Multiplex Network Analysis Public Transportation

Matteo Fasulo
Tecniche Informatiche per la gestione dei Dati
Libera Università Maria Santissima Assunta
Rome, Italy
info@matteofasulo.me

Matteo Sivoccia
Tecniche Informatiche per la gestione dei Dati
Libera Università Maria Santissima Assunta
Rome, Italy
m.sivoccia@lumsastud.it

Simone Flavio Paris
Tecniche Informatiche per la gestione dei Dati
Libera Università Maria Santissima Assunta
Rome, Italy
simone.flavio.paris@gmail.com

Lorenzo Bianchi
Tecniche Informatiche per la gestione dei Dati
Libera Università Maria Santissima Assunta
Rome, Italy
lorenzobianchi5082@gmail.com

***Abstract*—This paper contains the multimodal transport network of Ille-de-France (Paris) modelled with a multi-layer network structures. Each layer in the network represent a public transport network with last layer being the road layer. In order to connect two different transport layers, a spatial approach is proposed by making a new edge between two nodes on different layers. A set of nodes near point of interest is analysed to verify if the presence of a museum or a monument guarantees more centrality to those nodes.**

***Index Terms*—Multimodal transport networks, Multiplex Network, Transport Analysis**

I. Introduction

Public Transport Layer (PTN) contain multiple layers of traffic networks like: metro, train, road and tram network. The analysis of traffic flow in large metropolitan areas is often a challenging task. In the analyses of the PTN, complex graphs and networks are used with focus on spatial analysis of the network considering the individual transport level or the overall transport level. In this context, data has recently become an interesting source of information and analysis on mobility behaviour since it is difficult for city authorities to act on an unified view of mobility patterns. In the analyses made on PTNs, which use graph theory, the networks are represented as graphs with nodes representing the stops of means of transport such as trains and trams and edges representing the connection of the stops of the respective PTNs. Depending on the type of edge, the graph can be modelled in a space S, where there is an edge between two stops if they are consecutive. The majority of passengers use multiple transport nodes to reach their destination or even different transport layers and the interaction between network layers may lead to a better understanding of the entire network. In this paper we make a contribution to the research and analyses carried out on the transport structure of Ille-de-France (Paris).

II. Related work

Kivela et al. [1] introduced the representation and notation for the monoplex (mono-layer) networks, multiplex (multi-layer networks), interdependent networks, interconnecting networks, networks of networks, etc. Tomasini [2] also introduced measure for multi-layer analysis and Zanin [3] demonstrated that the results of the analysis of single projected layer might yield a biased understanding of the actual network by comparing with its multi-layer functional network counterpart.

III. Multiplex Networks

Recalling a graph G is a set of nodes V and edges E, $G = (V, E)$. Our work propose the spatial analysis by representing

$$G = (V(x, y), E) \tag{1}$$

with $V = \{n_i(x_i, y_i), x_i = latitude, y_i = longitude\}$ and $E = e_{ij} \to (n_i(x_i, y_i), n_j(x_j, y_j)$. For instance we assume that a particular node is always identified with its latitude and longitude so $n_i(x_i, y_i)$ will be a unique value representing n_i. In transport network analysis the inspection of spatial embedding of nodes manifest locations in which two or more nodes are almost overlapped. In these specific cases, if d_{ij} (the distance between two nodes n_i and n_j) satisfy the condition

$$d_{ij} < d_{th} \tag{2}$$

then n_i and n_j will be connected with a *inter-layer edge*. Indeed our graph structure will be

$$G = (V(x, y), E, E_L) \tag{3}$$

with E_L being *inter-layer edges*. The distance d_{ij} is calculated with the Haversine formula [4] and d_{th} is set to be 455 m assuming that it is a walkable distance to reach a station. Considering the previous graph structure we define the multilayer network, M, as follows

$$M = (V_M, E_M, L) \tag{4}$$

$L = \{L_\alpha\}_{\alpha=1}^{d}$ is the set of layers with d dimensions. For $d = 1$ (single dimension), the network reduces to a mono-layer aspect. In this paper, $d = 4$ with an elementary layer and an additional layer. $E_M \subseteq V_M \times V_M$ is the edge set with bot *intra-layer* and *inter-layer* edges. Among all layers, the metro layer is considered the elementary layer since it is the most preferred transport mode in Ille-de-France. Another interesting aspect is that the transport network structure is a layer-disjoint multi-layer network so each node exists in at most one layer

