A visual data-driven and network-based tool for transportation planning and simulation

Michele Ferretti
King's College London
London, United Kingdom
michele.ferretti@kcl.ac.uk

Luca Pappalardo
University of Pisa
ISTI, CNR
Pisa, Italy
lpappalardo@di.unipi.it

Gianni Barlacchi
University of Trento,
SKIL, Telecom Italia
Trento, Italy
barlacchi@fbk.eu

Bruno Lepri
Fondazione Bruno Kessler
Trento, Italy
lepri@fbk.eu

ABSTRACT

The availability of massive data describing human mobility offers the possibility to design simulation tools to control and improve transportation systems. In this perspective, we propose a visual and data-driven simulation tool based on a multiplex network representation of mobility data, where every layer describes people's movements with a given transportation mode. We then develop a visual application which provides an easy-to-use interface to explore the mobility fluxes and the connectivity of every urban zone in a city. Our application allows the user to visualize changes in the transportation system resulting from the addition or removal of transportation modes, urban zones and single stops. We show how our visual application can be used to explore mobility in Singapore, by using data provided by the CIKM challenge 2017 and mobility data obtained from external sources. The application allows to simulate the reaction to changes in the public transportation system and to assess the resilience of the transportation network to the removal of single subway/bus stops.

KEYWORDS

urban science, data science, human mobility, complex systems, network science, multiplex networks

ACM Reference Format:

Michele Ferretti, Luca Pappalardo, Gianni Barlacchi, and Bruno Lepri. 2017. A visual data-driven and network-based tool for transportation planning and simulation. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 PROBLEM STATEMENT

Nowadays the availability of massive data describing human movements allows us to face relevant urban computing challenges [2, 4, 6, 7, 15]. For example, the observation of mobility flows offers the possibility to investigate the resilience of urban transportation systems, thus uncovering weak points and potentially sub-standard routes. By combining the methods from machine learning and network science, we can design powerful models and simulation tools for what-if analysis of different urban planning scenarios [12, 13].

Specifically, Singapore is a global city where such tools would be particularly valuable, due to the complexity of its transportation system [11]. Despite Singapore's renowned efficiency, its transport services still face daily challenges which might undermine its economy and negatively impact the well-being of its inhabitants. Common problems in the public transportation system are related to a non-optimal positioning of bus stops or subway stations; to prolonged waiting times at such stops; or to misaligned interconnections and inter-modal routes between different transportation networks (e.g., subway and bus). Further, as recently empirically demonstrated by Xu and González [16], a slight re-routing of a fraction of daily rush-hour car commutes across metropolitan areas produces more-than-proportional reductions in traffic, alleviating the overall transport system's congestion state. In this perspective, as highlighted in the Intelligent Transport System Strategic Plan for Singapore [1], the development of big data analytics tools can help to control the transportation system and improve both the customer's travel experience and the system's overall efficiency. Starting from these considerations, we address the following questions: *(i)* what and where are the weakest transportation routes in a city? *(ii)* Given some changes in the transportation system, what scenarios are likely to occur and what is their impact on human mobility? In the literature, Çolak et al. [6] investigate the interplay of number of vehicles and road capacity on their routes to determine the level of congestion in urban areas. They explain that the ratio of the road supply to the travel demand can explain the percentage of time lost in congestion. De Domenico et al. [8] show that the efficiency in exploring the transportation layers depends on the layers' topology and the interconnection strengths. Although these works doubtless shed light on interesting aspects about the structure of urban transportation, they do not provide easy-to-use tools for exploring a city's demand for mobility and the efficiency of the related transportation system. Giannotti et al. [10] partly overcome this problem by proposing a querying and mining system (M-Atlas) for extracting mobility patterns from GPS tracks. However, M-Atlas does not allow to investigate the transportation system's resilience with respect to a city's mobility demand.

We propose a visual, data-driven and network-based simulation tool to highlight and explore the weaknesses in a public transportation system. Our tool is based on a multiplex network representation of mobility data [8], where every layer describes people's movements with a given transportation mode, e.g., buses, metros, taxis. A node in a layer represents a zone of the city, edges indicates routes between zones and edge weights indicate the amount of people moving between two nodes in a given time window. From the multiplex network we extract a set of measures indicating a layer's carrying capacity and its ability to satisfy the overall mobility needs in a given urban area [3]. We then develop a visual application which provides an easy-to-use interface to explore the mobility fluxes and the connectivity of every urban zone in a city. Our visual application allows the user to visualize changes in the transportation system resulting from the addition or removal of transportation modes, urban zones and single stops. We show how our visual system can be used to explore human mobility in Singapore, by using data provided by the CIKM challenge 2017 and Singapore mobility data obtained from external sources. The application allows to point out weak routes among urban areas in the city (i.e., routes where public transportation does not meet the needs of the city users), and simulate changes in the capacity of public transportation to satisfy needs of citizens when specific events occur in the city, e.g., closing/adding transportation modes or subway/bus stops. Our approach is highly flexible since it uses only data about transportation and mobility flows. Given that many open datasets of such nature are publicly available [1], our approach can be potentially applied to any other urban area to simulate traffic changes due to specific events, such as the impact of adding or removing transportation modes or stops, the impact of closing the access to an urban area, or the organization of city-wide public events.

