Empirical Study of Long-Range Connections in a Road Network Offers New Ingredient for Navigation Optimization Models

Pu Wang[1], Like Liu[1], Xiamiao Li[1], Guanliang Li[2] and Marta C. González[3,4*]

[1]*School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, 410000, P.R. China*

[2]*Center for Polymer Studies, Boston University, Boston, MA, 02215, USA*

[3]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*

[4]*Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*

Navigation problem in lattices with long-range connections has been widely studied to understand the design principles for optimal transport networks, however, the travel cost of long-range connections was not considered in previous models. We define long-range connection in a road network as the shortest path between a pair of nodes through highways and empirically analyze the travel cost properties of long-range connections. Based on the maximum speed allowed in each road segment, we observe that the time needed to travel through a long-range connection has a characteristic time T_h~29min, while the time required when using the alternative arterial road path has two different characteristic times T_a~ 13min and 41min and follows a power-law for times larger than 50min. Using daily commuting OD (origin-destination matrix) data, we additionally find that the use of long-range connections helps people to save about half of the travel time in their daily commute. Based on the empirical results, we assign more realistic travel cost to long-range connections in two-dimensional square lattices, observing dramatically different minimum average shortest path <*l*> but similar optimal navigation conditions.

1. Introduction

While each city has its specific constraints in geography, history and socio-economic mechanisms that shape its structure [1, 2], the road networks from very diverse cities, such as Brasilia, Cairo, London and Los Angeles, have similar topological properties measured in terms of the efficiency and the total length of the entire network [3, 4]. To understand road network topology, complex network studies [3-8] so far have investigated connectedness [4], spatial accessibility [5], price of anarchy [6] and betweenness centrality [7-8]. As another important issue in this field, navigation problem has also drawn much attention in recent years [9-19]. Theoretical works dedicated to the problem of navigation explored the network topology with optimal transport performance, where the average shortest path length $< l >$ is usually the navigation variable to be optimized [10]. On square lattices these studies discovered the strategies to minimize the average shortest path length by adding long-range connections [9-14]. Yet, a square lattice with long-range connections, where all links have exactly the same travel cost, is unrealistic and not able to fully represent the properties of some actual transport networks [20].

Similar to a two-dimensional square lattice with long-range connections, the road network in modern cities is typically composed of two layers: one layer is the *highway layer* formed by highways and the other layer is the *arterial layer* formed by arterial roads (figure 1(a)). The highway layer (with high speed limit) works like the long-range connections, providing fast channels for long-distance travels (figure 1(b)). The arterial layer (with low speed limit) has functions similar to a two-dimensional square lattice, densely spreads across the whole region, connecting a location with its periphery areas (figure 1(c)). In this study, we empirically analyze the properties of the Bay Area road network to understand how long-range connections are embed on the underlying arterial layer in an urban road network, which has been shaped by complex mechanisms such as geography, history and socio-economy for a long time. Using the daily home-work commuting OD data, we further predict the traffic flows in the road network and investigate the functionality and the usage patterns of the long-range connections. Finally, based on the empirical results we improve the two-dimensional square lattice model by assigning more reasonable travel cost to long-range connections.

2. The long-range connections in the Bay Area road network

The Bay Area road network is provided by NAVTEQ, a commercial provider of geographical information systems (GIS) data [21]. The data encapsulate the attributes of roads, such as length and speed limit. In this road network, each link represents a road segment (24,408 in total) and each node represents an intersection (11,309 in total). To have a preliminary understanding of the network properties, we first measure the length l and the free travel time t (length divided by speed limit) of each road segment. We observe that most road segments are densely located in the cities, having a small length l, while a few long road segments sparsely distribute in rural areas, having a length $l > 10$ miles (figure 1(d) and (e)). The longest arterial road segment and the longest highway road segment are roughly 15 miles and 6 miles respectively. However, given the arterial roads' lower speed limit, the maximal free travel time of the arterial road segments is over 35 minutes, which is four times larger than that of highway road segments (figure 1(f) and (g)). The distributions of length l and free travel time t are also plotted in log-log graphs (insets of figure 1(d)-(g)). The length l and free travel time t of road segments follow power-law distributions in wide ranges, showing similar topological features with many practical networks, from the airline transportation network [22] to human communication networks [23].