$$(n_i)_\alpha, (n_i)_\beta \in V_M \Rightarrow \alpha = \beta \tag{5}$$

where a node is present either in layer α ($L_1 = \alpha$) or β ($L_2 = \beta$). The layer-disjoint property signify an important observation that there exists no edges between the layers in the actual network structure, and the layers are normally connected virtually (by a small walking distance) and not physically. In order to connect these layers we employ the method of spatial join with a radius of 455 m. The overlapped nodes in the 455 m region will be edge-connected and a weight attribute will be added to those edges. The weight attribute added is an estimation based on both distance between nodes and transfer speed (T_s) (i.e. walking speed to reach destination) plus a basic cost (for simplicity B_c) representing the mean time waiting the transportation vehicle. Hence the weight will be

$$w_{ij} = \frac{d_{ij}}{T_s} + B_c \tag{6}$$

with T_s set at 5 km/h (mean walking speed) and B_c set at 11 minutes. [5]

IV. Dataset description

The analysed dataset contains information about the public transport lines of Paris, by splitting them in four different layers: metro, train, tram, road.
The data are collected from different sources, like *OpenStreetMap* [6] (OSM) and *The National Institute of Geographic and Forest Information* [7] (IGN), a french public administrative establishment, and then gathered to perform the analysis. Table I has a summary of every layer with average degree and reference database.

The collection of data is grouped into a **GeoJSON** file containing both nodes and edges.

A. Nodes properties

Every node in the dataset has a *coordinate* attribute indicating *latitude* and *longitude* to identify node's position as well as a *layer's* attribute to identify its source.

B. Edges properties

Every edges has a name attribute (that could be road name, train line, subway line or metro line), direction indicating if the edge is directed (i.e. one way) or undirected (i.e. two way) and a layer attribute to identify its source. Among *intra-layer edges* there is also a *length* parameter that measure the distance between nodes of the edge itself. There is also a fifth layer named *crosslayer* that groups all the edges between nodes standing on different layer, the *inter-layer edges*.

C. GeoJSON format

There are also information about the coordinates of the different transport hubs, that are of type *Point* for the nodes and type *LineString* for the edges.
The *Points* are lists in the format $[lat, lon]$ while *LineString* contain coordinates of both source and target node (i.e u and v) in the format $[[u_{lat}, u_{lon}], [v_{lat}, v_{lon}]]$
The latitude and the longitude format follows the standard coordinate system *WSG 84*.

D. Transport lines

- *Train* (5) $\Rightarrow$ RER A, RER B, RER C, RER D, RER E
- *Tram* (7) $\Rightarrow$ T1, T2, T3a, T3b, T4, T5, T7
- *Metro* (16) $\Rightarrow$ M1, M2, M3, M3bis, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14
- *Road* (8563) $\Rightarrow$ A86, D1, D19, D9, D11, D17, D130, D14, etc

V. Analysis

The first part on the analysis regards the distribution of nodes and edges between levels. The road network has over 95% of all nodes as well as over 82% of edges (figure 3). An interesting aspect is that crosslayer layer represent 15% of edges (4071 edges) so there are a lot of junctions points where it is possible to change to different transport mode. The average degree values ranges from 1.9 (tram) to 3.0 (road). This is not surprising since, except for specific stations, the metro, train and tram networks are rail based so the majority of stations will have a degree value of 2. Road network, instead, has more parallel ways and crossings so its degree value is higher.
Both the node map and the edge map are available at GitHub repository under section *maps* [8]. Then we continued with a top-down approach of analysis, starting from large scale measures to arrive at those focused on single nodes. Table II summarises the maximum degree among all layers as well as density, shortest path length, diameter, etc. It is clear that *road* network has the highest degree thanks to the fact that it has the majority of nodes and edges (table I). Recalling that density is calculated as

$$d = \frac{2m}{n(n-1)} \tag{7}$$

with n nodes and m edges. Road layer has the lowest density since $n(n-1) \sim n^2$ with $n = 14804$. One factor of interest could be the shortest path length being ~ 20 in *tram* network but the main reason is due to the fact that rail based transport mode rarely have parallel ways and not even crossings. There could be only one single path to reach the destination station and so on. Both *shortest path* and *bridges* are not available for road layer since, while metro, train and tram are undirected graphs, this network is modelled as a directed graph.
Also the average cluster coefficient underlines the property of roads of being more clustered while obviously others are not clustered at all (i.e. train and tram) or poorly clustered.
A bridge is an edge whose removal causes the number of

connected components of the graph to increase. It can be noticed that rail based transport mode tend to have an high number of bridges due to the fact that even a single edge removal could cause the collapse of the entire network. As a special case, the tram network has a number of bridges equals to the number of edges (on its maximum connected component by length).

