Visual Exploration of (French) Commuter Networks

Yves Chiricota, Guy Melançon, Trung Tien Phan Quang, Patrice Tissandier

Visual exploration of (French) commuter networks

Yves Chiricota
Universite du Quebec a Chicoutimi, Canada
Yves_Chiricota@uqac.ca

Guy Melancon, Trung Tien Phan Quang,
Patrice Tissandier
CNRS UMR 5800 LaBRI & INRIA Bordeaux – Sud-Ouest, France
{melancon, phanquan}@labri.fr

1. INTRODUCTION

Although inhabitants of greater cities generally work in the same town they live in, most non urban citizen commute to find work out of their hometown. As a consequence, governments must plan transportation and territorial accessibility accordingly, turning commuters into a first order subject of study for geographers. Commuters induce a network structure on the territory linking cities. Some cities appear as attractors because they offer good job opportunities, others appear as worker broadcasters because of their suburban activity.

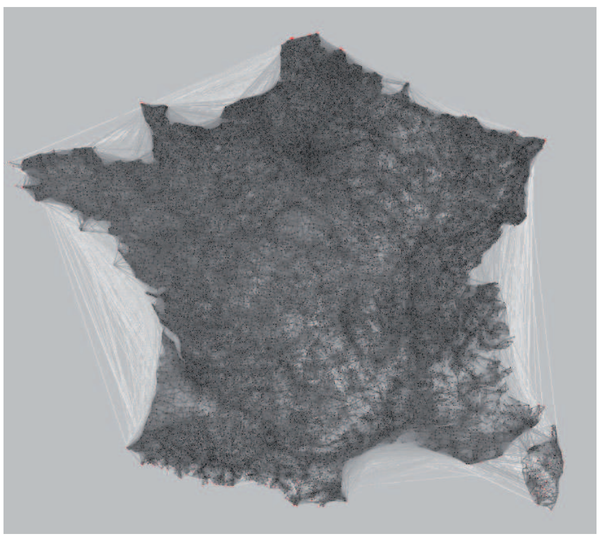

Figure 1: Commuter network over all of France (1982). The contour of the country results from the drawing of towns based on their geographic coordinates, while the density of links reflect the amount of commuting activity.

Figure 1 show the entire network of commuters formed over France in 1982. This network seen as a graph is a directed and weighted graph and the density of the image suggest the load of the underlying data formed of about 36 000 nodes (towns) and up to more than 2 million edges (connections between cities through commuters). The interplay of commuter moving from one city to another creates a complex dynamic between cities and towns. Geographers study commuter data trying to understand the structure of the network, as a contribution towards urban planning. The analysis of commuter networks helps answering natural questions such as:

- What cities act as “poles” implicitly defining attraction bassin (regions where people would be more inclined to move to) ?
- How does the network organize around these “poles” and can how we assess of a multilevel structure (“poles of poles”) ? Indeed, suburbs not only export workers in greater cities, but create jobs as well to fulfill the needs of their local population. Hence, workers commuting to a second order town actually contribute to the dynamic of the nearby city.

Most studies looking at commuter traffic have used simplified representations, such as egocentric networks concentrating on the exchange of workers going to the same city or region [3] [2] [1]. To our knowledge, no visualization has been designed in order to visualize the entire network of commuters [7] [6] [8]. Our approach combining the strength of visual analytics and graph combinatorics provides a novel approach to address the study of commuter traffic.

Our approach is more specifically designed to help geographers identify multiscale phenomenon, reflecting how the activity occurring at a local scale actually contribute to higher order dynamics. As suggested earlier, flows occurring between nearby suburbs may indeed be seen as local contribution of greater cities taking place at a national scale. Although our visualization and the analytical tools we have designed do not by themselves provide sound answers to these questions, they definitely contribute and help geographers when designing/validating hypothesis.

2. GEOMAPPED COMMUTER TRAFFIC

The questions listed in the introduction challenges visual analytics and requires to design a visualization helping geographers to grab the big picture, and at the same time be able to access to a more detailed view in the neighborhood of any specific city – instead of simply looking at salient properties of only a small sample of the whole dataset.

Although the data under study is described as a graph, its visualization through standard node-link graph drawing and clustering turns out to be a difficult task for several reasons. The large size of the commuter network is a challenge in itself. Our first attempts to identify a multiscale structure in the network through hierarchical graph clustering (and visualization) has been defeated by the size and complexity of the data. We opted instead for an approach taking advantage of the graphic processing unit (GPU), concentrating on edge filtering in an effort to "see" the inherent structure in the data. Edge filtering is implemented with the help of clipping (performed at the GPU level) against the view volume implemented in the 3D graphics pipeline.

