Studying Urban Transportation Networks Using a Multi-Layered Approach

Ciência das Redes Complexas

Daniel Dias 102124
Guilherme Pires 102132
Miguel Figueiredo 98722

November 2021

Abstract

Transportation networks play a huge role in urban mobility and those networks have their own means such as bicycle paths, pedestrian paths, roads and railways, all of those paths can be combined in various ways leading to a multiplex transportation network. The paper *Extracting the Multimodal Fingerprint of Urban Transportation Networks* studies precisely this topic and groups different cities into clusters considering their networks' overlap census. Here we will try to replicate the paper, with only 14 of those cities, and apply further studies to the network. We try to find a relation between the networks' nodes and edges and the cities' population. We also study TomTom's 2017 *Traffic Congestion Ranking* and try to predict where the rest of the cities that don't appear there would land in that ranking. Our work may help cities to improve their urban mobility in the future. We find that the overlap census has a certain impact on a city's Traffic Congestion and that a city's population doesn't have impact on the size of its rail and bike networks.

1 Introduction

We can describe each urban mobility network as a mathematical object, a multiplex network. A city's multiplex network is a multi-layered network composed by different layers, such as bicycle paths, pedestrian paths, roads and railways, that keep that city moving. The paper *Extracting the Multimodal Fingerprint of Urban Transportation networks* [1] studies "15 world cities and develop an urban fingerprinting technique based on multiplex network theory to characterize the various ways in which transport layers can be interconnected, identifying the potential for multimodal transport."(Orozco et al. 2020) The paper also finds classes of cities, regarding their transportation priority using clustering algorithms on the resulting urban fingerprints. We replicate the findings of this paper, with only 14 of those cities, and go even further by trying to find a relation between each cities' population and its different networks, try to understand the placement of the cities in the TomTom's 2017 *Traffic Congestion Ranking* [2] and predict the position of the cities that don't appear on the ranking.

2 Methods and Data

For the networks' dataset we used the same dataset [3] as the one used by the authors in [1]. Each network is represented by nodes and edges where each node represents a keypoint/intersection on the urban mobility network(Eg.: Train station, Store, Monument...) and each edge represents a connection between two nodes(Eg.: Road, Bridge, Cicleway...). We extracted each city's Bike, Drive, Pedestrian and Rail networks, a total of 55 different networks($4 * 14 - 1$ since Beihai doesn't have a bike network even though we use that network as an empty graph in our code). Since these networks are all real and urban networks, they cannot be described as scale-free networks: their degree distribution doesn't follow any power law, as can be seen in Fig. 1. For the population data we also used the paper's [1] population info.

To study the data we used python [4], including some specific libraries like networkx [5], matplotlib [6], altair [7] and pandas [8]. We also used three functions made by the paper's [1] authors that are available on their GitHub [9]. All of our code is available in GitHub [10] as well as the data used and a readme file explaining how to retrace our steps.

Before going into the overlap census, the ratio of nodes that are exclusive to that network in relation to the city's multi-layered network with all the layers combined, we calculated every network's average degree($\langle \mathrm{k} \rangle$) by dividing the sum of each node's degree by the total number of nodes in that network and inserted that information as well as every network's nodes and edges and every city's population in Table 1. We also combined each city's network into a multi-layered network and extracted the number of nodes and layers of each network, we calculated their average degree($\langle \mathrm{k} \rangle$) as seen in Table 2

2.1 Overlap Census

For the overlap census, for each city we gathered the Bike, Rail, Pedestrian and Drive networks and used every network combination, leaving us with 15 different networks for each city. After that each city's networks' overlap census was calculated and plotted the results to each city.

2.2 Population X Network

For the analysis of a pattern between the cities' population and its network we extracted every city's population and data(node count and link count, including from multi-layered networks) from Tables 1 and 2 to a json file and plotted two different graphs to see if we could find any relation between the cities' population and any network's nodes and/or edges.

