Network Analysis of Commuting in France

I. Literature Review

In 2014, the French government introduced a bill destined to reorganize the local administrative structures. France indeed has numerous and redundant scales of administrative and political authority, making coordination and efficiency difficult [1].

The first element of the bill was a redrawing of the boundaries of the "Régions", rough equivalent to U.S. States, albeit much less independent and powerful. The rationale to reduce the number of regions from 22 to 13 was that existing Regions were too small, especially compared to German Länder [2], to play their role in economic development policies.

The second element was the establishment of "Metropolitan Areas", regrouping several cities to increase synergies. This was driven by the idea that economic growth is nowadays *de facto* driven by Metropolitan Areas [3], so that cooperation between cities within each Area should be reinforced. New Regions, containing usually two or three such Areas, should help diffuse the economic development to less central areas.

The decision of Cities to regroup with each other in Metropolitan Areas has been largely accomplished on the basis of existing cooperation structures, but many choices have been driven by the will to preserve the political influence (by staying away from a neighbor Metropolitan Areas around a larger Core City), rather than to reflect economic links. Analyzing the discrepancies between the new administrative boundaries and the actual economic network seems thus necessary to spot Metropolitan Areas which have been kept "artificially" separated, and should seek to cooperate on economic development matters.

In this paper, I perform a network analysis on one metric of economic interdependence between cities: commuting flows. The analysis of networks, stemming from Graph Theory, has spread in recent years to Social Sciences, as a tool to understand the structure of complex systems. Mobility is one of the fields lending itself nicely to this type of analysis, as shown by the recent studies on commuting between German Kreise [4], or the use of telecommunication data to understand mobility networks [5].

Section II. of this paper describes the data sources, and the structure of the network. In section III. I fit spatial models to the pattern of French commuting. Section IV presents the results of a community detection, and the comparison of the results with new Metropolitan Areas limits.

II. Data and Network Structure

1. Sources and Data Cleaning

The French National Institute for Statistics and Economic Studies (INSEE) computes 5-year rolling estimates of mobility patterns through the Census [6]. The most recent dataset is from 2010 and contains, for each pair of cities, the number of people living in one and working in the other. Only flows superior to 100 are reported. The dataset also contains the number of employed workers for each city.

For statistical purposes, the INSEE has defined "Urban Units" as a group of cities where no building is farther than 200m away from the closest [7]. Of the 9,398 "source" cities, 3,847 are in no Urban Unit. But those are mostly small cities, so that the number of commuters living in those cities is only 9.46% of the total number of commuters. Moreover, as the analysis is focused on Metropolitan Areas, I can ignore those cities outside Urban Units.

In terms of network definition, a node in this network is an Urban Unit, and there is an edge between two nodes if there are commuters between those units. Each edge is directed (links are not symmetrical), and is weighted by the number of commuters.

Finally, in order to fit some models to the network, I pulled distances and duration between Urban Units from the Google Maps API [8]. The distance or duration between two units is thus between the "centroids" of those Units, as defined by Google Maps. This is a first approximation, but washes away the more granular dynamics between large Urban Units and their periphery: workers who commute from the Northern suburbs of Paris to Paris probably work in the Northern half of Paris. Moreover, the distance and duration are computed for a person driving to work, which might be a problem in the Paris Area where a large share of workers commute by public transportation [9].

2. Network structure

The network is composed of 1,936 nodes and 3,908 edges. Of those edges, 3,373 are between different nodes, and 535 are "loop edges", counting the number of people living and working in the same Urban Unit. Unless specified, those edges are ignored, as a large isolated Urban Unit could look very connected because of its own population.

A node is first described by its degree. In case of a directed network, in-degree and out-degree are computed: the former is here the number of people commuting to the Urban Unit, and the latter the number of people commuting from the Urban Unit. The average in- and out-degree (by definition equal) is 1045, meaning that, on average, each existing pair of Urban Units is connected by 1045 commuters. But the in- and out-degrees have different distributions in our network [Fig 1]: the distribution of in-degrees is both more concentrated (the standard deviation is a third of the one for out-degrees [Table I]), but has more extreme values (the maximum value for in-degrees is four times the maximum out-degree value). This reveals the fact that some nodes are much more attractive than others. This concentration feature is also seen in the cumulative distribution of commuter flow [Fig. 2]: the 14.8% (500) biggest edges represent 47.2% of the share of total commuters.