2 DATA SOURCES

We use heterogeneous data sources to simulate, by using our visual tool, transportation changes in the city of Singapore. In particular, we use data about bus lines available at the website www.mytransport.sg [2]. For every bus line, we retrieve information about its stops and the GPS traces describing the bus route. For every stop, we retrieve its GPS position. The bus lines data provide information to build a transportation network describing the displacements of inhabitants between different zones of Singapore. We split Singapore in urban zones by using the shape files provided at the website data.gov.sg [3], where different administrative divisions of the city are provided. We use the most fine-grained division and assign every bus stop to the corresponding urban zone. We hence obtain for every urban zone z: the bus stops z contains, the bus lines passing through z and all the urban zones connected to z. We define two urban zones z_1 and z_2 to be connected if there is at least one bus line connecting z_1 and z_2. Finally, we use data indicating both the presence of people in every urban zone and the fluxes of people between urban zones at a given date and time, downloaded from the API provided by DataSpark[4] for the CIKM AnalytiCup 2017 [5]. We use these data to estimate the number of people moving by bus between two urban zones in a given time window, since official information about the number of users traveling on the buses is not available.

3 METHODOLOGY

In an urban area, two zones can be connected through several transportation means (bus, taxi, subway, etc.). To express this kind of information we introduce the concept of urban multiplex network:

Definition 3.1. An **Urban Multiplex Network (UMN)** is a network in which two nodes represent zones of an urban area and can be connected, at the same time, by multiple edges that belong to different dimensions. We model such structure with an edge-labeled multi-graph denoted by $G = (V, E, L)$ where: V is a set of nodes (urban zones); L is a set of labels (public transportation means); E is a set of labeled edges, i.e., a set of triples (u, v, d) where $u, v \in V$ and $d \in L$ is a label. We use the term *dimension* to indicate a *label*.

The multidimensional connectivity of two zones in an urban area is a combination of two elements: connection intensity and connection redundancy [14]. We define the intensity of the connection between two zones on a single dimension as:

Definition 3.2. **Connection intensity**

$$h_d(u, v) = w_d(u, v)\frac{|\Gamma_d(u) \cap |\Gamma_d(v)|}{min(|\Gamma_d(u)|, |\Gamma_d(v)|)}, \tag{1}$$

where $w_d : V \times V \times L \to \mathbb{N}$ is a weight function representing the mobility flux between two zones on dimension d, and Γ_d is the set of neighbours of a zone.

Connection intensity consists hence of two factors: the first factor, w_d, indicates how many people move between the two zones using transportation layer d; the second factor, $\frac{|\Gamma_d(u) \cap |\Gamma_d(v)|}{min(|\Gamma_d(u)|, |\Gamma_d(v)|)}$, is the percentage of common neighbours, $|\Gamma_d(u) \cap |\Gamma_d(v)|$, with respect to the most selective zone, $min(|\Gamma_d(u)|, |\Gamma_d(v)|)$. The idea is that, on each dimension, the connection intensity is influenced by both the number of displacements between the two zones, weighted by the value of selectiveness of the more selective zone, i.e., the probability that the cluster shared by the two zones is the main one for the zone with the smallest set of neighbours. The second element of multidimensional connectivity is connection redundancy, which takes into account the relevance of a dimension for a zone, i.e., to what extent the removal of the links belonging to a dimension affects the capacity to reach a zone's strong connections.

[1] https://data.gov.sg/group/transport
[2] https://www.mytransport.sg/content/mytransport/home/dataMall.html
[3] https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-web
[4] https://datasparkanalytics.com/
[5] http://cikm2017.org/CIKM_AnalytiCup_task2_dataset.html

Definition 3.3. **Connection Redundancy**

$$r_d(u, v) = (1 - DR(u, d))(1 - DR(v, d)), \tag{2}$$

where dimension relevance DR is the fraction of neighbours that become directly unreachable from a zone if all the edges in a specific dimension were removed [5]. We give a higher score to the edges that appear in several dimensions, so we are interested in the complement of those values. If the two areas are linked in more than one dimension, the score is raised until a maximum of 1. We combine connection intensity and connection redundancy taking into account the multidimensionality of connectivity: a greater number of connections on different dimensions is reflected in a greater chance of having a strong connectivity.