The long-range connection in the road network is not as obvious as that in a square lattice. Some highway road segments do not share road intersections with arterial road segments, thus they fail to define shortcuts. We explore the long-range connections in a road network by first finding *connecting* *nodes*, which are the intersections connecting both arterial roads and highways. Consequently, we define a long-range connection as the shortest path (measured in travel time) in the highway layer between a pair of connecting nodes. Similarly, we define the long-range connection's alternative arterial road path as the shortest path between the same pair of connecting nodes through the arterial layer. The shortest paths are calculated by the Dijkstra algorithm [24]. The times needed to travel through a long-range connection and its alternative arterial road path are denoted as T_h and T_a, where T_a is a similar measurement with the Manhattan distance r_{ij} in a two-dimensional square lattice [10]. A long-range connection or its alternative arterial road path is constituted by one or several road segments of the same kind. As shown in figure 2(a), intersection A and intersection B are two connecting nodes that connect both arterial roads and highways, the long-range connection from A to B is highlighted by the thick purple line (highway road segments h1, h2, h3, h4, h5) and its alternative arterial road path is highlighted by the thick blue line (arterial road segments a1, a2, a3, a4).

Measuring travel time T_h and T_a between each pair of connecting nodes, we find that 92% of the long-range connections have alternative arterial road paths (8% of them serve as the only path). An important distinction is that in previous works on a square lattice, all long-range connections have the same travel cost regardless of the Manhattan distance r_{ij} between their two endpoints [9-14]. However, in the studied road network the average travel times $< T_h >$ and $< T_a >$ are 31.36 minutes and 54.15 minutes respectively, implying that in average the time cost when we use a long-range connection is about 58% of that cost when we use its alternative arterial road path. Interestingly, not all long-range connections have shorter travel times than their alternative arterial road paths ($T_h < T_a$), we observe that 16% of the long-range connections have $T_h > T_a$. This could be resulted from highways' limited spatial coverage (see figure 1(b)), which generates time consuming detours (figure 2(b)).

We next analyze the probability density functions (PDF) of T_h and T_a. As figure 2(c) and (d) shows, the travel time T_h follows a Gaussian distribution (fit 1) with a characteristic time $T_h \sim 29$ minutes, while the travel time T_a has two different characteristic times $T_a \sim$ 13min and 41min and can be approximated by two different fitting functions for large and small T_a (dashed lines are plotted to guide the eyes):

fit 1: $P(T_h) = 0.023\exp(-\left(\frac{T_h - 28.8}{26.0}\right)^2)$

fit 2: $P(T_a) = 0.009\exp(-\left(\frac{T_a - 12.7}{10.1}\right)^2) + 0.011\exp(-\left(\frac{T_a - 40.8}{29.7}\right)^2)$ when $T_a \leq 50$ minutes

fit 3: $P(T_a) = 21{T_a}^{-1.9}$ when $T_a > 50$ minutes

According to these empirical results, first we can conclude that using T_h to quantify a long-range connection's travel cost is more realistic than assuming the travel cost to be unit for all shortcuts. Next we find that the distribution of the travel time T_a decays much slower than following fit 2, which could be caused by the time consuming detours in the alternative arterial road paths.

3. The usage patterns of the long-range connections

To quantify the effect of long-range connections in actual road usage, we use the Bay Area daily home-work commuting OD data. The OD data are provided by US census bureau [25] and record the number of trips from residents' home locations to work locations at a street-block level. The highly refined spatial resolution creates too many zones, thus we group street blocks into the census tracts (1,398 in total) they are located in and generate the OD in a census tract resolution. As figure 3(a) shows, the number of trips between a pair of OD follows a power law distribution $P(n) \sim n^{-2.88}$, implying that trips are heterogeneously distributed between origins and destinations.

In people's daily commuting, they use different transportation modes which include car (drive alone), carpool, public transportation, bicycle and walk. Based on the mode split data [26], we calculate the vehicle using rate (VUR) in a census tract as follows: $\mathrm{VUR}(i) = P_{car\ drive\ alone}\,(i) + P_{carpool}\,(i)/S$ where $P_{car\ driver\ alone}\,(i)$ and $P_{car\ pool}\,(i)$ are the probabilities that residents in census tract *i* drive alone or share a car (the average carpool size $S = 2.25$ in California [27]). We randomly assign the transportation mode (vehicle or non-vehicle) to the residents living in each census tract according to the calculated VUR. We then filter out the trips that are not made by vehicles.

To assign trips to the road network, we map each OD pair from census tract based OD to intersection-based OD. We find the road intersections within a census tract and randomly select one intersection to be the origin or destination in the intersection-based OD. When no intersection is found in a census tract, we assign a trip's origin or destination to a randomly chosen intersection in the nearest neighbouring census tract. With the intersection based OD calculated, we use the Dijkstra algorithm [24] to find the path with the shortest travel time $T(\text{all})$ between the origin and destination of each trip and calculate the traffic flow in each road segment. In figure 1(a) and figure 3(b), we show the estimated traffic flow, which follows a power law distribution $P(V) \sim V^{-1.48}$.