2.3 Network Rankings and Prediction

For the analysis network/traffic rankings and placement prediction we gathered the TomTom's 2017 *Traffic Congestion Ranking* [2] and inserted the data from the cities we are studying in Table 3 and predicted the position of the cities that aren't in the ranking.

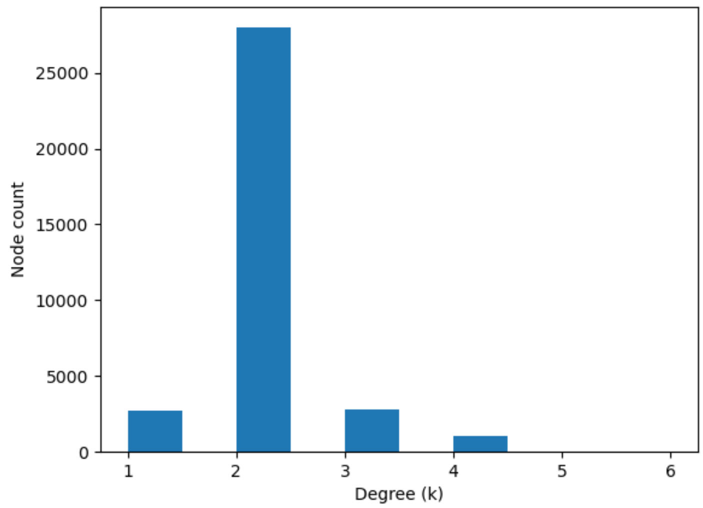


Figure 1: Degree Distribution to Amsterdam's Bike Network. This figure shows how many nodes are characterised by each degree value. Other layers show similar degree distributions.

	Bike			Drive			Pedestrian			Rail			Population
	Nodes	Links	⟨k⟩	Nodes	Links	⟨k⟩	Nodes	Links	⟨k⟩	Nodes	Links	⟨k⟩	
Amsterdam	34,529	35,619	2.06	15,125	21,722	2.87	23,321	33,665	2.89	1096	1655	3.02	872,680
Barcelona	7,553	7,647	2.02	10,393	15,809	3.04	20,203	30,267	3.00	249	249	2.00	1,600,000
Beihai	0	0	0.00	2,192	3,209	2.93	2,026	2,978	2.94	59	62	2.10	1,539,300
Bogota	9,760	9,651	1.98	62,017	91,197	2.94	81,814	121,038	2.96	166	165	1.99	7,412,566
Budapest	10,494	10,318	1,97	37,012	52,361	2.83	73,172	106,167	2,90	1,588	1,964	2.47	1,752,286
Copenhagen	13,980	13,988	2.00	15,822	20,451	2.59	30,746	41,916	2.73	276	369	2.67	2,557,737
Detroit	3,663	3,626	1.98	28,462	45,979	3.23	47,828	78,391	3.28	20	21	2.1	672,662
Jakarta	248	231	1.86	138,388	188,637	2.73	140,042	191,268	2.73	58	54	1.86	10,075,310
Los Angeles	14,577	14,428	1.98	71,091	101,692	2.86	89,543	128,757	2.86	173	221	2.55	3,792,621
London	62,398	60,043	1.92	179,782	219,917	2.45	270,659	351,824	2.60	2,988	3,535	2,37	8,908,081
Manhattan	3,871	3,777	1.95	5,671	9,379	3.08	13,326	21,447	3.21	349	436	2.50	1,628,701
Mexico City	5,218	5,278	2.02	95,375	140,684	2.95	108,033	158,425	2.93	370	364	1.97	8,918,653
Phoenix	35,631	35,979	2.02	73,688	102,139	2.77	111,363	157,075	2.82	105	138	2.63	1,445,632
Portland	24,252	24,325	2.01	35,025	49,062	2.80	50,878	72,958	2.87	230	340	2.96	583,776

Table 1: Every city's population networks' nodes, links and average degree - All the results match the paper's results except Manhattan's Drive Average Degree.