Definition 3.4. **Connectivity.** Let $u, v \in V$ be two nodes and L be the set of dimensions of an urban multiplex network $G = (V, E, L)$. The connectivity of two urban areas u, v is defined as:

$$c(u, v) = \sum_{d \in D} h_d(u, v)(1 + r_d(u, v)). \tag{3}$$

The measure proposed can be used to estimate the strength of the connection also in mono-dimensional networks, where r_d is zero and the overall sum is h_d.

4 APPLICATION DESIGN

The overall application, comprising both the network representation and the interactive interface, is currently running on private hosting solution. An *ad hoc* release to interested third parties and potential collaborations might be considered in the future. The application design follows closely the analytical framework described in Section 1, while implementing a client-server architecture pattern. The back-end component of such structure is responsible for implementing and serving the models presented in Section 3, which the client application then consumes via a REST service exposed by the same server. This API is responsible not only for data provisioning, but also acts as the communication layer between the user and the network models. It is worth noting that, given its modularity, our application can be re-purposed as an agnostic provider of services to other consumers. In particular, the exposed methods consist in:

- a query endpoint returning the network features' geometries for a given urban zone ID, and additional information used to populate the geographic map application;
- a second query endpoint returning for each given urban zone ID the computed network metrics, i.e., *Connection intensity, Connection Redundancy, Multidimensional Connectivity* (Section 3).

The front-end application is thus the entry-point for quickly interacting and prototyping future transportation scenarios. The User Interface (UI), visible in Figure 1, has been developed with easiness of use and clarity of interpretation its standard pillars. It allows a non-technical audience to inspect the number of routes connecting an urban zone to the rest of the city simply by clicking on an urban zone in the city panel (Figure 2), and control them via an interactive menu (Figure 1). Upon addition/removal of one or more routes in the menu, the user can trigger the calculation of the network metrics, which are promptly displayed in three separate windows. Every window shows a 3D map of the city, where an urban zone's height is proportional to its average networks value computed over the connected urban zones (Figure 3). The menu and the bottom windows allow the user to simulate how the city's connectivity changes after, for example, the construction of a new route or the temporary closing of an existing one.

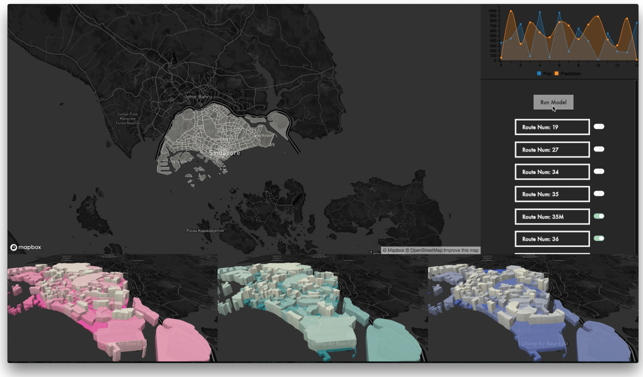

Figure 1: The Application User Interface (UI) displaying the 3D network metrics for the city of Singapore. The interactive menu on the right allows to select and deselect transportation routes. The "Run Model" button allows to calculate the intensity, redundancy and connectivity measures and visualize them in the three bottom windows.

Figure 2: The city panel of the application user interface. It visualizes all the urban zone in the city (Singapore). When clicking on a urban zone, the system shows in the menu all the bus routes passing through that urban area.

All of the geographic map components are fully interactive and built with the latest web-mapping technologies built on WebGL standards with 3D capabilities; as such, they allow for a most fluid and seamless experience. This is a crucial factor that allows the application to not only hide the complexity of the models, but also lets the technology move out the background, making space for generating discussions and streamline decision making processes.

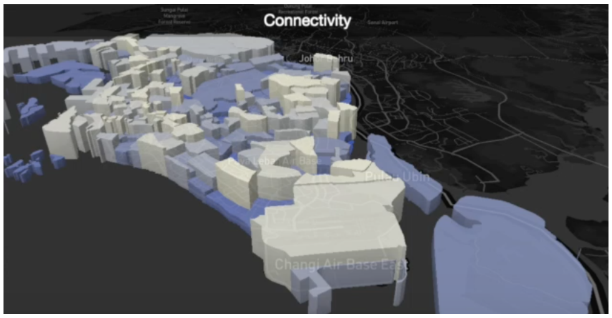

Figure 3: The 3D map in the rightmost bottom window. The height of an urban zone (in Singapore) is proportional to the average value of its multidimensional connectivity, computed across all the connected urban zones.