First of all, nodes (towns) are mapped to their geographical coordinates. As a consequence, the relative position of nodes reproduces the overall shape of France, and in a lesser fashion of regions and departments – regions form larger administrative units gathering a set of departments[1]. As for edges (t_1, t_2) connecting two towns t_1 and t_2, the idea consists in drawing them in 3D using line segments whose (x, y) coordinates of endpoints correspond to geographical coordinates of towns t_1, t_2.

Now, the dataset comes equipped with several edge attributes such as the distance between towns. Another attribute we are given is the commuter traffic (the number of persons residing in town t_1 and working in town t_2) – observe that in the first case the graph can be considered undirected while in the second case we distinguish the edge (t_1, t_2) from the edge (t_2, t_1). Using one of these attributes, and other metrics as we shall see later, each segment is additionally assigned a z coordinate. An orthographic projection is then used to obtain the final 2D image on the screen, where the projection plane is perpendicular to z axis (see Figure 2).

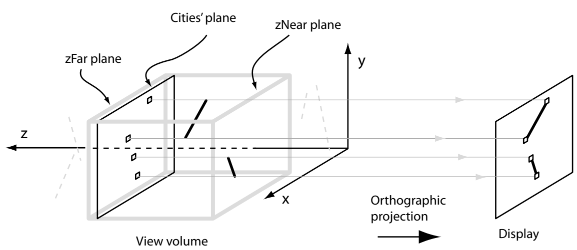

Figure 2: Visualization setup.

Line segments (edges) are drawn onto the plane through the z buffer, using alpha blending to simulate transparency to improve readability of the resulting 2D map. Using simple interactors, the user can filter edges by moving the $zNear$ and $zFar$ planes. It is also possible for the user to pan and zoom to get a closer view of a subregion on the map (Figure 3). On the application's side, this is accomplished by changing the viewport position in the projection plane. Thanks to current graphic adapters, all computations remain real-time and interaction is very fluid.

Towns are represented by small squares in a plane sitting slightly in front of the zFar plane, with color and size that can be mapped to any value. Positioning the town plane this way makes it sure it does not get clipped.

3. GRAPH ANALYTICS

Filtering edges based on attributes such as distance between cities or the raw number of persons commuting from one city to the other does not provide a clear idea on how the flows organize. Nowadays commuting does not mean

[1]Bordeaux is for instance part of the Gironde department, itself part of the Aquitaine region.

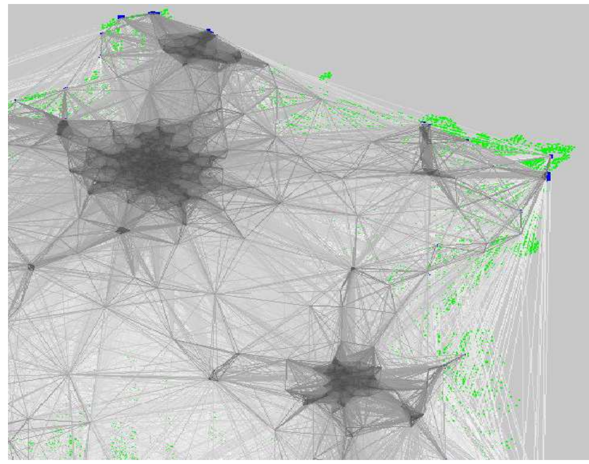

Figure 3: A closer view of the commuter network showing the effect of edge filtering, revealing lower scale structure organizing around greater cities (Paris, Lyon, Lille).

travelling small scale distances because many people take high speed trains, going from Marseille or Bordeaux in 3 hours for instance, or even fly. Moreover, distance does not bring in ingredients to analyze the importance of the flows taking place over the territory. It is the number of persons moving in and out of cities we wish to "see". However, the raw numbers themselves must be interpreted in terms of the relative mass of cities. A thousand people moving in and out everyday day might be a considerable subset of the population for a city like Bordeaux, although it is more or less unnoticeable for a city like Paris.

To this end, we borrowed and adapted a metric first introduced by Guimerà et al. in [5] in the context of the study of the worldwide air passenger traffic. Given an *undirected* graph $G = (V, E)$ and a clustering $C_1, \ldots, C_k$ of the graph (a partition of its set of vertices V), one can define the participation coefficient of a node $v \in V$ as:

$$p(v) = 1 - \sum_i \big(\frac{d_{C_i}(v)}{d_G(v)}\big)^2 \tag{1}$$

where $d_G(v)$ is the degree of v in G (number of neighbor nodes) and $d_{C_i}(v)$ is the number of neighbors of v belonging to C_i. The participation coefficient is null exactly when all neighbors of v belong to the same cluster C_i. On the opposite, $p(v)$ is maximum when its neighbors are equally distributed among clusters[2]. Based on this metric together with betweenness centrality on the entire network, they were able to classify cities as being either hubs, transit cities or simply acting as second order airport.