	Multi-layered Network		
	Nodes	Links	⟨k⟩
Portland	69,041	100,864	2.92
Phoenix	139,869	196,577	2.81
Detroit	50,834	89,986	3,54
Amsterdam	58,571	79,018	2,70
Budapest	83,855	127,751	3.04
London	332,790	447,249	2,69
Los Angeles	103,127	150,332	2.92
Barcelona	27,679	41,245	2,98
Copenhagen	39,839	57,267	2.87
Jakarta	141,217	194,402	2.75
Mexico City	114,405	169,683	2.97
Bogota	91,986	138,099	3.00
Manhattan	16,979	27,638	3.25
Beihai	2,294	3,355	2.93

Table 2: Every city's multi-layered network's nodes, links and average degree.

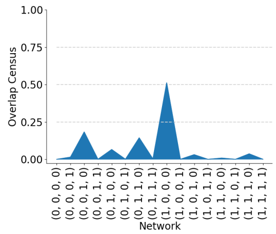
#	World Rank	City	Congestion Level
1	3	Bogota	62%
2	4	Jakarta	61%
3	8	Mexico City	52%
4	19	Los Angeles	42%
5	34	London	36%
6	79	Budapest	31%
7	101	Barcelona	28%
8	127	Portland	27%
9	146	Amsterdam	25%
10	203	Copenhagen	22%
11	291	Detroit	17%
12	294	Phoenix	17%

Table 3: TomTom's 2017 Traffic Congestion Ranking.

3 Results

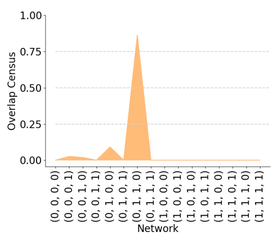
3.1 Overlap Census

We plotted 14 different graphs, one for each city and decided to group the graphs in 4 different groups, the first one has the cities that have a predominant Bicycle network - Fig.2; the second one has the cities that have a predominant Pedestrian-Drive network that translates to car-centric cities - Fig 3; the third one has the cities that have two predominant networks, the Pedestrian only and the Pedestrian-Drive, that also translates to car-centric cities - Fig 4; and the final one has the cities that have a balance between three networks, the Pedestrian only, the Bike only and the Pedestrian-Drive, that translates to mobility active cities - Fig 5.

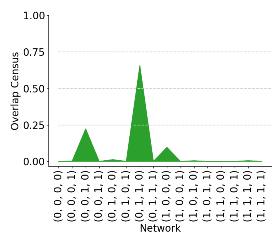


(a) Amsterdam

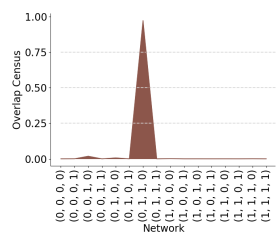
Figure 2: City that has a predominant Bike network.The x-axis represents each network combination (represented by a combination of zeros and ones where if it is 0, that network isn't considered and if it is 1 that network is considered) and the y-axis represents the overlap census that can be a number between 0 and 1. The order of the networks is Bike, Drive, Pedestrian and Rail. Eg.: x = (1,0,0,0) represents the Bike network and x = (0, 1, 1, 0) represents the Drive-Pedestrian network.



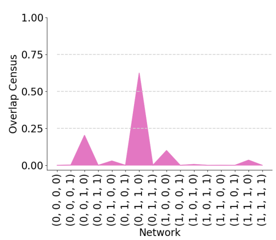
(a) Beihai



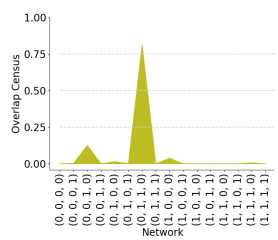
(b) Bogota



(c) Jakarta

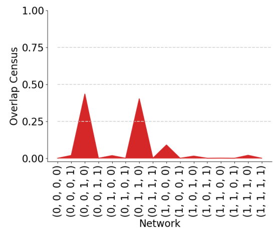


(d) Los Angeles

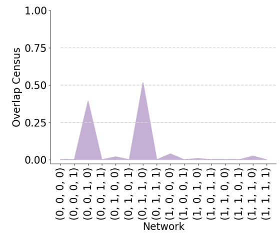


(e) Mexico City

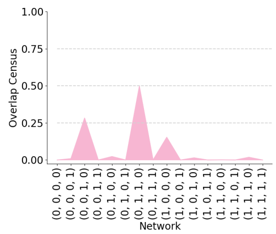
Figure 3: Cities that have a predominant Pedestrian-Drive network.



