**A comparative study of topological analysis and temporal network analysis of a public transport system – case study in Ho Chi Minh City, Vietnam**

Nam Huynh  
Fulbright School of Public Policy and Management, Fulbright University Vietnam  
105 Ton Dat Tien Street, District 7, Ho Chi Minh City, Vietnam  
Email address: nam.huynh@fulbright.edu.vn

**Abstract.** This research investigated the difference in operational characteristics of a public transport network calculated under two equally popular analytical approaches, namely the topological network analysis and the temporal network analysis, using the bus network in Ho Chi Minh City (HCMC), Vietnam as a case study. As existing studies provided little discussion on their adoption of one approach over the other, we aimed to answer the question “What is the degree of discrepancy and/or loss of information, if any, in analyses using the topological network approach compared to those from the supposedly more computationally demanding and time consuming temporal network approach?”. Stop-based accessibility metrics and most used infrastructure (bus stops and bus routes) were calculated, compared and reported for both approaches. Different to many existing studies in which shortest path between stop pairs were computed by optimising only travel time, we adapted the multi-criteria profile connection scan algorithm to minimise both time and number of transfers during the computation. In representing the HCMC bus network, which had 253 directed routes servicing 4,350 stops, the number of nodes and the number of edges in the temporal network of a 24-hour period were a few orders of magnitude larger than the those in the topological network. The computational time for computing shortest paths between all stop pairs in the temporal network over a 3-hour period was at least an order of magnitude longer than that in the topological network. Nonetheless, results from both representations showed similar qualitative trends, in terms of the number of bus stops (and the corresponding average distances) accessible by inner city stops within a given range of travel time. Quantitatively, results from the topological network were almost always larger than those from the temporal network, which could be partly attributed to its inability to capture temporal heterogeneity and intra-day variation in a public transport network operation.

**Keywords.** Public transportation, topological network, temporal network, accessibility, resilience

1. **Introduction**

A user-friendly and efficient public transport (PT) system has been widely regarded as key to effectively tackling urban planning challenges, most notably of which include increased road traffic congestion, worsen air quality, long commute time, less affordable housing, and the reduced overall livability of the urban environment. Public transport systems in numerous cities around the world therefore have attracted a growing number of studies investigating their accessibility and other operational characteristics, especially from a complex network perspective.

Among most common approaches in those studies was the use of graph theoretic concepts in which a PT network was modelled as a graph Shanmukhappa et al. (2019). Various graph spaces exist for different analysis needs. In an L-space graph, a station (e.g. a train station or a bus stop) in the network is represented as a node, and two nodes are connected by an edge if there is at least one PT route connecting them. In a B-space graph, nodes represent both routes and stations. An edge is created only between a route node and each of its associated station nodes. In a P-space graph, nodes represent stations and edges are created between stations that can be travelled to/from each other without transfers. In a C-space graph, nodes represent routes and an edge exists between two nodes if they serve at least one common station. None of the graph spaces explicitly accounts for geospatial attributes of real-world PT networks. Keen readers are referred to studies by Shanmukhappa et al. (2019), Kurant and Thiran (2006), and Ferber et al. (2009) for more detail on graph spaces and their applications in PT network analysis.

L-space graphs were reportedly the most widely used topological representation, primarily thanks to their reflection of the structure of the real-world networks Shanmukhappa et al. (2019). Assumptions behind how features of the real PT networks were mapped to nodes and edges in an L-space graph varied between studies. While earlier studies (Kurant and Thiran (2006), Ferber et al. (2009), Sienkiewicz et al. (2005), Xu et al. (2007), Chen et al. (2007), and Soh et al. (2010)) assumed one to one mapping between a PT station and a node in the L-space graph, more recent studies tended to take a more sophisticated approach. In the analysis of multimodal PT network in five Hungarian cities (London et al., 2015), an L-space node represented duplicating stations (i.e. those sharing similar names or belonging to a major interchange). Similarly, Shanmukhappa et al. (2018) combined stations within 100 metres to each other into a so-called “supernode”, which was then represented by a node in their L-space graphs. An edge was created between two nodes if they were consecutive along at least one timetabled service. Edges were either undirected (Kurant and Thiran (2006), Ferber et al. (2009), Sienkiewicz et al. (2005), Xu et al. (2007), Chen et al. (2007)) or directed (Shanmukhappa et al. (2018), Feng et al. (2016), Chatterjee et al. (2016)), unweighted (Ferber et al. (2009), Sienkiewicz et al. (2005), Xu et al. (2007), Chen et al. (2007), Regt and Ferber (2018)) or weighted by various operational attributes over a specific period, such as an indicative travel time between the corresponding stations (Regt and Ferber (2018)), the total number of passengers travelled on the link (Soh et al., 2010), or an indicative capacity of PT vehicles traversing that link (Kurant and Thiran (2006), Feng et al. (2016), Chatterjee et al. (2016)).

The analysis focus of these studies was on statistical attributes of the network (e.g. topological efficiency, small-world behaviour, average path length, average clustering coefficient, and average degree) and those of network nodes (e.g. degree distribution and various measures of node centrality). For a comprehensive review of topological attributes commonly calculated for PT networks under the complex network theory paradigm, keen readers are referred studies by Shanmukhappa et al. (2019) and by Zhang et al. (2018).

It is noted that while timetables were part of the input data into some of the above studies, they were used primarily for reconstructing the topological graphs (Kurant and Thiran (2006), Soh et al. (2010)) or in calculating (average) travel time as edge weight (Regt and Ferber (2018)). The graphs and their attributes remained temporally static.

The temporally dynamic nature of PT operations was addressed to different extents in studies which investigated PT-based accessibility of urban areas. Salonen and Toivonen [ref], in evaluating door to door accessibility disparity by private vehicles and by PT to various destinations in Helsinki, approximated the variation of the PT schedule by four time points inside and outside of rush hours of a normal weekday. Tenkanen et al. [ref] employed the similar approximation to construct a matrix of travel time by PT, by cars, and walk for all pairs of statistical grid squares in Helsinki. Goch et al [ref] evaluated PT-based accessibility to district and local centres in Warsaw for every hour between 6.00 am and 9.00 am. Instead of calculating PT travel times for only a number of departure times in the day, other studies addressed time-varying PT travel times in a more continuous manner. For example, Lei and Church [ref] integrated bus services timetable as an arc attributes in a GIS as part of an analysis tool which evaluated and compared accessibility by PT and by private vehicles to various parts of the Santa Barbara of California. In studies by Farber et al [ref] and by Owen and Levinson [ref] which evaluated accessibility by PT to supermarkets and to jobs, respectively, the variation of PT travel times was accounted for at all times of the day. Farber and Fu [ref] extended the continuous accessibility calculation in form of a data object called PT travel time cube, which contained the estimated shortest travel time between pairs of traffic analysis zones at different departure times of a representative day. Stepniak et al (2019) provided a comprehensive review of recent studies which included the temporal dimension in evaluating PT-based accessibility, and noted that such inclusion was reportedly computationally expensive especially at high temporal resolutions.