Obviously, the participation coefficient can be similarly defined for *directed* graphs, by restricting the degree to incoming or outgoing edges. That is, we define the *outgoing participation coefficient* by setting:

[2]This makes the participation coefficient quite analoguous to entropy as originally defined by Shannon [9] or Burt's hierarchy index [4].

$$p^+(v) = 1 - \sum_i \left(\frac{d^+_{C_i}(v)}{d^+_G(v)}\right)^2 \quad (2)$$

where $d^+_G(v)$ is the number of outgoing edges having v as source node, and $d^+_{C_i}(v)$ is the number of outgoing edges having v as source and having their target node in C_i. The *incoming participation coefficient* $p^-(v)$ is defined similarly. Hence, outgoing participation coefficient should reveal how much a town participates in sending workers around, or said differently, how much it acts as a residential area feeding other regions with workers. Conversely, the incoming participation coefficient should reveal how much a town, city or region act as an attractor, bringing workers into its local industry – all computations being performed relatively to the total flow of people going out of or coming in a town. For our purpose, we have considered two different clustering of the whole space, namely the division of France into departments or into regions.

Now, it turns out that one can define a participation coefficient of a node without any reference to a clustering of the graph, we shall call a *flat* outgoing participation coefficient. That is, we set:

$$\bar{p}^+(v) = 1 - \sum_{w \in N^+_G(v)} \frac{\omega(v,w)}{d^+_G(v)} \quad (3)$$

where $N^+_G(v)$ is the set of successors of v (neighbors that can be reached *from* v), and $\omega(v,w)$ is the weight (number of commuters) of the edge (v,w). The flat incoming participation coefficient is defined similarly (using $N^-_G(v)$ and $d^-_G(v)$ instead).

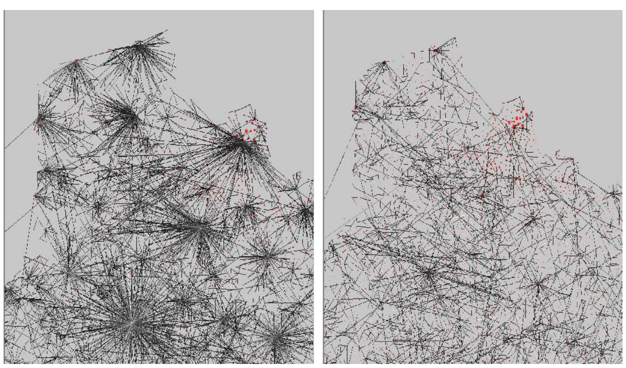

Figure 4: Edges filtered based either on their outgoing (left) or incoming (right) participation coefficient.

These two different coefficients provide distinct knowledge on the dynamics of the data. Figure 3 is obtained by filtering edges on the northern part of France. The left image reflects how outgoing flows organize, the star-like patterns clearly showing that some towns indeed provide significant flows of workers in their nearby territory. On the opposite, the right image does not show any regularity hence meaning that no town seems to act as a pole. That is, the workers emigrating from the star-like towns do not concentrate on any particular place in that region of France.

As another example, Figure 5 shows how the flat outgoing participation coefficient can be used to filter edges and explore the structure of the data. The image provides a close up view around Lyon (on the left part of the image). The edges shown here are those with a high coefficient, thus revealing high flows of workers leaving their hometown to work. Observe that these edges actually *do not* connect to Lyon but occur between suburbs. It is important to note that this observation *could not be captured based on an ego-centric approach* essentially looking at data directly related with Lyon itself (as in [2] or [1] for instance). What is found here is evidence that part of the activity around Lyon occurs at a lower scale, although taking place at a short distance form the city. On the contrary, observe on the right how Chambery experiences high flows taking place on longer distances towards the inner part of the country.

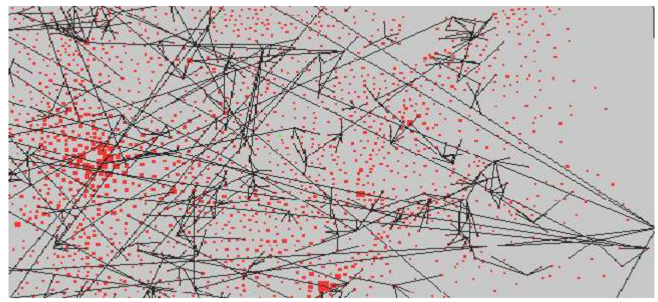

Figure 5: Flows of workers around Lyon (red spots forming a denser region on the left of the image).

Another issue of interest when studying commuter data is to study the activity occurring on a whole department (or region). The application allow this type of filtering, restricting edges going out (or coming into) a given department. Figure 6 shows how the flow of workers leaving the Gironde department distribute, with varying intensity (grayscale).

