 5.0 and a speed of 0.1. Run the algorithm until the nodes are a little more spaced out and therefore visible.

Like before, go to “Overview” to refine your visualisation. Set the edge opacity to 70 and make sure the edges are curved. Your network should look very similar to that in Figure 7. Press refresh and export the image as “Network.svg”. Remember where you save your resulting image, as you will need this information for the final and most exciting step! You can now close Gephi.

Finally, open a new document on Inkscape. Go to “File”, “Import...” and choose the file Map-base.svg that you downloaded at the beginning of the practical session. To crop the image around the map, select the “Create rectangles and squares” tool on the left menu (it is the square icon, just above the circle icon). Draw a square around the map and make sure the square is selected. Go to “Edit”, then “Resize page to selection”. Select the square you just drew with the “Select and transform object” tool (it is the arrow icon, two above from the square icon). Right click and select “Delete”.

Next, import the “Network.svg” file that you created on Gephi. Adjust the size so that it matches the background map. Make sure the layer with the names of the cities is on top of all the others. And that’s your final result! It should look similar to Figure 8.

References

[1] Grandjean, M.: GEPHI – Introduction to Network Analysis and Visualization. https://www.martingrandjean.ch/gephi-introduction (2105)

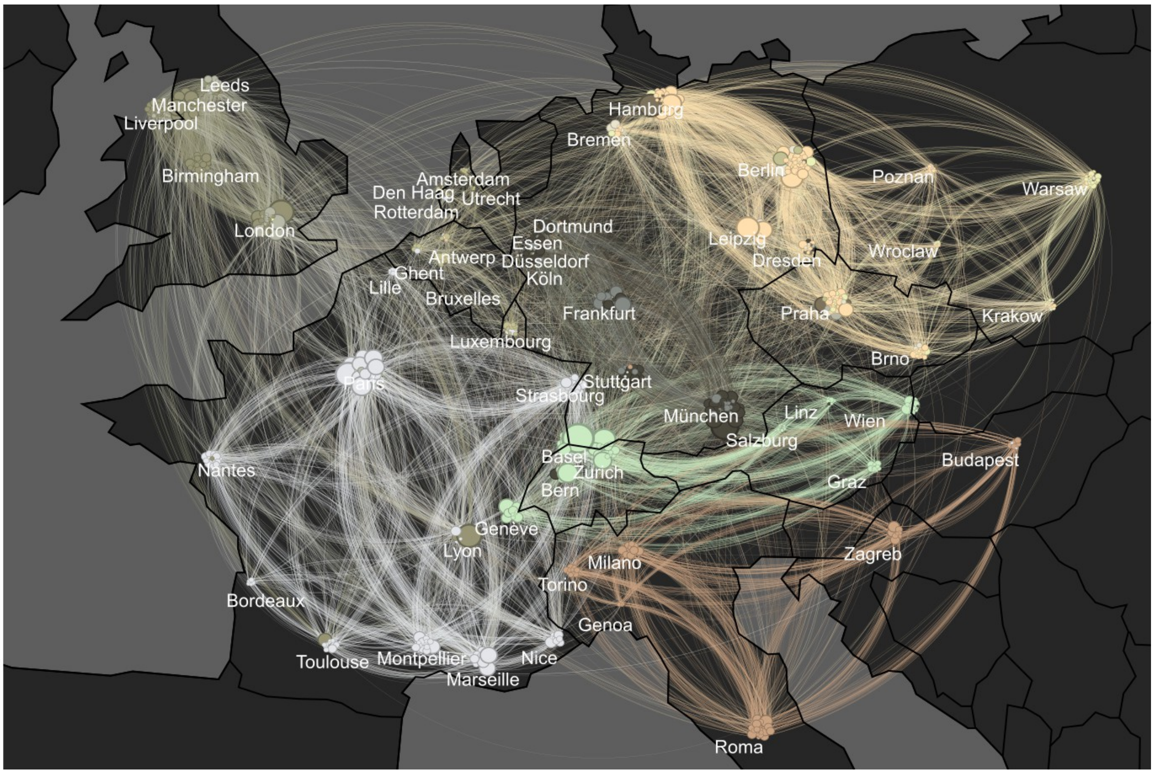

Figure 8: Result of visualisation, with base map.