The above review revealed two apparently equally popular and increasingly active approaches to investigating connectivity and accessibility of a PT network, namely topological (temporally static) network approach and the temporal network approach. There appeared to be a trend that most studies adopting the topological network approach were published in journals in the field of physics and complex networks, whereas those adopting the temporal network approach were primarily published in journals dedicated to the field of transportation. Regardless, existing studies provided little discussion, if at all, on their adoption of one approach over the other. Also, accessibility measures were dominantly the focus of many existing research which accounted for the temporal dimension of PT operation. Analysis of temporal variability of PT network elements critical to maintaining the network connectivity, which is an important input towards improving the network’s resilience and robustness, has not attracted a similar level of interest.

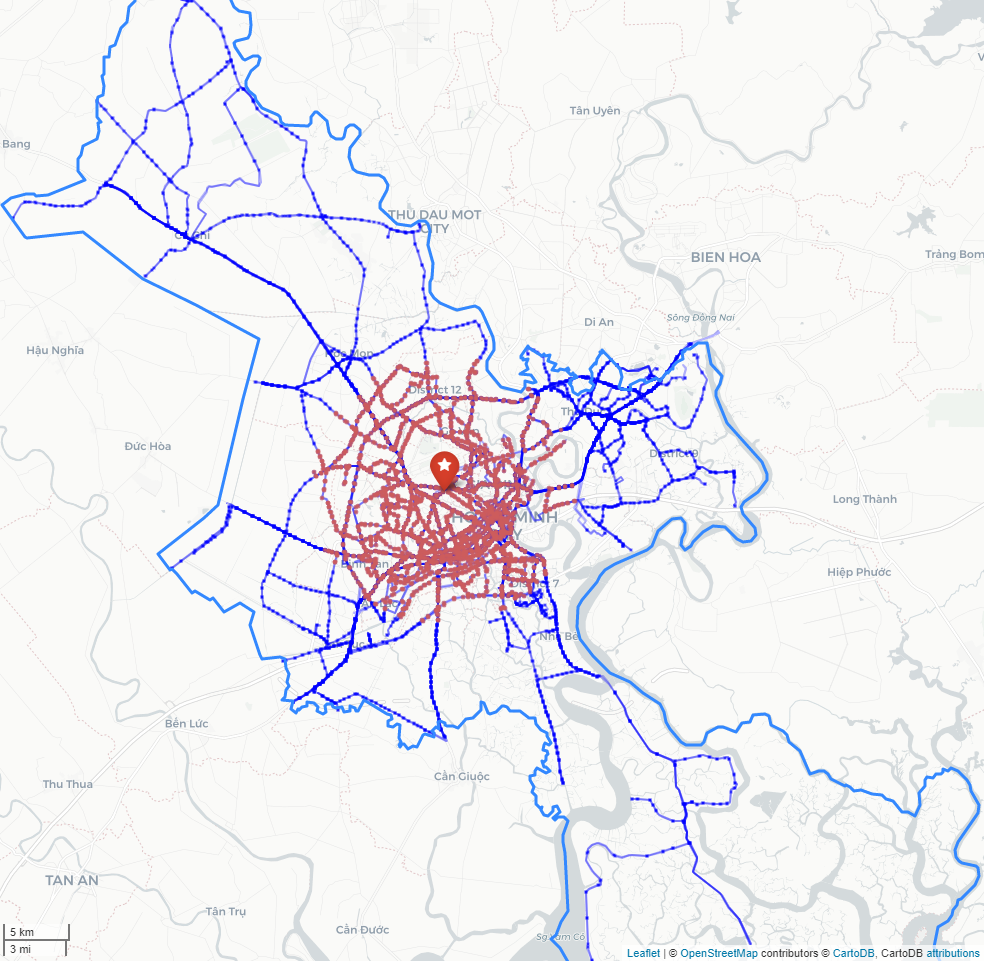
In this work, we examined the accessibility and most-used infrastructure of the bus network in Ho Chi Minh City (HCMC), Vietnam, and provided a direct comparison of these characteristics calculated under the topological network approach and under the temporal network approach. More specifically we aimed to answer the question “What is the degree of discrepancy and/or loss of information, if any, in analyses using the topological network approach compared to those from the supposedly more computationally demanding and time consuming temporal network approach?” For this purpose, we built two representations of the HCMC bus network, an L-space directed and weighted graph and a temporal network model following the methodology by [50] and Kujala et al. (2018).

In this study, the accessibility of a bus stop was measured by stop-level metrics *proximity density* and *proximity average distance*, respectively defined as the number of bus stops within a given total travel time of the bus stop and the corresponding average distance between these bus stops and bus stop . The metrics therefore can be classified as infrastructure-based in the review of accessibility measures by Geurs and Wee (2004). The most-used infrastructure was defined as bus stops and bus routes traversed by the shortest path between the highest number of all stop pairs, and served as a proxy to our analysis of network elements important to maintaining the network connectivity.

Computing shortest paths was instrumental to our analyses. It should be noted that shortest paths in the studies incorporating the temporal dimension of PT networks were computed primarily by minimising travel time and did not explicitly account for transfers (Kujala et al., 2018), which was reportedly among key factors affecting comfort and convenience of passengers and the success of a PT network [44-46]. Kujala et al. (2018) addressed this issue by adapting the multi-criteria profile connection scan algorithm originally developed by Dibbelt et al. [ref] to compute shortest paths in their temporal network model. In this study, we made further changes to this algorithm so that shortest paths in the temporal network were computed at a desired departure time on the basis of earliest departure earliest arrival while minimising both travel time and number of transfers.

The remaining of the paper is structured as follows. Section 2 describes the bus network data available for HCMC, design features of the topological model and the temporal model of the network, and definition of network metrics used in our analyses. Section 3 briefly discusses computational implementation of the network models. Section 4 presents and compares results calculated from the temporal model and the topological model. Please note that while the temporal network model was capable of calculating network metrics at any given departure time, we present only results for departure times representative of the morning peak hours (07.00), the off peak hours (12.00), and the afternoon peak hours (17.00). Conclusion remarks and future research direction are provided in Section 5.

1. **Material and methods**
   1. *Bus network in HCMC as a case study*



**Figure 1.** Layout of directed routes in the HCMC bus network which will be used for analyses in this study. The star marker denotes the centre of mass (i.e. centroid) of all considered bus stops. Plotted in red are bus stops within a 10 km radius of the network centroid, referred to as the inner city bus stops.