To better understand the functionality of the long-range connections in people's daily commute, we try to find the shortest path in the arterial layer for each OD pair and compare the travel time $T(\text{arterial})$ with the shortest travel time using the whole network $T(\text{all})$. For 51% of the trips, we fail to find paths only composed of arterial roads, indicating the vital role that long-range connections play in people's daily commute. For the other 49% of the trips, paths in the arterial layer exist and the ratio of $T(\text{all})$ and $T(\text{arterial})$ is found to peak at 0.5, suggesting that the use of long-range connections can help people save about half of their travel time in the daily commute (figure 3(c)). For the shortest path of each trip, we further analyze the fraction of highway use measured in length and in travel time. As figure 4(a) shows, for 16% of the trips, people use arterial roads only. It is also observed that a driver is unlikely to intensively use arterial roads and occasionally use highways in his/her trip. In another word, a driver normally uses highways to complete a large fraction of his/her trip if he/she uses highways. As for the highway use measured in travel time, we obtain similar results (figure 4(b)). As figure 4(c) & (d) illustrate, the fraction of highway use increases sharply with travel length (travel time) when the trip distance is small and gradually saturates to a value near one as the trip distance keeps increasing. The average fraction of highway use has already reached 65% when the travel distance is only 5 miles, note that highways only represent 25% of the road segments in the Bay Area road network. This indicates that the paths of moderate and long distance trips are dominated by highways, while arterial roads are heavily used in very short trips. This result is consistent with the usage patterns of infinite incipient percolation cluster (IIC superhighways) in the ER networks, the SF networks and the square lattices [5].

4. Optimal navigation condition using more realistic travel cost information

In former models dedicated to the navigation problem in lattices, the travel cost of a long-range connection equals to one regardless of the spatial locations of the underlying nodes it connects, thus highly overestimating the shortcuts' ability to reduce travel length (cost). Yet, in the studied road network the ratio of the travel times using highways and arterial roads peaks at $T_h/T_a \sim 0.5$, indicating that a long-range connection typically saves about half of the travel time comparing to its alternative arterial path (figure 2(b)). Indeed, the long-range connections that connect distant nodes in many transport networks are not so 'short' as previously modelled. It is necessary to explore the optimal navigation conditions and calculate the average path length under more realistic travel cost scenarios.

We generate a regular two-dimensional square lattice with N=1,000,000 nodes, pairs of nodes i and j are then randomly selected to receive long-range connections with probability proportional to the Manhattan distance ${r_{ij}}^{-\alpha}$ (figure 5(a)), where α is the variable exponent controlling the number and the length of long-range connections. The addition of the long-range connections stops when the total length (cost) $\sum r_{ij}$ reaches N. Different from the model presented in Ref. [10], the travel cost of each long-range connection is assigned in our model. We make a reasonable assumption that the travel length (cost) l of a long-range connection scales linearly with the Manhattan distance between the two nodes it connect, which is denoted by $l = \beta r_{ij}$. In a road network the scaling exponent β quantifies the fraction of travel time saved by using highways. As illustrated in figure 5(a) the Manhattan distance between node i and j is six, the travel cost of the shortcut is three when the scaling exponent $\beta = 0.5$.

The optimal conditions were discovered at $\alpha = 0$ and $\alpha = 2$ for navigation using global or local information if no total cost constraint exists in adding connections [9]. The optimal navigation condition was found at $\alpha = 3$ for a system subject to reconstruction cost [10], implying that more short (low-cost) connections are preferred when one has limited resources. Similar to former modelling frameworks, we use the average shortest path $< l >$ as the navigation variable to be optimized. Three scenarios $\beta = 0.5$, $\beta = 0.2$ and $\beta = 0.8$ are studied, which correspond to the cases that long-range connections have moderate, low and high travel cost respectively. Given that links have different travel cost in our model, the shortest path between a pair of nodes is calculated by Dijkstra algorithm [24].

Although different minimum $< l >$ are found for the three scenarios due to the different travel cost of long-range connections, similar optimal navigation conditions are found at $\alpha \sim 3$ (figure 5(b)). Comparing with the minimum average shortest path $< l >$ found by assuming $l = 1$ for all connections, the minimum $< l >$ is much larger for the moderate travel cost scenario, again validating that long-range connections' ability to reduce travel cost was overestimated in previous models. Finally, as the scaling exponent β increases, the differences between the average shortest path $< l >$ at different variable exponent α decrease. When the scaling exponent β reaches one, long-range connections fail to improve the navigation efficiency at all. In conclude, the optimal navigation condition will not dramatically change when adding realistic travel cost to long-range connections, demonstrating the generality of the classic model raised in Ref. [10]. However, adding realistic travel cost to long-range connections will largely improve the accuracy of the estimation of $< l >$, indicating that travel cost is an important parameter to be considered when a long-range connection's transport efficiency is comparative with the underlying lattices.