In Figure 1 there are tree different plots and also the network draw of road layer. First graph represents the *complementary cumulative distribution function* modelling the probability of finding a node of degree k or higher. Usually these functions have a heavy tail for the presence of hubs while the most portion of distribution represents nodes with low degree values. As we can see this cumulative distribution has not a long-tail, perhaps, in all the networks we have analysed (individual layers) is the more attack resistant. In fact, more a network has a heavy tail more is resilient to random removal of nodes, also called failure) but vulnerable to the removal of the hub nodes (targeted attacks). In this scenario the probability of finding an hub with more than 10 edges is less than 2%, we can evince that better in the degree histogram where the vast majority of nodes have a very low degree and only a few hubs have a degree larger than average.

Thanks to the spatial join this network have a much greater maximum degree value 38 (table III) than others (table II). We can also notice that the principal hub of the network is a tram station near the ceramic museum "*Manufacture nationale de Sèvres*" (POI) and one of the main road junctions on the west side of the Seine that is also connected with another relevant node junction on the other side of the bridge.

The *degree histogram* instead could be approximated with a Poisson distribution with $\lambda = 3$. Both CCDF and degree distribution suggest that our multi-layer network is not so far away from the Watts–Strogatz model (WS). In the figure 2 are showed Newman-Watts–Strogatz (NWS) model with $n_{NWS} = n_M = 15492$ $e_{NWS} = 26330$ (while the multi-layer network has $e_M = 27071$) and the WS model with $n_{WS} = n_M = 15492$ $e_{WS} = 30984$. We can see that the complementary cumulative distribution function has a similar trend as well as the degree histogram similar to a Poisson. Also betweenness distribution would have been similar to WS model if we have limited the *x-axis*. The WS model is the best one for the fitting of the CCDF distribution and for the degree histogram. It was difficult to find a good model that fit the betweenness distribuition, in fact the multi-layer one is like a zero inflated.

The node with the maximum Between Centrality is located near the airport and connects the south-east transportation. The "*Villeneuve St Georges*" stop connects train lines A, B and D with Bus lines N, J1, G1, G2 and H.

Right in the centre of this metropolitan city there are two stops at 505 meters away from each other representing *closeness centrality* and *eigenvector centrality* nodes, respectively a train station and a metro station.

In close proximity to the first one, there are several Points of Interest: the *Nelson Mandela Garden*, the *Bourse de Commerce "Pinault Collection"* Museum and many stores. Also the train station is well connected with the "*Les Halles*" metro stop which is only 93 meters away. The second one is the metro station for "1", "4", "7", "11" and "14" lines, and it's near the "*Musée du Louvre*", "*Square de la Tour Saint-Jacques*" and "*Sainte-Chapelle*".

Generally, the four measures of centrality are correlated to each other. This happens because the degree centrality and the closeness one are directly involved in the number of edges that a node has, while the betweenness is mostly involved in the cardinality of the hops that participate at the shortest path. Nevertheless not always we can find true that degree and closeness are closely correlated. In fact, it could exist a hub which is not placed in a strategic position and therefore is not central. We note a slight correlation between degree and closeness, confirmed in our figure 5. In this specific case the two nodes with the maximum closeness and eigenvector result close to each other, but it happens just for a real casualty because their correlation is completely neutral ($\sim [0.152]$). Probably it occurs because in the proximity of the node with the maximum eigenvector centrality (that is about the prestige score: per definition is a measure of the influence of a node based on the concept that connections to high-scoring nodes contribute more to the score of the node in question), there are other nodes with an elevated degree also by virtue of the fact that their position is very central (for example near Musée du Louvre and Avenue des Champs-Élysées). As for the node with the maximum closeness, it is trivially the most central of all the nodes in our network and therefore is located around the Parisian centre.

TABLE I: Different transportation networks with their properties

Layer	Nodes	Edges	Avg.Degree	Reference
Metro	303	356	2.350	OSM
Train	241	244	2.025	OSM
Tram	144	140	1.944	OSM
Road	14804	22281	3.010	IGN

TABLE II: General measures of each layer (max. connected component)

Metric	Metro	Train	Tram	Road
Max Degree	8	4	3	12
Density	0.00778	0.00844	0.0136	0.0001
Shortest Path	12.21	16.54	20.74	None
Diameter	34	47	61	187
Avg. Cluster Coefficient	0.008809	~ 0	~ 0	0.03289
Transitivity	0.0186	0	0	0.0694
Bridges	115	152	140	None
Max Degree Centrality	0.02649	0.0167	0.02098	0.0008
Max Betweenness Centrality	0.32723	0.2269	0.1041	0.0222
Max Eigenvector Centrality	0.34506	0.50485	0.50028	0.3877
Max Closeness Centrality	0.1277	0.06554	0.03634	0.01078