Out of the total of 131 bus routes reportedly operational in HCMC (as of March 2020), we removed those primarily for tourist purposes (routes ‘DL01’ and ’72-1’), the water bus route ‘WB01’, and bus route ’70-5’ which had only two bus stops one of which was outside of the HCMC boundary. For the remaining bus routes, we removed any end stops outside of the city boundary. We also noticed that the two directions of each bus route did not always share similar road links, thus treated them as two separate bus routes in our analyses and referred to each as a directed route. Figure 1 presents the layout of the final bus network for our analyses which had in total 253 directed routes servicing 4350 bus stops.

Unlike other PT network analysis studies in which the PT network data were available in the standardised General Transit Feed Specification (GTFS), the bus network data for HCMC used in this study was from the city’s PT information portal <https://busmap.vn>. The information available for each directed route on a typical weekday included the stop sequence, geographical location (latitude and longitude) of each stop, geographical location of path points along road links between consecutive stops, whole-route travel time, and scheduled departure time of each bus service running this directed route throughout the day (i.e. service frequency of each directed route at different times of day). Because the information portal provided only a single value of whole-route travel time of a bus route, we assumed the same travel time for both directions of the route. Stop-by-stop time schedules were also not available, thus needed to be approximated for each directed route from the whole-route travel time and road distance between consecutive stops. In approximating such a stop-by-stop timetable, apart from the apparent assumption that the travel time was constant along each directed route, we also assumed a constant dwell time of 6 seconds for all buses at each bus stop.

* 1. *Topological network model*

The bus network was topologically modelled by means of a directed L-space graph in which each node represented a stop in the timetable and an edge existed between two nodes if the corresponding bus stops were consecutive along at least one directed route. Node attributes included essential bus stop information (e.g. geographical location and description of the stop), the number of bus lines and the total number of bus services (on all lines) passing that stop over a day as specified in the timetable. Edge weights included the average road distance and the average travel time between the two corresponding bus stops by all bus services as specified in the timetable.

The traditional Dijkstra’s algorithm was applied on the directed L-space graph to calculate the shortest path travel time (i.e. using the average travel time as edge weight) between all stop pairs that were not within walk distance of each other. The maximum walk distance assumed in this study was 300 metres. For stop pairs that were within a walk distance of each other, the shortest travel time is calculated by , in which is the geographical distance between the two bus stops and is the constant walk speed, assumedly 1.3 m/s in this study.

* 1. *Temporal network model*

|  |  |
| --- | --- |
|  |  |
| (a) Schematic layout of sample directed routes | (b) Temporal network model of sample directed routes |

**Figure 2.** Elements of the temporal network model

Elements in the temporal network model representing the HCMC bus network is illustrated in Figure 2 by two sample directed routes, each run by 2 bus services. Stop G and stop D were assumed within a walk distance to each other. Different to the L-space graph in which a node simply represented a bus stop, a node in the temporal network represented an event made by a bus service (i.e. arrival or departure) at a bus stop. A temporal network node therefore was uniquely defined by a combination of three elements, the associated directed route, the corresponding bus stop, and the time of the bus event at that stop. These nodes are represented by red circles (‘departure’ nodes) and blue circles (‘arrival’ nodes) in Figure 2b. The solid black segments and the solid green segments are called transit edges and represent bus services along the directed route [A, B, C, D, E] and the directed route [F, B, G, H], respectively. Among these solid segments, the horizontal ones represent dwelling of a bus service at a bus stop, which as mentioned in Section 2.1 was assumed a constant of 6 seconds in this study.

Different to other studies (for example by Kujala et al (2018)) in which transfers were allowed between all stop pairs within walk distances, a transfer edge was defined in our temporal network to model the following transfer types.

* Waiting, i.e. connecting the arrival of a bus service at a bus stop and the departure of another bus service at the same bus stop if the two bus services are of different directed routes and the time of the departure event must satisfy . is the time of the arrival event, and is the maximum time a passenger would wait at a bus stop to board a bus service, assumedly 60 minutes in this study. Wait transfer edges are demonstrated by the two dashed curves connecting nodes at station B in Figure 2b.
* Walking and waiting, i.e. connecting the arrival of a bus service at a bus stop and the departure of another bus service at another bus stop if the two bus services are of different directed routes, the two stops are within a maximum walk distance (assumedly 300 metres in this study) and the time of the departure event must satisfy . is the walk time from the arrival bus stop to the departure bus stop calculated by , where is the geographical distance between the two bus stops and is the constant walk speed, assumedly 1.3 m/s in this study. Walking and waiting transfer edges are demonstrated by the three dashed curves connecting ‘arrival’ nodes at stop D to ‘departure’ nodes at stop G in Figure 2b.

A transfer value of 1 was assigned to each transfer edge and of 0 was assigned to each transit edge. Travel time between a stop pair within a walk distance (i.e. no farther than 300 metres in this study) is calculated by as above. In calculating the travel time between a stop pair that are more than 300 metres apart at a desired departure time , the algorithm adapted by Kujala et al. (2018) was closest to ours, except that we narrowed our algorithm to solving the problem of earliest departure – earliest arrival while minimising the number of transfers. Specifically, our algorithm for shortest path travel time can be summarised in the below steps.

**Step 1.** Select ‘arrival’ nodes (i.e. those associated with an arrival event) in the temporal network which belong to the destination stop and have their arrival time .

**Step 2.** For each of these nodes in ascending order of time (i.e. earliest possible arrival time), select ‘departure’ nodes (i.e. those associated with an departure event) in the temporal network which belong to the origin stop and have their departure time satisfying , where is the time at the current ‘arrival’ node.

**Step 3.** For each of these ‘departure’ nodes in ascending order of time (i.e. earliest possible departure time), search for the path connecting the ‘departure’ node with the current ‘arrival’ node that has the smallest number of transfers using the traditional Dijkstra’s algorithm with transfer weight as the edge weight.

**Step 4.** Repeat from Step 3 and then from Step 2 (if necessary) until a path is found.

**Step 5.** If a path is found between a ‘departure’ node and an ‘arrival’ node, record the nodes along this path (which also have information of the corresponding bus stops and directed routes). The total travel time between this stop pair equals the time at the ‘arrival’ node minus the time at the ‘departure’ node, not the desired departure time , i.e. the pre-journey wait time is ignored.

Please note that in our temporal network, transfer edges were not created between stops along a directed route even if they were within a walk distance, further reduced the possibility of unnecessary transfers in calculating shortest path travel time between stop pairs.

* 1. *Bus network metrics*

We used two bus stop-level metrics to compare the bus network characteristics calculated by the topological network model and by the temporal network model. The first metric measures bus stop accessibility (to/from other bus stops) and the second measures how important a bus stop is in maintaining connectivity across the network.