Various technological and natural networks, from transportation networks [8, 20] to social networks [23] and epidemic spreading networks [28, 29], are characterized with two-layer structures. For many of them, the transport efficiency of a long-range connection is comparable with that of short-range connections (e.g. the travel time taking a plane is comparable with the travel time driving a car if the travel distance is small). Therefore, it is necessary to empirically estimate the actual transport efficiency of long-range connections, and build up models that incorporate this important information to understand how the optimal transport condition is affected under different travel cost scenarios. In this study, we find that the travel time using highways (T_h) is about half of that using their alternative arterial paths (T_a) in this real world transportation network, thus this gives us a reasonable justification to assign travel cost in long-range connections. Moreover, the empirical results on the distribution of T_h/T_a and the heterogeneously distributed travel demand allow for detailed information encapsulated in future models. The empirical investigations of the properties and usage patterns of long-range connections in practical networks offer us the way to introduce more realistic link properties and a guidance to generate practical models dedicated to navigation optimization.

5. Conclusions

The optimization of a transport network's navigation efficiency has great impacts not only in the traffic engineering, but also in computer science and information spreading. We define long-range connections in a road network, analyze the time needed to travel through them and the time needed to travel through their alternative arterial road paths, which, we believe can enrich our understanding of the road network structure and provide useful information for transport networks' optimal design. We investigate the navigation problem by building a new model that encapsulates more realistic travel cost information. We find the new optimal transport networks have similar optimal navigation conditions but different average shortest path comparing to the scenario that all connections have equal unit travel cost. Due to the different populations in traffic zones and the different distances between traffic zones, travel demands are always not homogeneously distributed in an urban area [20]. In future models, not only network properties but also travel demands are the necessary ingredients that need to be considered when evaluating or improving a transport network.

The studied road network possesses similar topological features with many practical networks, such as the airline transportation network [22] and human communication networks [23], where the lengths of links also follow power-law distributions. Therefore, the empirical findings of this work could be generalized to this broader set of networks. The travel (transport) cost of long-range connections is also ubiquitous in different kinds of networks. Our model employs a general scaling exponent β to incorporate the adjustable travel (transport) cost of long-range connections into the classic optimal navigation models, which, we believe can provide a general modeling framework for navigation optimization in diverse problems related to network flows in science and engineering.

Acknowledgements

The authors thank A.M. Bayen, T. Hunter and Y. Yang for valuable discussions. The work is supported by the National Natural Science Foundation of China (No. 51208520 and No. 71071165), New England UTC Year 23 grant, awards from NEC Corporation Fund and the Solomon Buchsbaum Research Fund. P. Wang acknowledges support from Shenghua Scholar Program of Central South University.

References

1. J. Jacobs. The economy of cities. *Vintage*, February 12 (1970).
2. M. Batty. The size, scale, and shape of cities. *Science* **319**, 769-771 (2008).
3. M. Barthélemy & A. Flammini. Modeling urban streets patterns. *Phys. Rev. Lett.* **100**, 138702 (2008).
4. M. Barthélemy. Spatial networks. *Physics Reports* **499**, 1–101 (2011).
5. Z. Wu, L.A. Braunstein, S. Havlin & H.E. Stanley. Transport in weighted networks: partition into superhighways and roads. *Phys. Rev. Lett.* **96**, 148702 (2006).
6. H. Youn, M.T. Gastner & H. Jeong. Price of anarchy in transportation networks: efficiency and optimality control. *Phys. Rev. Lett.* **101**, 128701 (2008).
7. M.E.J. Newman. A measure of betweenness centrality based on random walks. *Social Networks* **27**, 39–54 (2005).
8. M. Kurant & P. Thiran. Layered complex networks. *Phys. Rev. Lett.* **96**,138701 (2006).
9. J. M. Kleinberg. Navigation in a small world. *Nature* 406, 845 (2000).
10. G. Li *et al*. Towards design principles for optimal transport networks. *Phys. Rev. Lett.* **104**, 018701 (2010).
11. S. Carmi *et al*. Asymptotic behavior of the Kleinberg model. *Phys. Rev. Lett.* **102**, 238702 (2009).
12. C. Caretta Cartozo and P. De Los Rios, Extended navigability of small world networks: exact results and new insights. *Phys. Rev. Lett.* **102**, 238703 (2009).
13. M.R. Roberson and D. ben-Avraham, Kleinberg navigation in fractal small-world networks. *Phys. Rev. E* **74**, 017101 (2006).
14. M. Boguna and D. Krioukov, Navigating ultrasmall worlds in ultrashort time. *Phys. Rev. Lett.* **102**, 058701 (2009).
15. K. Komidis *et al*. Structural properties of spatially embedded networks. *Europhys. Lett.* **82**, 48005 (2008).
16. C.F. Moukarzel and M. Argollo de Menezes, Shortest paths on systems with power-law distributed long-range connections. *Phys. Rev. E* **65**, 056709 (2002).
17. R. Guimera *et al*. Optimal network topologies for local search with congestion *Phys. Rev. Lett.* **89**, 248701 (2002).
18. M.C. Santos *et al*. Optimization of random searches on defective lattice networks. *Phys. Rev. E* **77**, 041101 (2008).
19. M.T. Gastner and M.E.J. Newman, Optimal design of spatial distribution networks. *Phys. Rev. E* **74**, 016117 (2006).
20. P. Wang *et al*. Understanding road usage patterns in urban areas. *Scientific Reports* **2**, 01001 (2012).
21. Navteq Official Website. http://www.navteq.com/
22. G. Bianconi, P. Pin & M. Marsili. Assessing the relevance of node features for network structure. *PNAS* **106**, 11433 (2009).