Finally, we can try to rank the different nodes (Urban Unit) in terms of their centrality in the network. In presence of a mobility network, the pagerank centrality is often used [10], as it incorporates the fact that an area is more "central" if it is connected to other well-connected areas [Appendix A]. Unsurprisingly, the top ten nodes in terms of centrality are the cities considered as the cores of the future Metropolitan Areas [Table II]. The presence of Serris and Claye-Souilly, two very small suburbs of Paris, in the top ten, speaks however of the big premium given by the pagerank centrality to small units connected to very large ones. Moreover, the relationship between the pagerank of a node and its centrality score is best fitted by a power law [Fig. 4], but very imperfectly. Even the most suited centrality analysis presents serious shortcomings, calling for caution in its use for the analysis of commuting networks.

III. Spatial models

1. Gravity Model

A classical model, already estimated for France [11] stems from Isaac Newton's law of universal gravitation: each Urban Unit generates an attraction force proportional to its size (i.e its number of active workers), and the strength of the link between two Units decreases with the distance between the two. The fit of the model is rather imperfect, with a R-squared of only 28.8% [Table III]. I test some other specifications, with duration instead of distances, and only on cities of more than 40,000 inhabitants, without improving the fit [Table III]. All the results suggest that, when the distance between two locations doubles, the percentage of decrease in the flow of commuters is between 47% and 72%.

2. Gravity model with local normalization

Given the poor performance of the simple gravity model, I incorporate a local normalization to it, following the study on German Kreise [4]. This model [Appendix B] fits the variability of the data much better, with a R-squared of 58.9% [Table III]. This suggest that a model with local normalization is better suited for the analysis of the complete network of commuting (without having to slice it depending on the size of Urban Units). The estimated coefficient γ for distance is -0.57, meaning that the flow of commuters between two cities decrease less than proportionally with distance. This might be driven by a few cities far away but connected by high-speed rail (e.g. Paris and Marseille).

Table III: Regression results for $log(E_{ij}))\sim \alpha log(W_i) + \beta log(W_j)) + \gamma log(D_{ij})$[1]

Specification	γ	95% CI	R^2
(i) Gravity Model (distance)	-0.4744	[-0.516, -0.432]	0.288
(ii) Gravity Model (duration)	-0.5905	[-0.651, -0.530]	0.265
(iii) Gravity Model on cities of more than 40,000 workers (dist.)	-0.7268	[-0.858, -0.596]	0.267
(iv) Gravity Model with local normalization (distance)	-0.577	-	0.589

IV. Community Detection

There are several algorithms available to detect "communities" in a network. A community is defined as a group of nodes strongly connected between themselves, and (relatively) weakly connected to the rest of the network. I use the Louvain Method of community detection on the 1,936 Urban Units in the network, based on the commuting links

[1] With E_{ij} the flow between two units, W_i the number of active workers in unit i, and D_{ij} the distance between the two units

between them. The overall modularity score [Appendix C] is 0.947, which indicates a set of internally strongly connected communities (given that the score is bounded between -1 and 1). The superposition between the detected communities and some of the new Metropolitan Areas shows that, despite their size, and ambition, the new areas fall short to regrouping all the Urban Units within a "commuting community" [Map 1].

V. Conclusion

The commuting pattern in a country lends itself nicely to a network analysis, as it enables to describe the interaction between the numerous parts with simple concepts, without having to lose too much in generality by describing each very particular setting.

In the French case, professional mobility reveals a network with several well delineated communities. Each region seems to have a clear attractive core, although the imbalance between the core centers and the peripheries is less important than it would have been had the centrality rank followed a power law. The analysis also underlines the weight of Paris, whose score is three times the score of the second most central city. Despite decentralization efforts, the heritage of the monarchy and the centralized Republics is still visible.

Finally, it seems that the recent reform has not been able to create a level of government regrouping whole commuting sub-networks. This could be seen as a failure, but one must realize that, as seen in the case of Marseille, such an entity would have been bigger than the old Départements, a local government between cities and Régions. As the government had decided not to suppress them (for their necessity in rural areas and their long history), it was not possible to establish large enough Metropolitan Areas. Although this is understandable also for cooperation reasons (already difficult with the numerous cities), it obviously undermines the initial dual goal of simplifying the French administrative entities, and making them more in line with economic networks.