This study proposed *proximity density* *(PD)* and *proximity average distance (PAD)* as metrics for measuring accessibility of a bus stop. Specifically, we defined *origin proximity density* and *destination proximity density* of a bus stop as the number of other bus stops that could be travelled to and from that bus stop, respectively, within a given (range of) total travel time, e.g. under 30 minutes. Similarly, *destination proximity average distance* and *origin proximity average distance* of a bus stop was defined as the average distance from/to the bus stop to/from other stops, respectively, within in that (range of) total travel time. In the temporal network, the total travel time between a stop pair includes in-vehicle time and the time of all transfers and is calculated by the algorithm described in Section 2.3. The total travel time in the topological network consists of only in-vehicle time and is calculated by Dijkstra’s algorithm using average travel time as edge weight. Please note because we aimed to measure the accessibility of a bus stop from the bus operation perspective, stop pairs within a walk distance from each other (thus the presumed mode is walking) were not included in the calculation of PD and PAD.

The importance of a piece of infrastructure (a bus stop or a bus route) in maintaining the connectivity across the network was measured by the number of stop pairs that have their shortest path passing that bus stop or using part of the bus route, which to some degree is similar to the betweenness centrality widely used in network analysis studies.

1. **Calculations**

The L-space graph described in Section 2.2 consisted of 4,350 nodes (i.e. the number of bus stops) and 5,397 directed edges and were implemented in Python using the network analysis package *networkX* (version 2.3) [ref]. The computation of shortest paths weighted by average travel time between all stop pairs in the network was carried out by the package’s built-in implementation of Dijkstra’s algorithm.

The temporal network representing the timetable of the HCMC bus network as described in Section 2.3 had 1,579,972 nodes and 1,561,663 transit edges. The number of transfer edges was at least an order of magnitude larger. For example, the number of transfer edges added to the temporal network in the 07.00 –10.00 period alone was over 14 million. In order to generate and model such a large network, we used *python-igraph* (version 0.8.0) [ref] which is the Python interface of the C core network analysis package *igraph* and provides the computational efficiency not available in the Python-based *networkX*. Please refer to the study by Leskovec and Sosic [ref] for a thorough review and comparison of the performance of *networkX* and *igraph* on various network operations.

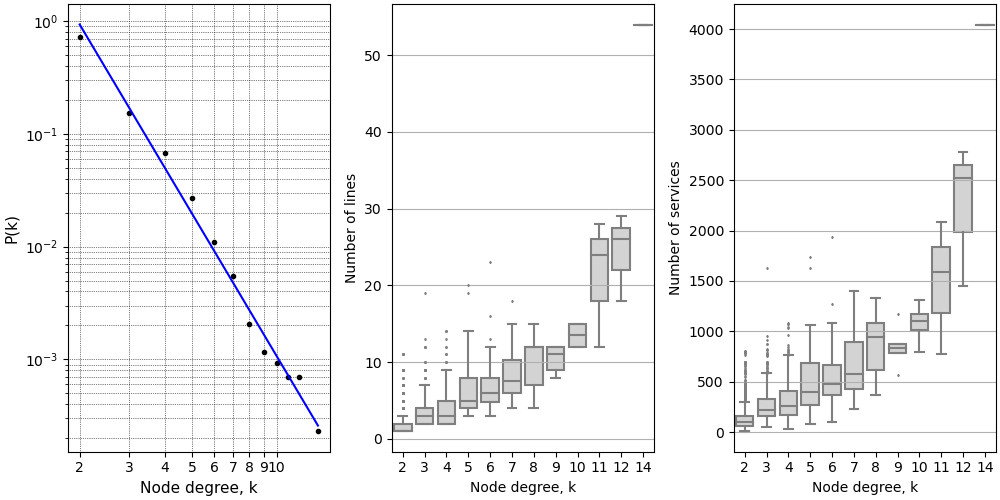
Please note because the temporal network explicitly represents any changes of service frequency and of stop by stop travel time along each directed route throughout the day, the shortest path travel time with minimum number of transfers between a given stop pair may be different at different desired departure times. In order to demonstrate such variation in the temporal network model (while minding the computational demand of the shortest path algorithm described in Section 2.3), we computed shortest path of all stop pairs in the HCMC bus network and the corresponding network metrics in Section 2.4 at three desired departure times 07.00, 12.00, and 17.00. The computation of the transfer-weighted shortest path between a stop pair in the bus network was carried out by the *python-igraph*’s built-in implementation of Dijkstra’s algorithm. We also limited the time window for a possible arrival time to within 3 hours from the desired departure time, i.e. the condition used in searching for ‘arrival’ nodes in Step 1 of the shortest path algorithm described in Section 2.3 became . Therefore, time window for searching an earliest departure, earliest arrival with minimum transfers path for the above three departure times were 07.00–10.00, 12.00–15.00, 17.00–20.00, respectively. In other words, we effectively assumed a maximum total travel time of 3 hours for a bus trip in HCMC. Transfer edges were generated for all eligible bus stops for the whole day but were only added to the temporal network for the above three time periods to lower the amount of network data kept in memory during the computation of the shortest paths.

The complete Python codes for the generation and analysis of the topological network and of the temporal network, together with the HCMC bus network data are available for download at link.

1. **Results and discussion**
   1. *Node degree and node strength distributions from the topological network model*

Degree distribution of a transit stop assumed a central role and appeared in almost every study investigating the topology of public transport networks [Shanmukhappa et al, 2019]. For a directed network, the total degree of a node is the total number of incoming edges and outgoing edges incident at the node. The degree distribution provides the probability of a node having a degree , mathematically expressed as , where is the total number of nodes in the network and the is the number of nodes having a total degree .

The plotting of versus for nodes in the HCMC topological bus network (black dots in Figure 3a) exhibits a strong power law, which is represented by the negative-sloped straight blue line in the log-log scale in Figure 3a. Mathematical equation of the line is or . This agrees well with the widely reported result that degree distribution of nodes in a topological transport network (especially in L-space) follows a power law [ref, Shanmukhappa et al, 2019].



**Figure 3.** Node degree and node strength distribution from the topological network model

Figure 3b and Figure 3c respectively present boxplots of the distribution of number of lines and number of services (node strengths) passing a node versus node degree . The overall trend is that a more connected stop (i.e. higher ) tends to bear more traffic (number of passing lines and services) and that the traffic tends to grow faster than increase of a node connectivity, as evidenced by the nonlinearity between the mean value in each boxplot versus node degree . Same observation was reported for the rail and bus transport systems in Singapore (Soh et al., 2010).