TABLE III: General measures of multi-layer network (max. connected component)

	Measures	Coordinates
Max Degree	38	—
Density	0.000226	—
Shortest path	37.615767	—
Diameter	115	—
Avg. Cluster Coefficient	0.121424	—
Transitivity	0.220804	—
Bridges	207	—
Max Degree Centrality	0.002453	—
Max Between Centrality	0.301238	—
Max Eigenvector Centrality	0.319027	—
Max Closeness Centrality	0.043421	—
Hub	tram_613	48.828750 2.225326
Node Degree Centrality	tram_613	48.828750 2.225326
Node Between Centrality	train_77	48.730151 2.446160
Node Eigenvector Centrality	metro_269	48.858339 2.346701
Node Closeness Centrality	train_194	48.862324 2.346637

VI. Conclusions

Compared to a normal graph representation, the multiplex graph representation offers a more specific way of analyse the transport network topology which not only helps in a realistic network analysis, but also takes into account an accurate representation of the network without ignoring inter-layer interactions. From this type of analysis it can be observed that the centrality of nodes differs significantly in the monolayer and multilayer analyses of the network. It therefore seems clear that is important to consider all aspects of the multimodal and multilayer network to understand the behaviour of an urban transport system and therefore to understand the effects of transport on other characteristics of interest. Although these studies are still very theoretical, they convincingly show that reasoning with only one mode of transport can be extremely misleading (i.e. leading to a biased view of the network) and therefore it is better to have a spatial multi-layer approach.

Further works

A better analysis by splitting nodes in neighborhoods and grouping them by macro indicators like working population, job opportunities, number of points of interest in that area, etc. Add a weight on every edge based on the average time needed to move between stations or roads.
Add the bus layer since a lot of people use buses in Paris.
Implement a merge method based on spatial join. Instead of adding an edge when $d_{uv} < d_{th}$ manteining original nodes, a new node will be created as centroid of old nodes positions so the topology of the network would result smaller and easier for further analysis.

Acknowledgment

We have dedicated this to the achievement of a *Ille-de-France Multi-layer Public Transport Network Analysis* inspired by [9].

References

[1] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," Journal of complex networks, vol. 2, no. 3, pp. 203–271, 2014.
[2] M. Tomasini, "An introduction to multilayer networks," BioComplex Laboratory, Florida Institute of Technology, Melbourne, USA, pp. 1–14, 2015.
[3] "Can we neglect the multi-layer structure of functional networks?" Physica A: Statistical Mechanics and its Applications, vol.430, pp. 184–192, 2015.
[4] C. Veness, "Calculate distance and bearing between two latitude/longitude points using haversine formula in javascript," Movable-Type Scripts, 2011.
[5] "Waiting Time In Paris,"Moovit Website, [Online; accessed 21-May-2022].
[6] "Open Street Map,"Open Street Map Website, [Online; accessed 21-May-2022].
[7] "National Institute of Geographic and Forest Information,"IGN Website, [Online; accessed 21-May-2022].
[8] "Multiplex Network Analysis Public Transportation,"Repository, [Online; accessed 21-May-2022].
[9] "Multiplex Metro," Riyadh Metro Multiplex Analysis, [Online; accessed 21-May-2022].

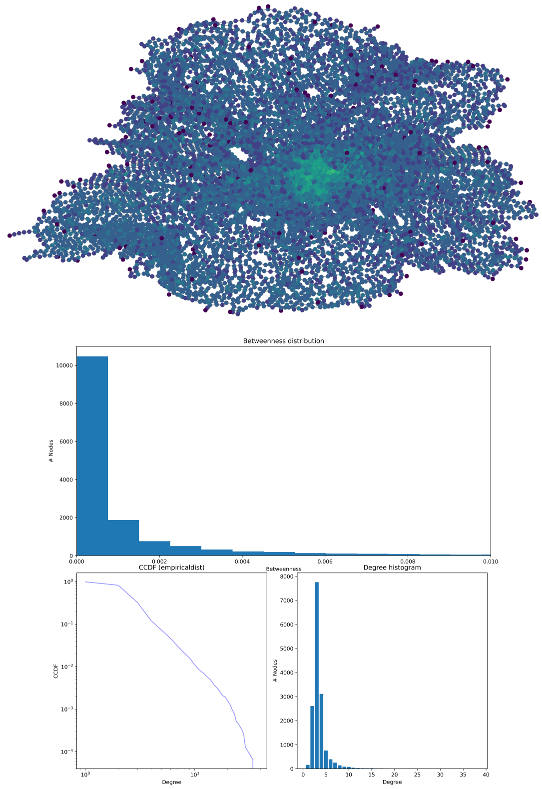
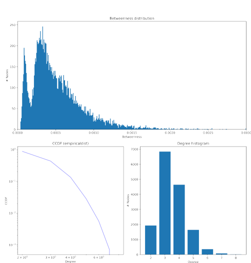
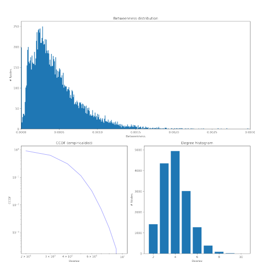