23. R. Lambiotte *et al*. Geographical dispersal of mobile communication networks. *Physica A*, **387**, 5317 -5325 (2008).
24. E.W. Dijkstra. A note on two problems in connexion with graphs. *NumerischeMathematik* **1**, 269–271 (1959).
25. http://www.census.gov/geo/www/tiger/tgrshp2010/tgrshp2010.html
26. Travel demand modeling with TransCAD 5.0, User's Guide. Caliper (2008).
27. State averages for private vehicle occupancy, carpool size and vehicles per 100 workers. http://www.nctr.usf.edu/clearinghouse/censusavo.htm
28. D. Balcan *et al*. Mutiscale mobility networks and the large spreading of infectious diseases. *PNAS* **106**, 21484 (2009).
29. P. Wang, M.C. González, C.A. Hidalgo & A.-L. Barabási. Understanding the spreading patterns of mobile phones viruses, *Science* **324**, 1071-1076 (2009).

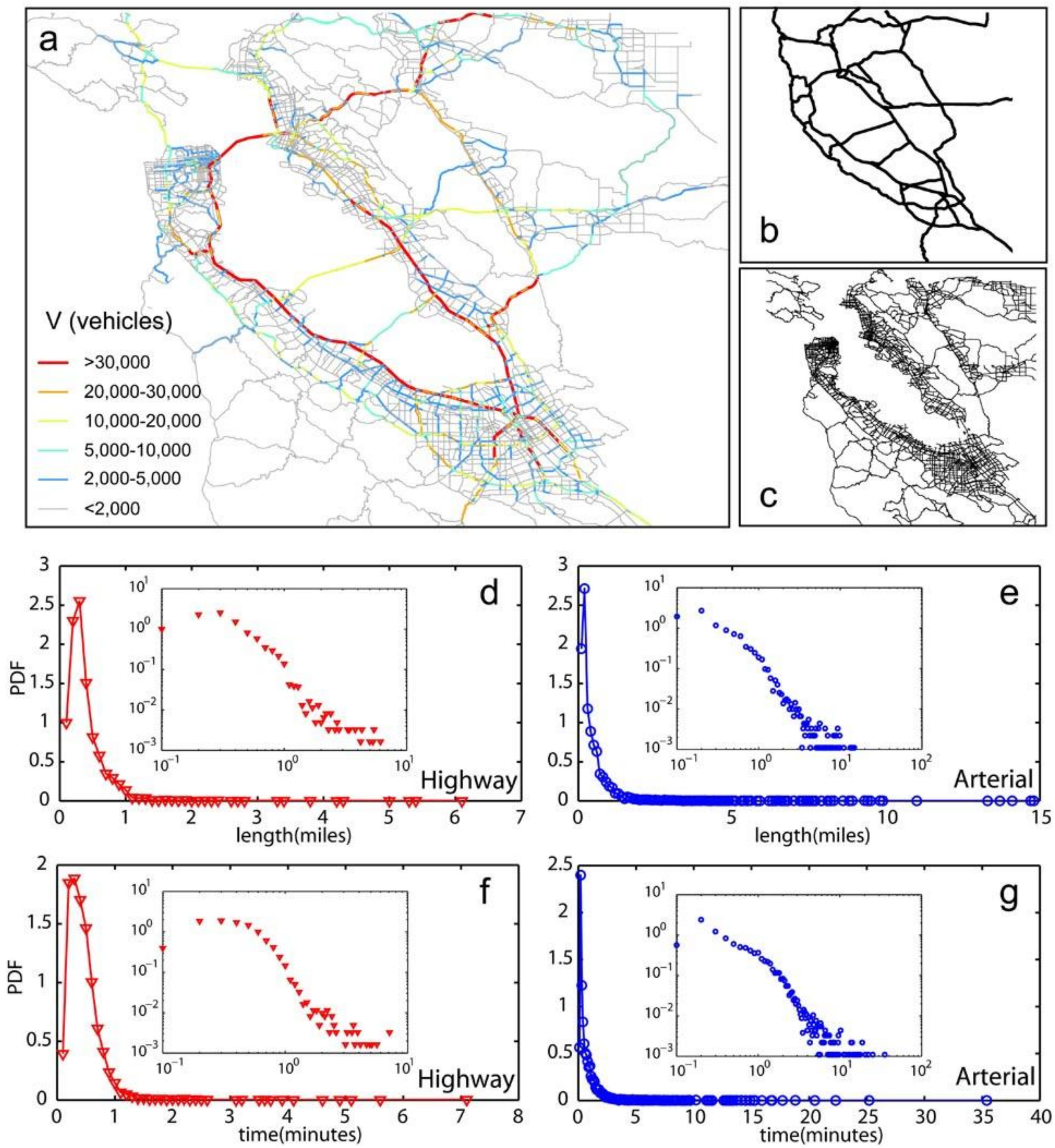

Figure 1. The Bay Area road network. (a) The color and the thickness of a link show the traffic flow in a road segment. The traffic flows are estimated based on the residents' daily home-work commuting OD data. (b) The highway layer formed by highway road segments (6,140 in total). (c) The arterial layer formed by arterial road segments (18,268 in total). (d) (e) (f) (g) The probability density function (PDF) of the length and the free travel time of arterial road segments and highway road segments. The insets show the results in log-log scales.