* 1. *Proximity densities (PDs)*

The shortest path travel time from/to a bus stop to/from every other stop in the network is classified into four categories of travel time, namely ‘under 30 minutes’, ’30-60 minutes’, ’60-90 minutes’ and ‘over 90 minutes’. The number of stops in each category is the destination PD and origin PD, respectively, of that bus stop in the corresponding travel time category. Figure 4 presents map plots of destination PD of all HCMC bus stops, normalised by a factor of where is the total number of bus stops in the network, for four categories of shortest path travel time, calculated from the topological network and at three desired departure times from the temporal network.

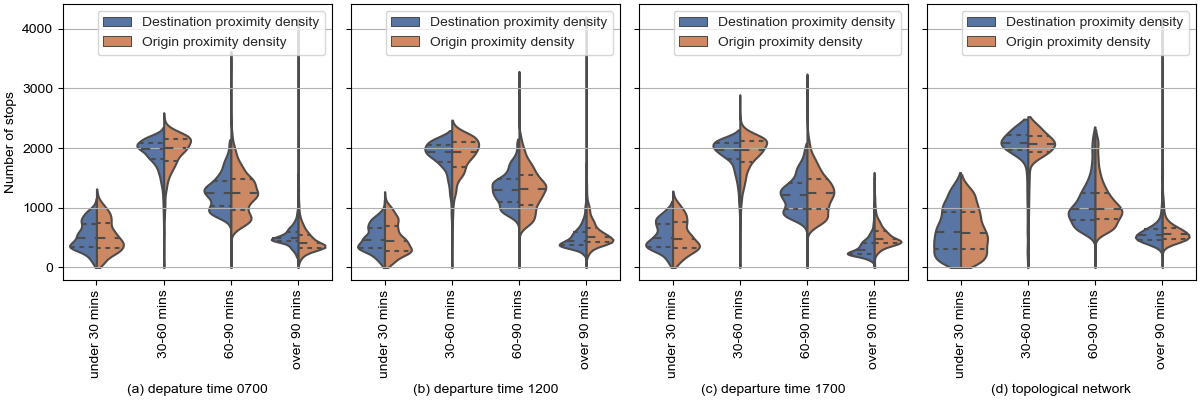
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Desired departure time in temporal model | 0700 | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens0700\destProxDens_u30_0700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens0700\destProxDens_3060_0700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens0700\destProxDens_6090_0700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens0700\destProxDens_o90_0700.PNG |
| 1200 | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1200\destProxDens_u30_1200.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1200\destProxDens_3060_1200.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1200\destProxDens_6090_1200.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1200\destProxDens_o90_1200.PNG |
| 1700 | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1700\destProxDens_u30_1700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1700\destProxDens_3060_1700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1700\destProxDens_6090_1700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDens1700\destProxDens_o90_1700.PNG |
| Topological model | | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDenstopo\destProxDens_u30_topo.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDenstopo\destProxDens_3060_topo.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDenstopo\destProxDens_6090_topo.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\proxDenstopo\destProxDens_o90_topo.PNG |
|  | | Under 30 minutes | 30-60 minutes | 60-90 minutes | Over 90 minutes |
|  | | Total travel time | | | |

**Figure 4.** Normalised destination PD of HCMC bus stops calculated from the topological model and for 3 desired departure times from the temporal model for 4 categories of total travel time. High resolution web maps are available at link.

Results from the topological network and from the temporal network appear to follow the same trends. For shortest path travel time less than 30 minutes, stops close to the network centroid (indicated by the star marker in Figure 1) are most accessible compared to other stops, i.e. they can reach more bus stops within this time frame compared to a stop at the boundary of the city. Their normalised destination PD are even higher in category ’30-60 minutes’, indicating that for a shortest path travel time between 30 and 60 minutes, these stops not only remain most accessible compared to other bus stops but also can reach even a higher number of other bus stops compared to when the travel time is less than 30 minutes. Indeed, 30-60 minutes is the travel time range in which the normalised destination PD of stops close to the network centroid are highest. As shortest path travel time increases (in categories ’60-90 minutes’ and ‘over 90 minutes’) stops that are farther from the network centroid become more accessible. Particularly in the ‘over 90 minutes’ category, stops close to the northern border of the city have normalised destination PD of over 0.9, meaning they can only reach over 90% of other stops in the network if the trip time is over 90 minutes.

The results of origin PD exhibit the similar trends observed in the results of destination PD. High resolution interactive map plots of normalised origin PD and normalised destination PD are available at link.

In order to demonstrate quantitatively the difference between results from the topological network and from the temporal network, let us focus on the bus stops that are within a 10 km radius of the network centroid (indicated by the star marker in Figure 1). The area covers almost all of inner districts of the city and includes 2,440 bus stops (approximately 58% of all stops) which are represented by red circle markers in Figure 1. Figure 5 presents the distribution of origin PD and destination PD of these inner city bus stops for four travel time categories, calculated for three desired departure times in the temporal network and from the topological network. The violin plots feature a kernel density estimation of the actual data points (i.e. origin PD and destination PD values) and quartiles of the underlying distribution.

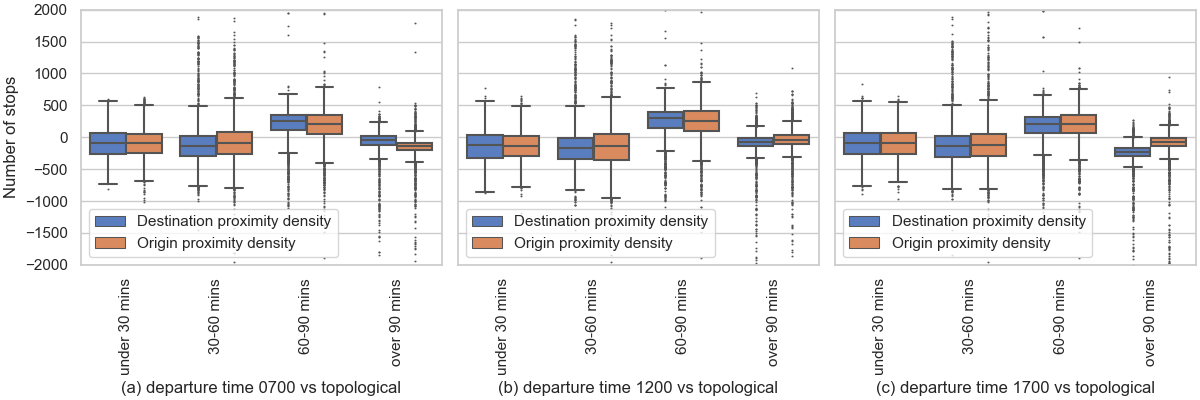


**Figure 5.** Distribution of PD calculated for each bus stop at different desired departure times in the temporal network and in the topological network

The plots for results from the topological network (Figure 5d) are highly symmetrical meaning that the distribution of origin PD and of destination PD of these inner city stops are very similar. On the opposite, the non-symmetry of violin plots for the results from the temporal network (Figures 5a to 5c) indicates the network’s capability to pick up differences of travel time in each direction between a stop pair, which was the combined result of heterogeneity in headway of directed routes, in the stop to stop travel time along them, and in the number and time of transfers. None of such heterogeneity was present in the topological network.