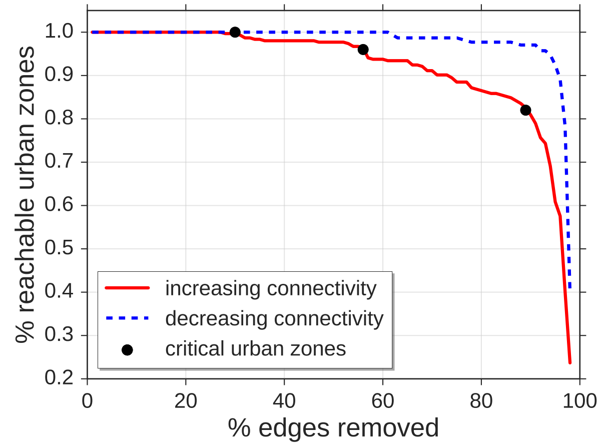

Figure 4: The stability of the urban network to link removal. The x axis shows the percentage of removed links. The y axis shows the size of the greatest network component.

5 EXPERIMENTS

We conduct an extensive connectivity assessment by observing Singapore's urban network resilience to the removal of high or low connectivity links. Figure 4 shows how the relative size of the largest network component changes with the removal of a given percentage of links, sorted by increasing (red solid line) or decreasing (blue dashed line) connectivity order. We find that Singapore has a resilient urban network, as the all nodes are still reachable when removing up to 30% of the links. However, the deletion of links in decreasing order of connectivity affect less the network's global connectivity, as more than 90% of the urban zones are still reachable after the removal of almost all the links (Figure 4, blue dashed line). In contrast, when deleting the links in increasing order of connectivity the network crumbles faster, as almost 20% of urban zones become unreachable after the removal of 90% of the links (Figure 4, red solid line). These results suggest that the proposed connectivity metrics can profitably be deployed to discover those urban connections whose existence is crucial for the network resilience. Moreover, those metrics can also help to evaluate the impact of changing in the city's mobility (e.g., the closure of a bus line). Figure 4 highlights the points where the accelerated network disassembly commences: up to those values the transportation system exhibits a fair resilience, but surpassing such thresholds provokes a rapid network fall-out. Further, while our simulated experiment has been conducted, due to time constrains, only on the bus transportation layer, it is straightforward to envisage and implement in practice a more comprehensive simulation. Testing the network resilience to the above stress conditions in a truly multi-modal perspective represents thus an important tool at the disposal of transportation planners to assess the current state of the transportation network; plan future operations; and keep running the system at its overall optimal capacity.

Acknowledgements. This work has been partially funded by the European project SoBigData RI (Grant Agreement 654024).

REFERENCES

[1] Intelligent transport system strategic plan for singapore. https://www.lta.gov.sg/ltaacademy/pdf/J15Nov_p04Chin_SmartMobility2030.pdf.
[2] Mohammed N Ahmed, Gianni Barlacchi, Stefano Braghin, Francesco Calabrese, Michele Ferretti, Vincent Lonij, Rahul Nair, Rana Novack, Jurij Paraszczak, and Andeep S Toor. A multi-scale approach to data-driven mass migration analysis. In *SoGood@ ECML-PKDD*, 2016.
[3] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.
[4] Gianni Barlacchi, Marco De Nadai, Roberto Larcher, Antonio Casella, Cristiana Chitic, Giovanni Torrisi, Fabrizio Antonelli, Alessandro Vespignani, Alex Pentland, and Bruno Lepri. A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific data*, 2:150055, 2015.
[5] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi. Foundations of multidimensional network analysis. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 485–489, July 2011.
[6] Serdar Çolak, Antonio Lima, and Marta C González. Understanding congested travel in urban areas. *Nature Communications*, 7:10793, 2016.
[7] Manlio De Domenico, Antonio Lima, Marta C González, and Alex Arenas. Personalized routing for multitudes in smart cities. *EPJ Data Science*, 4(1):1, 2015.
[8] Manlio De Domenico, Albert Solé-Ribalta, Sergio Gómez, and Alex Arenas. Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences*, 111(23):8351–8356, 2014.
[9] Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, and Roberto Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal*, 20(5):695–719, October 2011.
[10] S. Jiang, J. Ferreira, and M. C. González. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219, June 2017.
[11] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C. González. The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, 2016.
[12] Luca Pappalardo, Salvatore Rinzivillo, and Filippo Simini. Human mobility modelling: Exploration and preferential return meet the gravity model. *Procedia Computer Science*, 83:934 – 939, 2016. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.
[13] Luca Pappalardo, Giulio Rossetti, and Dino Pedreschi. How well do we know each other? detecting tie strength in multidimensional social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1040–1045, Aug 2012.
[14] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6, 09 2015.
[15] Yanyan Xu and Marta C. González. Collective benefits in traffic during mega events via the use of information technologies. *Journal of The Royal Society Interface*, 14(129), 2017.