Tables & Figures

Fig. 1 - Distribution density of in- and out-degrees for nodes in the network

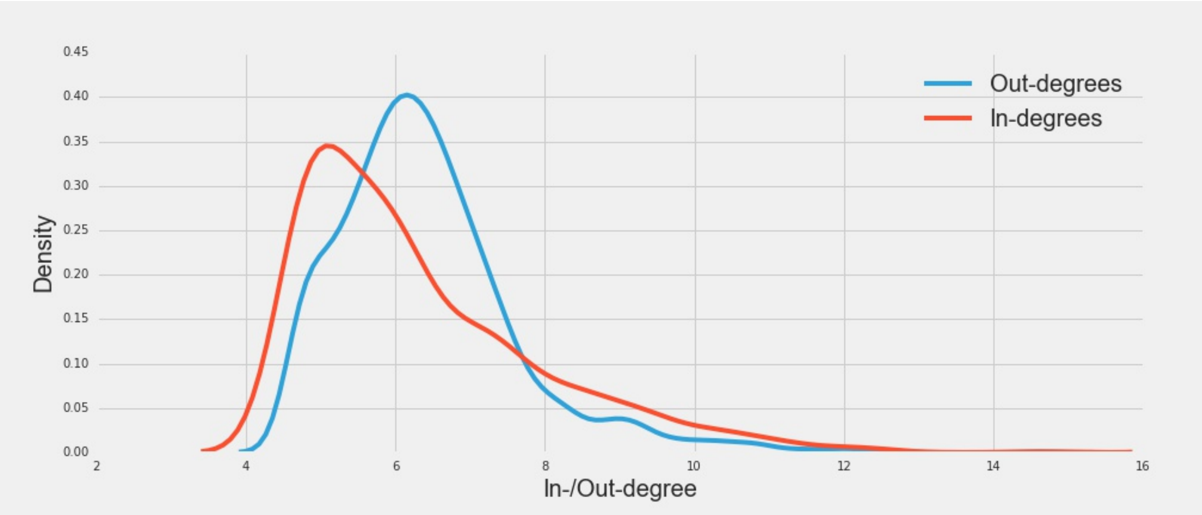


Table I - In- and Out-degrees

Variable	Median	Min	Max	Std.
In-degrees	106.5	0	47,820	937
Out-degrees	418	0	12,043	2815