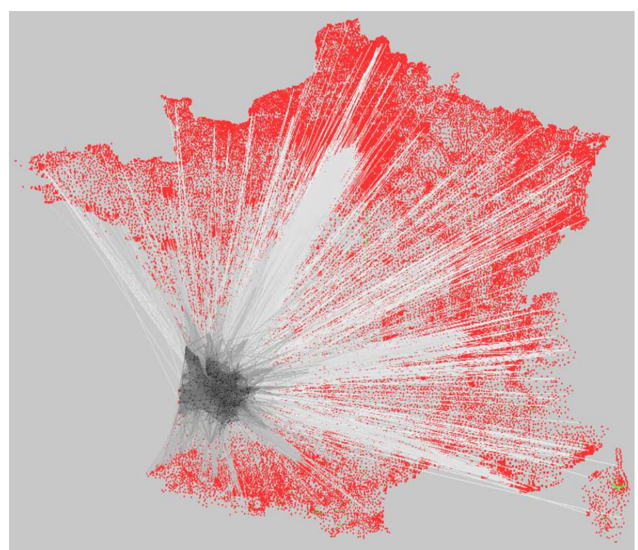

Figure 6: Flows of workers going out of Gironde.

As the image shows, the more intense flow take place at a local scale, although some workers travel the whole country going from Bordeaux to Lyon on a daily basis.

4. FUTURE WORK

The work presented in this abstract is ongoing.

Other variations of the participation coefficient can be used in order to sort out edges relatively to a department, a region, or any other cluster. As it stands, the participation coefficient, either Eq. (1) or Eq. (2), only provides information comparing edges incident to a given node. We are now investigating different metrics allowing the comparison of edges going out of a same department, for instance, thus taking the total flow emerging from a same cluster and looking at the relative weight of an edge with respect to the cluster it goes out from.

Indeed, it may well be that the number of people going from the Bordeaux region towards Paris compares with the number of people commuting between two small suburbs. Depending on the angle from which we look at the data, the flow going out of the suburbs may well appear as more important – which is what the actual participation coefficient assesses. From another angle, we might want to equate the two flows since they involve the same number of persons when looking at the department as a whole.

The interplay between the intensity of flows and their underlying physical distance is also an issue. The application should be enriched in order to allow filtering based on multiple dimensions. We plan to intensify the visual analysis of scatterplots, looking at possible correlations, but also driving the interaction from the scatterplots themselves through interactive brushing.

Finally, we wish to pursue both the analysis of the data and the development of our application in order to study the dynamics of commuter networks. Indeed, we do have access to the data covering four (French) national surveys (1975/1982/1990/1999), theoretically enabling us to address issues such as: can we identify places where flows experienced more changes over the years ? did new "poles" appear ? did some disappear, becoming simple actors of a subnetwork taking place at a larger scale ?

Also, animating the flows could potentially help in the visual exploration of such phenomenon. For now, the application allows the user to "play" an animation automatically piling up edges from lower to higher values. Although the animation reveals how poles builds from lower to larger scales, it still is relatively naive and straightforward.

Acknowledgements We wish to thank our colleague Céline Rozenblat, and more generally the whole SPANGEO group[3], for useful comments and discussions during the first stage of this work.

5. REFERENCES

[1] S. Berroir, T. Saint-Julien, and L. Sanders. Spécialisation fonctionnelle et mobilité : les pôles d'emploi de l'aire urbaine de paris. *Données Urbaines (Anthropos)*, 4:169–180, 2003.

[2] J. Binder, G. Haag, and G. Rabino. Analysis and modelling of commuter flows: application to the regions of stuttgart and turin. *Jahrbuch der Regionalwissenschaften*, 23:117–139, 2003.

[3] J. Bonnefoy. *Structures et fréquentations d'équipements, commerces et services en Languedoc-Roussillon.* Ph. D. Thesis, Université d'Aix en Provence, 1995.

[4] R. S. Burt. *Brokerage and Closure.* Oxford University Press, 2005.

[5] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7794–7799, 2005.

[6] J. Lévy and F. Dureau. *L'accès à la ville, les mobilités spatiales en question.* L'Harmattan, Paris, 2002.

[7] J. Orhan. Les déplacements quotidiens dans 50 villes. *Données Urbaines (Anthropos)*, 2:73–83, 1998.

[8] R. Patuelli, A. Reggiani, S. Gorman, P. Nijkamp, and F. Bade. Network analysis of commuting flows: A comparative static approach to german data. *Networks and Spatial Economics*, 7(4):315–331, 2007.

[9] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[3]See the URL `http://s4.parisgeo.cnrs.fr/spangeo/spangeo.htm`