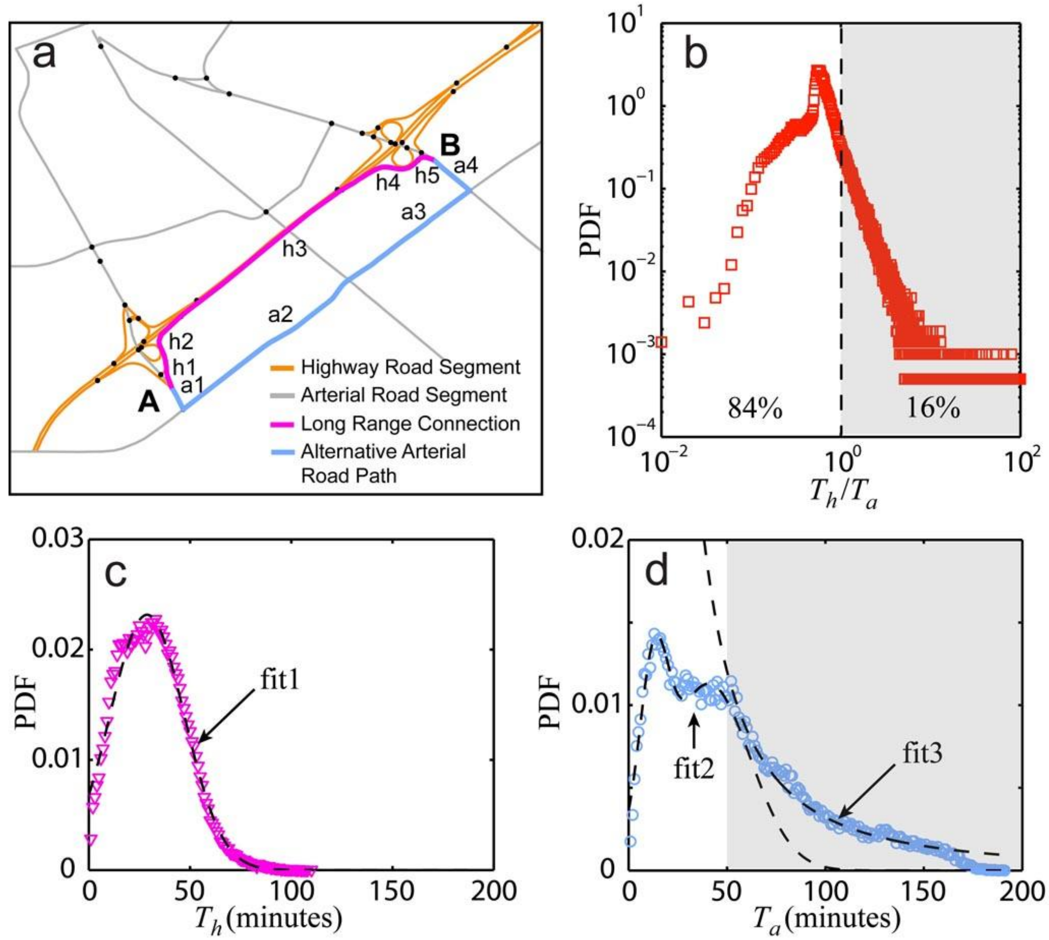

Figure 2. The long-range connections and the alternative arterial road paths. (a) Illustration of the long-range connection between two connecting nodes A and B. Orange links and gray links represent highway road segments and arterial road segments respectively. The purple line (formed by highway road segments h1, h2, h3, h4, h5) is a long-range connection defined in this paper. The blue line is the alternative arterial road path (formed by arterial road segments a1, a2, a3, a4) between A and B. (b) The times needed to travel through a long-range connection and its alternative arterial road path are denoted as T_h and T_a. For 84% of the long-range connections: $T_h < T_a$. (c) The probability density function (PDF) of T_h. (d) The probability density function (PDF) of T_a.

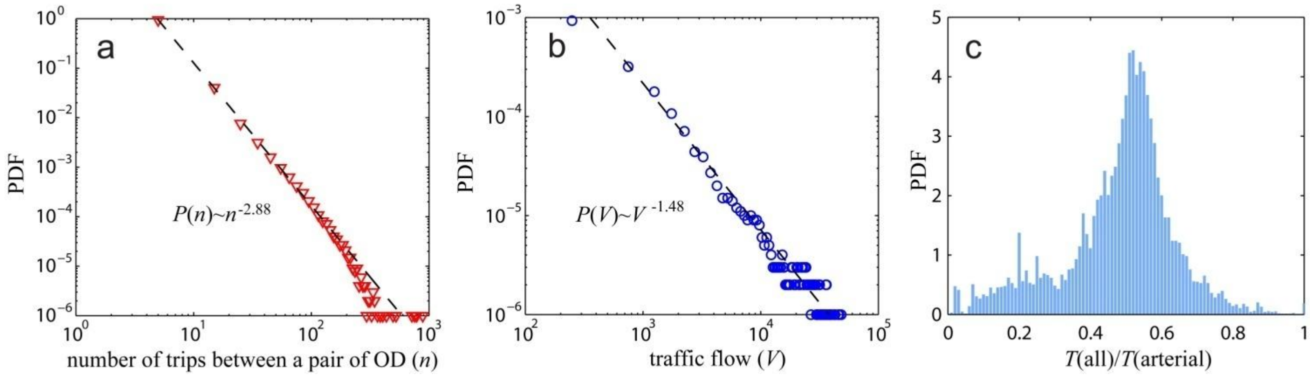

Figure 3. The functionality of the long-range connections is demonstrated by the daily commuting OD data. (a) The number of daily home-work commuting trips between a pair of OD can be well approximated by a power-law distribution. (b) The traffic flow generated by the daily commenting demands follows a power-law distribution. (c) The ratio of the shortest travel time *T*(all) and the shortest travel time in the arterial layer *T*(arterial). PDF represents the probability density function.