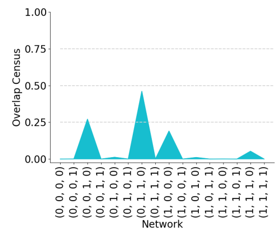
(a) Budapest



(b) Detroit

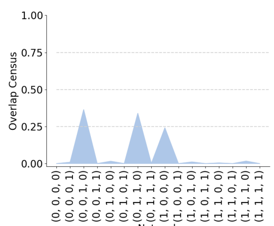


(c) London

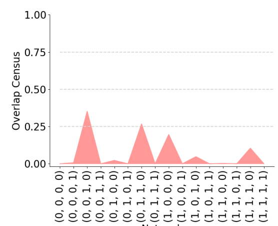


(d) Phoenix

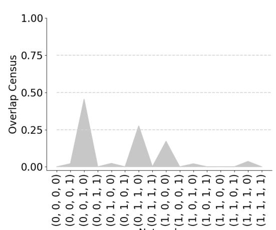
Figure 4: Cities that have two predominant networks, the Pedestrian only and the Pedestrian-Drive.



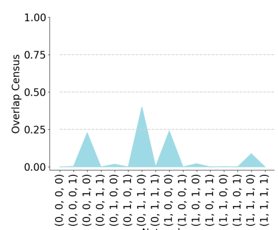
(a) Barcelona



(b) Copenhagen



(c) Manhattan

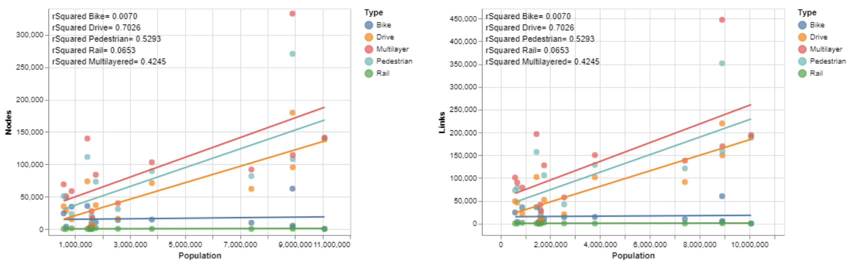


(d) Portland

Figure 5: Cities that have a balance between Bike only, Pedestrian only and Pedestrian-Drive networks.

3.2 Population X Network

We plotted 2 graphs where the x axis represents the population and the y axis represents the node count for Figure 6a. and link count for Figure 6b. and highlighted the Bike, Rail, Pedestrian, Drive and multi-layered(combination of all the previous layers) layers.



(a) Node count

(b) Link count

Figure 6: Relation between population over node and link count. Cities ordered by population, from lowest to highest: Portland, Detroit, Amsterdam, Phoenix, Beihai, Barcelona, Manhattan, Budapest, Copenhagen, Los Angeles, Bogota, London, Mexico City, Jakarta. The rSquared measures the amount of variance around the fitted values - if how much more closer to 1 the most the independent variable influences the dependent one.

3.3 Network Rankings and Prediction

By observing the Overlap Census information we predicted that Beihai would rank among the top 5 positions, since the other 4 cities that are more car-centric - Figure 3 - are all on top. We also predict that Manhattan would rank between the 7th and 11th since the other 3 mobility active cities ranked together between those positions.

4 Conclusion

4.1 Results Analysis

4.1.1 Population X Network

The only real relation we could find between the population and the networks is that every city's Rail and Bike networks seem to have, reasonably, a similar amount of nodes regardless of it's population so we can conclude that no city will have a smaller or bigger Rail and Bike networks regarding only their population, as you can see on Figure 6 the rSquared for the bike and drive networks in both graphs is very close to zero, which means that population doesn't influence those networks.