Fig. 1: Degree Analysis (CCDF and distribution) along with betweenness distribution. Lighter nodes have higher degree.

FIGURES

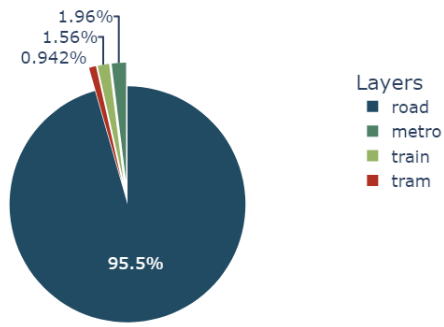


(a) NWS model

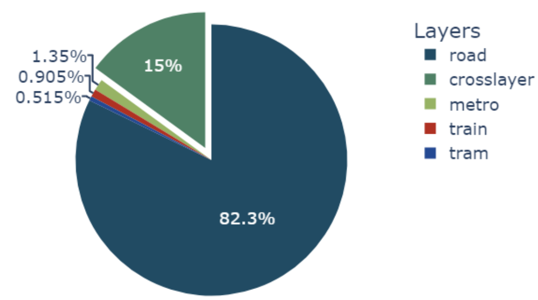


(b) WS model

Fig. 2: Degree analysis comparison between NWS model (a) and WS model (b)



(a) Nodes distribution



(b) Edges distribution

Fig. 3: Nodes and edges distribution over layers

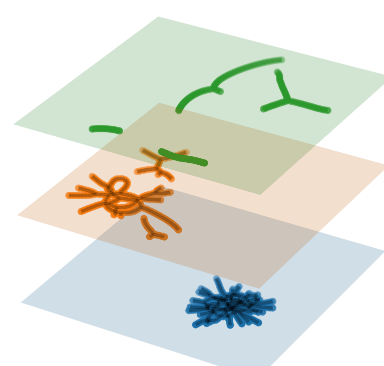


Fig. 4: 3d plot of metro (green), train (orange) and tram (blue) networks. Layers are not sharing any node, a confirm that it is a multi-layer network layer-disjoin

	degree_centrality	closeness_centrality	eigenvector_centrality	betweenness_centrality
degree_centrality	1.000	0.450	0.322	0.205
closeness_centrality	0.450	1.000	0.152	0.120
eigenvector_centrality	0.322	0.152	1.000	0.105
betweenness_centrality	0.205	0.120	0.105	1.000

Fig. 5: Correlation of centrality measure (cross-layer), darker color indicates strong correlation

	degree_centrality	closeness_centrality	eigenvector_centrality	betweenness_centrality
degree_centrality	1.000	0.443	0.496	0.715
closeness_centrality	0.443	1.000	0.671	0.551
eigenvector_centrality	0.496	0.671	1.000	0.607
betweenness_centrality	0.715	0.551	0.607	1.000

Fig. 6: Correlation of centrality measure (metro layer), darker color indicates strong correlation

	degree_centrality	closeness_centrality	eigenvector_centrality	betweenness_centrality
degree_centrality	1.000	0.402	0.227	0.362
closeness_centrality	0.402	1.000	0.610	0.594
eigenvector_centrality	0.227	0.610	1.000	0.087
betweenness_centrality	0.362	0.594	0.087	1.000

Fig. 7: Correlation of centrality measure (tram layer), darker color indicates strong correlation

	degree_centrality	closeness_centrality	eigenvector_centrality	betweenness_centrality
degree_centrality	1.000	0.312	0.360	0.513
closeness_centrality	0.312	1.000	0.471	0.673
eigenvector_centrality	0.360	0.471	1.000	0.564
betweenness_centrality	0.513	0.673	0.564	1.000

Fig. 8: Correlation of centrality measure (train layer), darker color indicates strong correlation