Plots in Figure 5 also echo the observation made from map plots in Figure 4 that inner city bus stops can be travelled to/from the highest number of other bus stops (i.e. are most accessible) for a travel time between 30 to 60 minutes. Let us take destination PD at the desired departure time 07.00 as an example. On average, approximately 2,000 bus stops are 30-60 minutes away from an inner city bus stop compared to the approximate of 300 stops, 1250 stops and 450 stops that are under 30 minutes, 60-90 minutes, and over 90 minutes away from an inner city bus stop, respectively. The same interpretation applies to plots of origin PD and destination PD calculated at other desired departure times.

Dissimilarities in the origin PD and destination PD distributions between three departure times are also noticeable but not significant, which can be attributed to our assumption of constant whole-trip travel time of each directed route, resulting in similar stop-by-stop travel time along each route throughout the day. The primary source of intra-day variation in the temporal model is the changing headway of each directed route.



**Figure 6.** Distribution of the difference of a bus stop’s PD calculated at a departure time in the temporal network and the stop’s PD calculated in the topological network

Finally, Figure 6 presents boxplots of the distribution of in which and are the PD at stop in travel time category calculated in the network and for the desired departure time in the temporal network, respectively. That a major part of the box plots is negative (except for those in ‘60-90 minutes’ category) was evident that the topological network tended to overestimate the number of stops reachable by an inner city stop compared to the temporal network in almost all travel time categories.

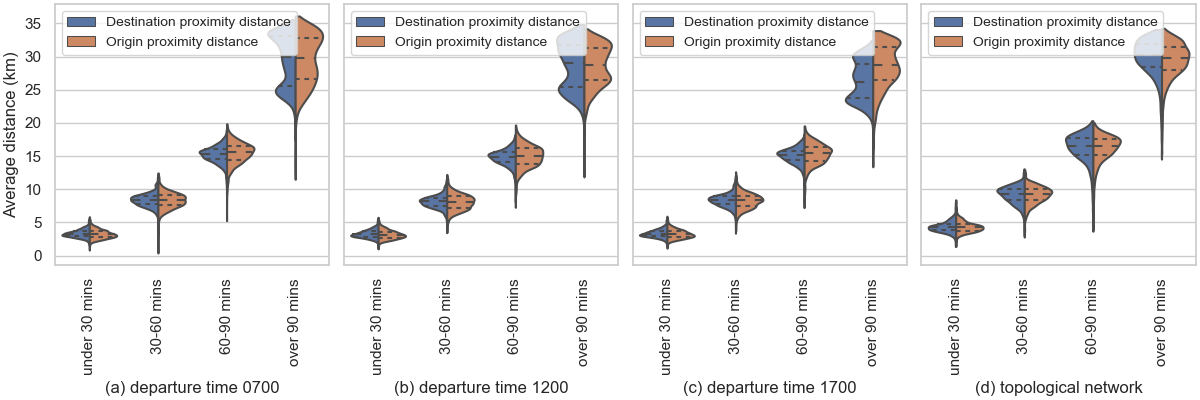
* 1. *Proximity average distances (PADs)*

The destination PAD and origin PAD of a stop in a travel time category was calculated by averaging the geospatial distances between that stop and stops that can be travelled from and to it, respectively, within the given range of travel time. Figure 7 presents the distribution of origin PAD and of destination PAD for different travel time categories calculated in the topological network and in the temporal network for the inner city stops. Key observations are summarised below.

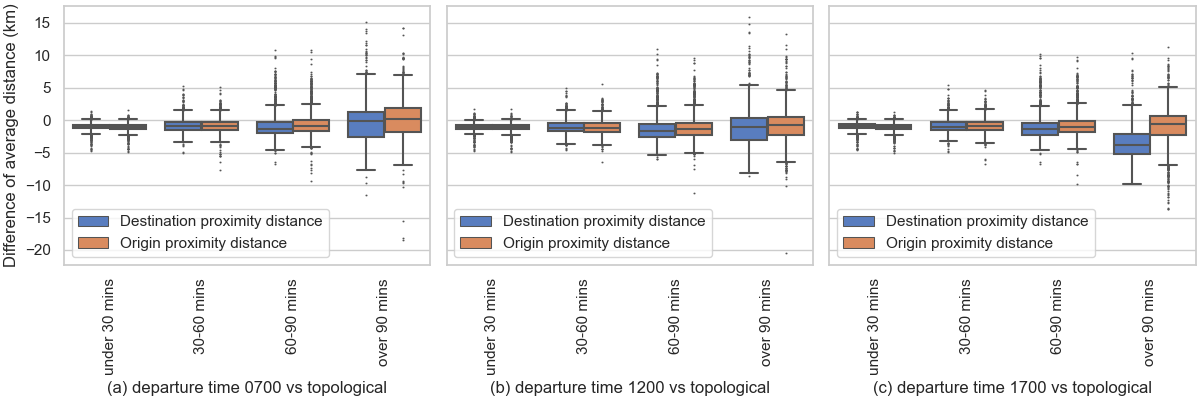
* Both destination PAD and origin PAD increase noticeably nonlinearly with larger travel time in the topological network and in the temporal network. It was also worth noting that even though on average an inner city stop can be reached by most other bus stops for 30-60 minutes of travel time (see Figure 5), the average geospatial distance between the stops was only approximately 8 km in the temporal network (Figures 7a to 7c) and approximately 10 km in the topological network (Figure 7d).
* The distribution of origin PAD and of destination PAD in the topological network were almost identical in all travel time categories, whereas their differences in the temporal network were noticeable, especially with larger travel times.
* The differences of destination PAD distribution and of origin PAD distribution between the three departure times in the temporal network were noticeable but not significant. This observation was similar to that of PD distributions. Again, this could be attributed to the fact that the primary source of intra-day variation came from heterogeneity of headway between directed routes. The stop by stop travel time along each directed route was assumed constant in our study due to the lack of information of whole-trip travel time variation of each route throughout the day.

On average, in the temporal network, for bus trips less than 30 minutes, the corresponding bus stops were approximately 3 km apart. This distance was approximately 8 km and 16 km for travel time between 30-60 minutes and between 60-90 minutes, respectively.

* Figure 8 presents boxplots of the distribution of in which and are the PAD at stop in travel time category calculated in the network and for the desired departure time in the temporal network, respectively. That the majority of each boxplot were less than 0 was evident that the average geospatial distances between stop pairs in the topological network tended to be higher than those in the temporal network, across all four travel time categories and three departure times.