Besides that, we couldn't find a pattern between the population and the rest of each city's networks, probably because there are many more factors in game, like the cities' area, economic situation and public transportation.

4.1.2 Network Rankings and Prediction

We found that, in general, the cities' congestion level grows with the Pedestrian-Drive network overlap census, as shown in Figure 7. The rSquared in the graph is over 0,6 which means the overlap census may have a moderate-high influence in cities' traffic congestion level.

We also noticed that there is a pattern between the traffic congestion level (ranking and info shown in Table 3) and the car-centric cities - Figure 3 - tend to have a higher congestion percentage when compared to other types of cities. We also conclude that mobility active cities - Figure 5 - tend to have a similar amount of traffic congestion (22-28%). For the car-centric cities that also have pedestrian pathways - Figure 4 - we weren't able to find a consistent pattern since the traffic congestion levels differ to much (17-36%). For the cities with a predominant bike pathways - Figure 2 - we weren't able to conclude anything because we have only data from one city of this type.

We believe that our predictions can be a little bit off because, as spoken before, there are other factors that we aren't taking into account like, in this specific case, the cities' public transportation availability, economy, lifestyle and average age.

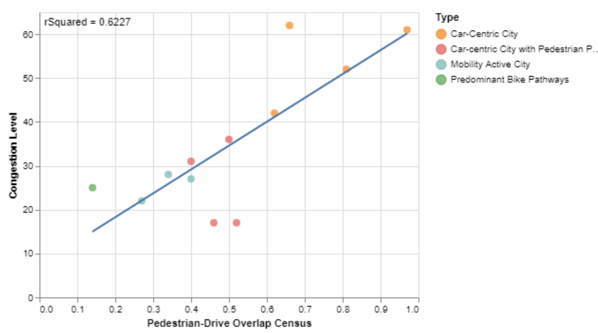


Figure 7: Relation between the cities' Pedestrian-Drive network overlap census and their Congestion Level. Each color represents a different city group by overlap census (Groups made in 3.2). The blue line is a linear regression that represents the growth of the cities' congestion level with the growth of their Pedestrian-Drive network's overlap census. The rSquared measures the amount of variance around the fitted values - if how much more closer to 1 the most the independent variable influences the dependent one.

4.2 Future Work

To continue this work, we think it would be very interesting to dig deeper into these topics by trying to figure out more about the influence of the overlap census in a city's traffic congestion as well as extracting more valuable data like the cities' public transportation schedule and frequency, area, economy, lifestyle and average age in order to get more accurate results and to see if we could discover a pattern here, be it for the population and networks's correlation or for the cities' congestion rankings.

References

[1] L. G. Natera Orozco, F. Battiston, G. Iñiguez, and M. Szell, "Extracting the Multimodal Fingerprint of Urban Transportation Networks," *Findings*, 6 2020.

[2] TomTom, "Traffic congestion ranking: TomTom Traffic Index."

[3] L. G. Natera Orozco, "Multiplex urban networks," 2019.

[4] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[5] NetworkX developer team, "NetworkX," 2014.

[6] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[7] J. VanderPlas, B. Granger, J. Heer, D. Moritz, K. Wongsuphasawat, A. Satyanarayan, E. Lees, I. Timofeev, B. Welsh, and S. Sievert, "Altair: Interactive statistical visualizations for python," *Journal of open source software*, vol. 3, no. 32, p. 1057, 2018.

[8] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference* (Stéfan van der Walt and Jarrod Millman, eds.), pp. 56 – 61, 2010.

[9] L. G. Natera Orozco, F. Battiston, G. Iñiguez, and M. Szell, "Extracting the multimodal fingerprint of urban transportation networks." `https://github.com/nateraluis/Multimodal-Fingerprint`, 2020.

[10] D. Dias, G. Pires, and M. Figueiredo, "Studying urban transportation networks using a multi-layered approach." `https://github.com/GuiFPires/ARCProject2021`, 2021.