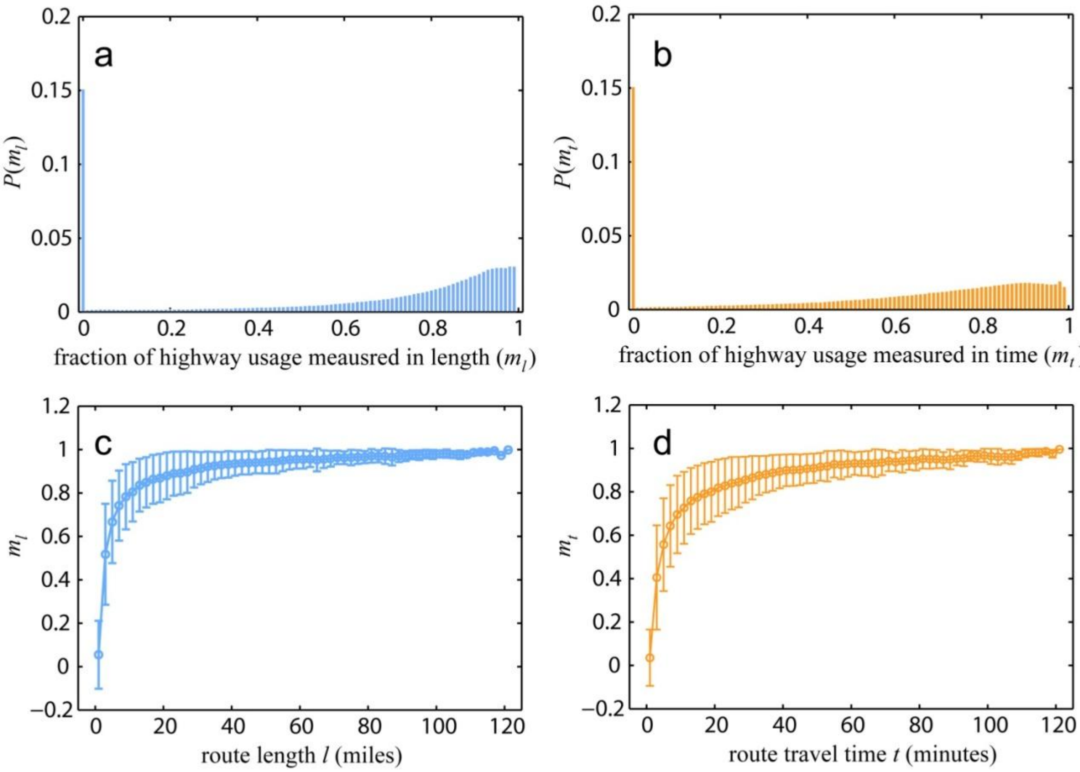

Figure 4. The usage patterns of the long-range connections. (a) The fraction of highway usage (measured in length) in the shortest paths of residents' daily home-work commuting trips. (b) Same as (a) but for the fraction of highway usage measured in travel time. (c) The fraction of highway usage increases with the travel length. The circles represent the average and the error bars stand for the standard deviation. (d) Same as (c) but for the fraction of highway usage measured in travel time.

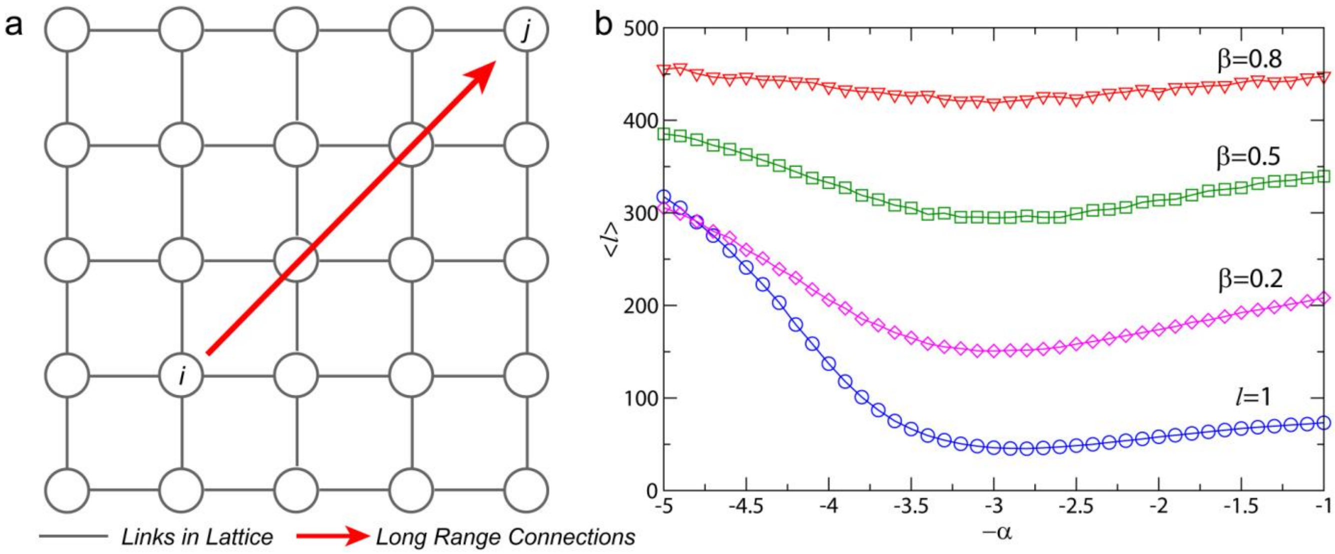

Figure 5. Average shortest path <*l*> at different variable exponents α. (a) A two-dimensional square lattice with long-range connections, where the travel cost of long-range connections is assigned following $l = \beta r_{ij}$. (b) The average shortest path <*l*> at different α values. When the travel cost equals to one for all long-range connections, the minimum <*l*> is found at $\alpha = 3$ (the blue symbols show the average of 100 realizations). When more realistic travel cost is assigned to long-range connections, similar α values are found for the optimal navigation conditions. The purple, green and red symbols represent the results (100 realizations) under the low, moderate and high travel cost scenarios.