Fig. 2 - Cumulative Distribution of Commuter Flow

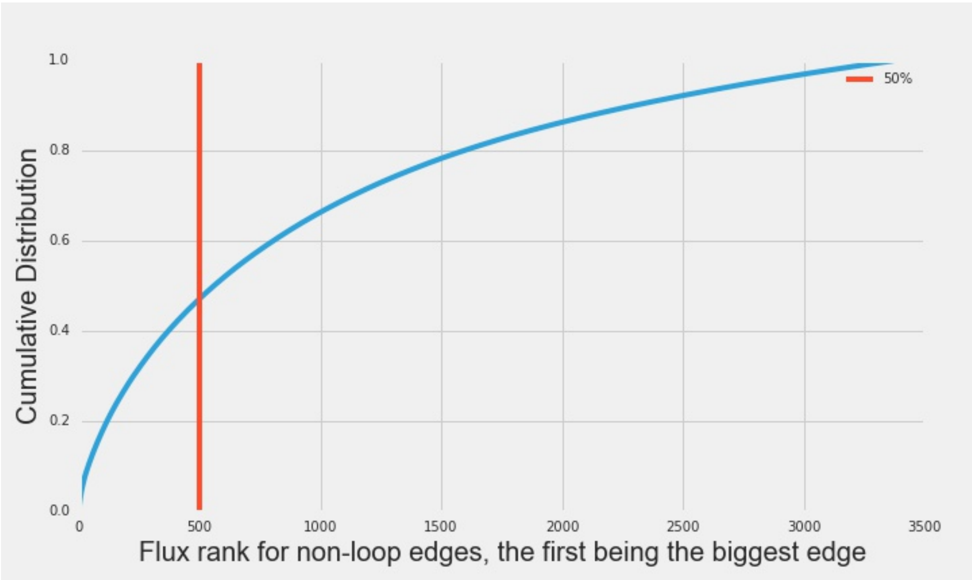


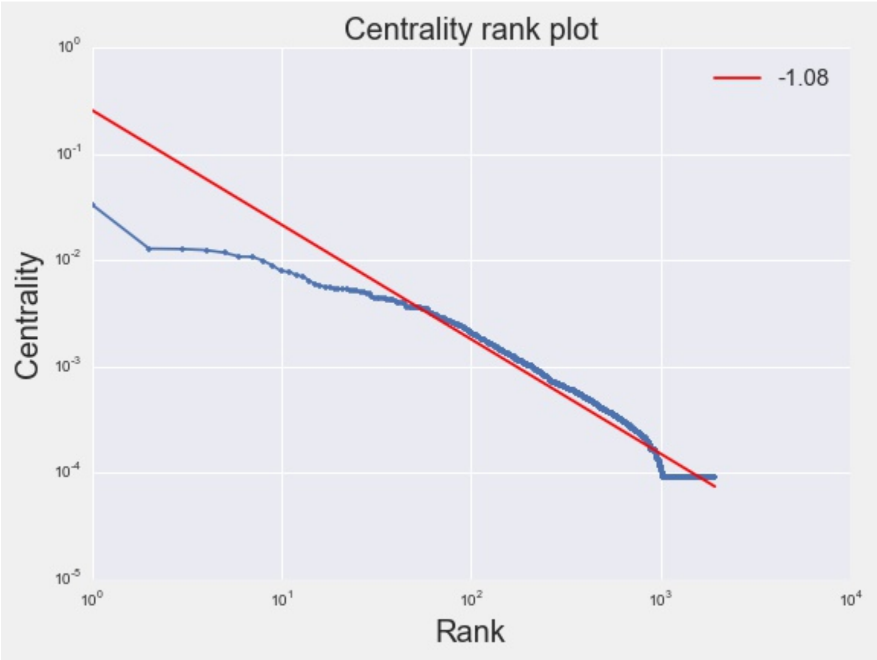
Table II - Top 10 Urban Unit Pagerank Scores

Rank	Urban Unit	Pagerank Score (10^{-2})
1	Paris	3.338
2	Strasbourg	1.286
3	Marseille – Aix-en-Provence	1.274
4	Rennes	1.246
5	Montpellier	1.176
6	Lille	1.082
7	Lyon	1.076
8	Nantes	0.989
9	Toulouse	0.888
10	Metz	0.786

Appendix A - For nodes a and b, with a probability $1 - \alpha$ of following a random walk, pagerank centrality is computed as following:

$$pagerank(a) = \alpha \sum_{(b,a)\in E} \frac{pagerank(b)}{out\_degree(b)} + \frac{1 - \alpha}{n}$$

Fig. 4 - Centrality rank plot (pagerank centrality)



Appendix B – Spatial Interaction Models

1. Gravity Model:

$$E_{ij} = k.W_i^{\alpha} W_j^{\beta} D_{ij}^{\gamma} \text{ with } k = \sum_{i \neq j} E_{ij} / \sum_{i \neq j} W_i^{\alpha} W_j^{\beta} D_{ij}^{\gamma}$$

2. Gravity model with local normalization

$$E_{ij} = T_i A_i W_j^{\beta} D_{ij}^{\gamma} \text{ with } A_i = 1/\left[\sum_{k \neq i} W_k^{\beta} D_{ik}^{\gamma}\right]$$