**Figure 7.** Distribution of PAD calculated for each bus stop at different desired departure times in the temporal network and in the topological network



**Figure 8.** Distribution of the difference of a bus stop’s PAD calculated at a departure time in the temporal network and the stop’s PAD calculated in the topological network

* 1. *Most used bus network infrastructure*

For each bus stop we calculated the number of stop pairs that have their shortest path passing the bus stop. To facilitate comparisons of this metric between results from different desired departure times in the temporal network and from the topological network, we divided this metric by a factor of , which is the total number of stop pairs in a directed network. A bus stop having this stop pair fraction of 1 could be considered the most critical piece of infrastructure because all traffic (in terms of shortest paths) travel through it. On the opposite, a bus stop with a very low stop pair fraction could be regarded of low importance because its removal would not cause much disruption to the network traffic.

Figure 9 presents map plots of stop pair fraction of each bus stop in the network. The alignment of most traversed bus stops (marked by circles with darker colour) into corridors was evident in both the results from the temporal network and the results from topological network. While the difference between three desired departure times was almost indiscernible in the map plots (Figures 9a to 9c), the difference between results in topological network and those in the temporal network was much more visible (Figure 9d versus Figures 9a to 9c).

|  |  |  |  |
| --- | --- | --- | --- |
| D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\stopPairsFrac_0700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\stopPairsFrac_1200.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\stopPairsFrac_1700.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\stopPairsFrac_topo.PNG |
| (a) departure time 0700 | (b) departure time 1200 | (c) departure time 1700 | (d) topological network |

**Figure 9.** Map plots of stop pair fraction at each bus stop calculated at three desired departure times in the temporal network model and in the topological network

To better illustrate the differences, Table 1 presents top 10 most traversed bus stops and their stop pair fraction calculated at each desired departure time and from the topological network. It was noteworthy that the four lists share many common bus stops, even though actual numerical value of stop pair fraction of the bus stops differed from one list to another (those from the topological network tended to be considerable higher than those in the temporal network). Specifically, bus stop ID 1239 at the An Suong Bus Station topped the lists as the most traversed bus stop at all desired departure times and also in the topological network. The stop was found in the shortest path of 11.19% of all stop pairs calculated at 07.00, 10.94% of all stop pairs calculated at 12.00, and 10.09% of all stop pairs calculated at 17.00. Interestingly, the results from both temporal and topological networks pinpointed stops associated with the An Suong Bus Station and the Cho Lon Bus Station among the most traversed. These stations were indeed two of the city’s major bus hubs for route interchange.

**Table 1.** Numerical value of stop pair fraction of 10 most traversed bus stops

|  |  |
| --- | --- |
| **Bus stop description** | **Stop pair fraction** |
| **Departure time 0700** | |
| An Suong Bus Station - 1B National Road 22, Hoc Mon District (stop ID 1239) | 11.19% |
| Tan Binh Industrial Zone - 932 Truong Chinh Street, Tan Binh District (stop ID 271) | 9.99% |
| Thanh Cong Textile – 8 Truong Chinh Street, Tan Phu District (stop ID 174) | 9.98% |
| District 12 Community Centre – National Road 22, District 12 (stop ID 1152) | 9.05% |
| District Hospital – 605 Hoang Van Thu Street, Tan Binh District (stop ID 510) | 9.04% |
| Cu Cai Intersection – 43/1 National Road 22, Hoc Mon District (stop ID 1234) | 8.79% |
| Tham Luong Bridge, Thien Hoa Supermarket – 21 Truong Chinh Street, District 12 (stop ID 169) | 8.63% |
| Cho Lon Bus Station, Le Quang Sung Street, District 5 (stop ID 8) | 8.22% |
| Vinh Phuoc Temple – 110 Truong Chinh Street, District 12 (stop ID 273) | 7.95% |
| An Suong Bus Station – F42 National Road 22, District 12 (stop ID 1115) | 7.25% |
| **Departure time 1200** | |
| An Suong Bus Station - 1B National Road 22, Hoc Mon District (stop ID 1239) | 10.94% |
| Tan Binh Industrial Zone - 932 Truong Chinh Street, Tan Binh District (stop ID 271) | 9.74% |
| Thanh Cong Textile – 8 Truong Chinh Street, Tan Phu District (stop ID 174) | 9.67% |
| District 12 Community Centre – National Road 22, District 12 (stop ID 1152) | 9.21% |
| District Hospital – 605 Hoang Van Thu Street, Tan Binh District (stop ID 510) | 9.04% |
| Cu Cai Intersection – 43/1 National Road 22, Hoc Mon District (stop ID 1234) | 8.73% |
| Tham Luong Bridge, Thien Hoa Supermarket – 21 Truong Chinh Street, District 12 (stop ID 169) | 8.70% |
| Cho Lon Bus Station, Le Quang Sung Street, District 5 (stop ID 8) | 8.40% |
| Vinh Phuoc Temple – 110 Truong Chinh Street, District 12 (stop ID 273) | 7.90% |
| Bui Mon Intersection, National Road 22, Hoc Mon District (stop ID 1160) | 7.81% |
| **Departure time 1700** | |
| An Suong Bus Station - 1B National Road 22, Hoc Mon District (stop ID 1239) | 10.09% |
| Thanh Cong Textile – 8 Truong Chinh Street, Tan Phu District (stop ID 174) | 9.19% |
| Tham Luong Bridge, Thien Hoa Supermarket – 21 Truong Chinh Street, District 12 (stop ID 169) | 8.24% |
| Cho Lon Bus Station, Le Quang Sung Street, District 5 (stop ID 8) | 8.19% |
| Cu Cai Intersection – 43/1 National Road 22, Hoc Mon District (stop ID 1234) | 7.99% |
| Tan Binh Industrial Zone - 932 Truong Chinh Street, Tan Binh District (stop ID 271) | 7.81% |
| District Hospital – 605 Hoang Van Thu Street, Tan Binh District (stop ID 510) | 7.51% |
| District 12 Community Centre – National Road 22, District 12 (stop ID 1152) | 6.62% |
| Van Lang Park, 132A Nguyen Tri Phuong Street, District 5 (stop ID 432) | 6.44% |
| Cho Ray Hospital, 357-359 Hong Bang Street, District 5 (stop ID 437) | 6.22% |
| **Topological network** | |
| An Suong Bus Station - 1B National Road 22, Hoc Mon District (stop ID 1239) | 14.39% |
| Trung Chanh Intersection – 30/10B National Road 22, Hoc Mon District (stop ID 1393) | 12.96% |
| District Hospital – 605 Hoang Van Thu Street, Tan Binh District (stop ID 510) | 11.89% |
| An Suong Bus Station – F42 National Road 22, District 12 (stop ID 1115) | 10.65% |
| Thanh Cong Textile – 8 Truong Chinh Street, Tan Phu District (stop ID 174) | 10.54% |
| Tan Binh Industrial Zone - 932 Truong Chinh Street, Tan Binh District (stop ID 271) | 10.45% |
| Cu Cai Intersection – 43/1 National Road 22, Hoc Mon District (stop ID 1234) | 10.17% |
| Tan Xuan School, 1/4 National Road 22, Hoc Mon District (stop ID 1235) | 10.17% |
| Linh Son Temple, G68A National Road 22, Hoc Mon District (stop ID 3635) | 10.17% |
| An Suong Bus Station, 142 National Road 22, District 12 (stop ID 1116) | 10.06% |

As described in Section 2.3, a node in the temporal network described an event (arrival or departure) made by a bus service at a bus stop, thus contained information of the ID of directed route, the ID of the stop and the time of the event. Because a shortest path computed in the temporal network (following the 5-step algorithm described in Section 2.3) was essential an ordered list of these nodes, the information on the directed route(s) used in the shortest path was readily available and allowed us to identify directed routes most used to connect stop pairs across the bus network. Figure 10 presents the map plots of the top 21 directed routes (10% of all directed routes) which were identified as most used to connect stop pairs across the bus network at different desired departure times. Please note that such an exercise was not possible in the topological network approach (at least with L-space graphs) because bus routes, together with their spatial and temporal characteristics, were by design not incorporated in the topological network model.

|  |  |  |
| --- | --- | --- |
| D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\mostUsedRoutes_0700_v2.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\mostUsedRoutes_1200_v2.PNG | D:\workDir\fsppm\VINIF\work\ptNetworkAnalytics\paper 1\figures\mostUsedRoutes_1700_v2.PNG |
| 1. departure time 0700 | 1. departure time 1200 | 1. departure time 1700 |

**Figure 10.** Top 21 routes (in red) most used in shortest paths calculated at three desired departure times in the temporal network. The complete bus network is plotted in the background in dark blue.

1. **Discussion and conclusions**

The results of proximity densities, proximity average distances and most used infrastructure presented in Section 4 facilitated a direct comparison of operational characteristics of the HCMC bus network under two commonly used approaches for PT network analysis, the topological network and the temporal network. Such an apple-to-apple comparison of the two approaches, if was not the first reported in the literature, would be among a very few which quantified their differences. Below are key points from our comparison.

* The temporal network approach was much more computational demanding not only in terms of the size and complexity of the resulting network (see Section 5 for the exact number of nodes and edges), which translated into large amount of data kept in memory during computation, but also the computing power and time to undertake the analyses. Computing shortest path between stop pairs was apparently the most time consuming part. It was much more so in the temporal network since the shortest path search was done multiple times for a stop pair to match the multiple options of departure time at the origin stop with options of arrival time at the destination stop. On the opposite, temporal dimension was completely absent in the topological network, resulting in only one node for the origin stop and one for the destination, thus the shortest path search was done only once for a stop pair. Even when we implemented the temporal network and its analyses using the package *python-igraph*, which was C-core and thus more efficient than the Python-based *networkX* used for the topological network analyses, the time to compute the shortest path of all stop pairs within a 3-hour window in the temporal network was at least an order of magnitude (days) larger than that in the topological network (hours).
* The temporal network approach was inherently able to capture the temporal heterogeneity and intra-day variation in a PT network operation. Examples of such heterogeneity were stop to stop travel time along directed routes and changes of their headway throughout the day. One implication in our study was that the shortest path travel time in the temporal network included both transfer time and in-vehicle time, instead of only in-vehicle time in the topological network.
* Nevertheless, results from both approaches showed similar qualitative trends. For example, bus stops close to the city’s northern border could only reach 90% of other stops for a total travel time of over 90 minutes (see Figure 4). Meanwhile on average an inner city stop was reachable by the highest number of other bus stops for a travel time between 30-60 minutes (see Figure 5). Both approaches came up with visually similar corridors of most traversed bus stops (in terms of shortest paths between stop pairs across the network, see Figure 9).
* The results from topological network tended to be larger than those from the temporal network in all metrics calculated, including the proximity densities (in all travel time categories except for the ‘60-90 minutes’, see Figure 6), the proximity average distance in all travel time categories (see Figure 8) and the number of stop pairs that have their shortest path passing a given bus stop (see Table 1).
* By design, the topological network approach (at least with L-space graphs) was not able to identify bus routes from the results of shortest paths. It was also unable to capture dissimilarities in travel time along each direction between a stop pair, as evidenced by the symmetrical distributions in the violin plots in Figure 5 and Figure 7.

Among major limitations of this study were the assumption of constant speed of all buses along a directed route throughout the day and the assumption of constant dwell time of 6 seconds applied to all buses at their scheduled stops. We made such assumptions because a GTFS-like stop by stop timetable was not available for the HCMC bus network (at least not publicly) at the time of the study. Ongoing negotiations with the city authority for access to the daily archive of the buses’ location would not only provide information to relax the assumptions but also open up new research opportunities. Specifically, the daily bus location data (if available over a long enough period) would enable analyses of temporal variation of bus services the along each directed route throughout the day (e.g. due to changing road traffic) and from one day to another. The findings in turn provide valuable evidence in devising operational strategies which improves the bus network’s resilience. As a case in point, many of the stops in Table 1, regarded as important to maintaining connectivity across network, were along the National Road 22. Understanding impacts of delays at these bus stops due to scenarios of (inevitably) increased road traffic along this corridor to travel time to and accessibility of the rest of the network would provide much needed insights to mitigate such disruptions.

The use of infrastructure-based metrics, i.e. proximity density and proximity average distance in this study, for measuring a bus stop’s accessibility reportedly had shortcomings because they excluded land-use components (Geurs and Wee, 2004). Indeed, the travel time in our temporal network analyses accounted for only in-vehicle time and transfer time. Incorporating demographics information, such as the geographic distribution of the city population by age, would allow for the calculation of walkable distance and time to bus stops by different age groups, thus more granular accessibility evaluation of the bus network by different population cohorts.

Finally, from a more technical viewpoint, improving the computational efficiency of modelling and analysing the temporal network, which would substantially reduce computing time, should also be a priority. A promising direction would be adapting the multi-criteria profile connection scan algorithm (originally developed by Dibbelt et al. (ref)) used in this study into GraphX, a module in Apache Spark for graphs and graph parallel computation (<https://spark.apache.org/graphx/>).

**Acknowledgement**

This research was supported by the Vingroup Innovation Foundation (VINIF) in project code VINIF.2019.DA20.

**References**