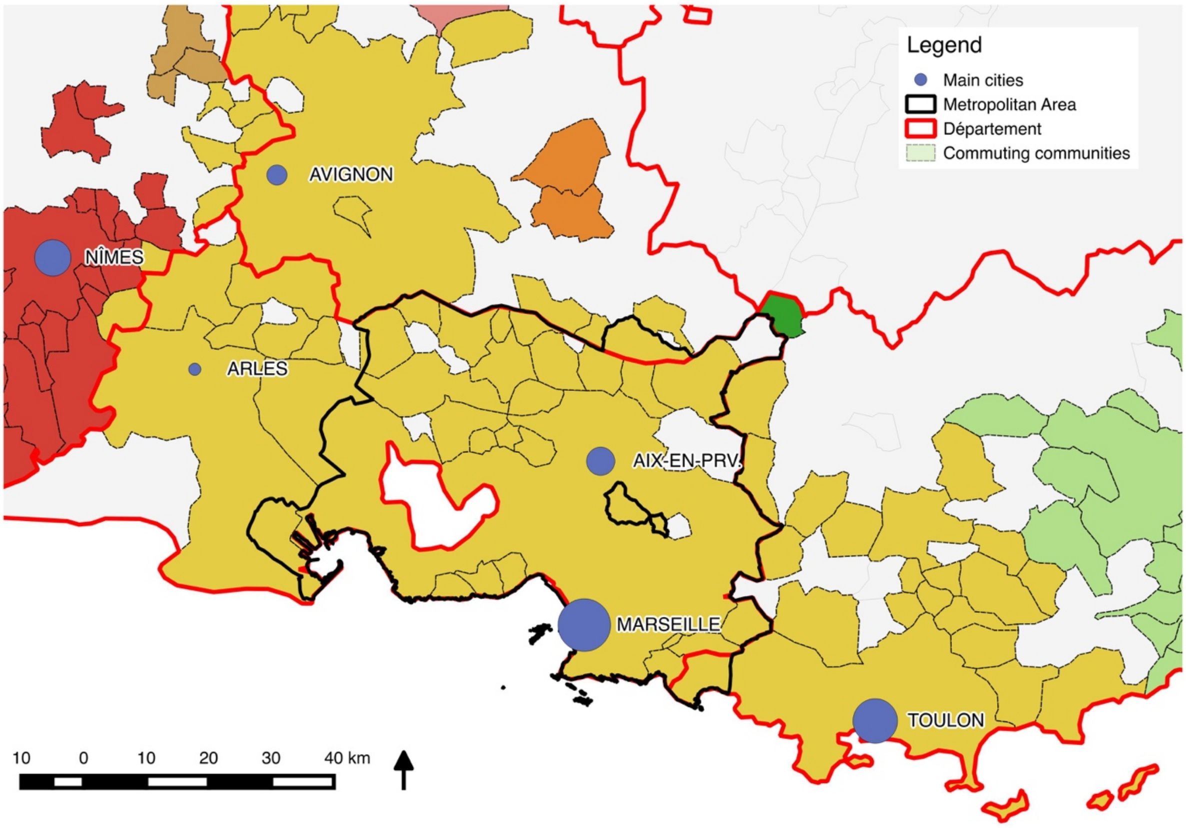
- E_{ij} is the number of commuters between Units i and j
- W_i is the number of active workers in area i
- D_{ij} is the distance (or duration) between nodes i and j
- T_i, is the total outgoing strength of node i

Appendix C – Modularity Score

$$Q(P) = \sum_{i,j,c_i=c_j} \left[\frac{E_{ij}}{T} - \frac{k_i^{out} k_j^{in}}{T^2}\right]$$

- For a given partition $P = c_i, i \in N$, where N is the network
- T is the total weight of the network,
- k_i^{out} is the outgoing strength of node i
- k_j^{in} is the ingoing strength of node j

Map 1 – Commuting community and Metropolitan Area around Marseille



Bibliography

[1] A. Smith and P. Heywood, “Regional Government in France & Spain,” August, 2000.

[2] J. M. Sellers, “Governing from below : urban regions and the global economy,” *Urban Region and the Global Economy*, pp. xii-407, 2002.

[3] S. Yusuf, “Will cities continue driving economic growth ?,” *Growth Dialogue Work. Pap.*, vol. 6, no. June, pp. v–26, 2013.

[4] R. Patuelli, A. Reggiani, S. P. Gorman, P. Nijkamp, and F. J. Bade, “Network analysis of commuting flows: A comparative static approach to German data,” *Networks Spat. Econ.*, vol. 7, pp. 315–331, 2007.

[5] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti, “Delineating geographical regions with networks of human interactions in an extensive set of countries,” *PLoS One*, vol. 8, no. 12, 2013.

[6] INSEE, Profesionnal Mobility Dataset http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/doc/presentation-fichiers-detail-mob-pro.htm

[7] INSEE, Definitions and Methods: Urban Units http://www.insee.fr/en/methodes/default.asp?page=definitions/unite-urbaine.htm

[8] Google Developers, Google Maps Distance Matrix API https://developers.google.com/maps/documentation/distance-matrix/intro

[9] T. Schwanen, “Urban form and commuting behavior: a cross-European perspective,” *Tijdschrift voor economische en sociale geografie,* vol. 93, pp. 336-343, 2002

[10] D. F. Gleich, “PageRank beyond the Web,” vol. 57, no. 3, pp. 321–363, 2015.

[11] J. Gingembre and J. Baude, “Les mobilités domicile-travail dans les réseaux d’agglomérations,” *EchoGéo*, vol. 27, pp. 1–14, 2014.