DelayFlow Centrality for Identifying Critical Nodes in Transportation Networks

Yew-Yih Cheng, Roy Ka-Wei Lee, Ee-Peng Lim, Feida Zhu
Living Analytics Research Centre
Singapore Management University
Email: yycheng.2012@mais.smu.edu.sg, roy.lee.2009@sis.smu.edu.sg, {eplim,fdzhu}@smu.edu.sg

***Abstract*—In an urban city, its transportation network supports efficient flow of people between different parts of the city. Failures in the network can cause major disruptions to commuter and business activities which can result in both significant economic and time losses. In this paper, we investigate the use of centrality measures to determine critical nodes in a transportation network so as to improve the design of the network as well as to devise plans for coping with network failures. Most centrality measures in social network analysis research unfortunately consider only topological structure of the network and are oblivious of transportation factors. This paper proposes a new centrality measure called *DelayFlow* that incorporates travel time delay and commuter flow volume. We apply the proposed measures on the Singapore's subway network and its about 2 million commuter trips per day, and compare them with traditional topology based centrality measures.**

Keywords. Network centrality, transportation network.

I. Introduction

Motivation. Transportation network is an important part of a complex urban city system. Each transportation network is expected to support efficient flow of people between different parts of the city. Failures in the network can cause major disruptions to commuter and business activities which result in both significant economic and time losses. As the transportation network continues to grow and interact with other parts of urban city, it is imperative to study how the network can cope with increase in human flow, new network nodes and connections, as well as new city developments.

One way to study a transportation network is to identify the *centralities* in the network which represent the more critical nodes that can have major impact to the network operations. Network centrality is a concept introduced in social science to analyze important nodes in social networks [4]. The key examples of network centrality measures include degree centrality, closeness centrality, betweenness centrality [3], and pagerank [1]. In a recent study by Derrible [2] on metro networks represented by transfer stations, terminal stations and their connections only, it was shown that betweenness is more evenly distributed among stations when the metro network is large. This allows the stations to share commuter load more equally. Our work is quite similar to that of Scheurer, et al., who also proposed a set of centrality measures for transportation networks [5]. Unlike ours, their measures do not consider commuter flow and travel time of alternative means between stations.

All the traditional network centrality measures consider network topology only but not factors associated with transportation. A node identified to be important topologically does not need to be important from the commuter flow and delay perspective. Consider the network example in Figure 1, node A has the highest degree, closeness and betweenness values by topology. If we know that the commuter flow between B and D far exceeds those of other connections, B and D should be deemed to be more central than the other nodes. If we further know that many commuters will take much longer travel time should B fails, one may even consider B to be more central than D.

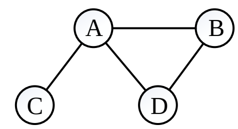

Fig. 1. Example Network

Research objectives and contributions. In this paper, we therefore develop a new centrality measure *DelayFlow* that incorporates *commuter flow* and *travel time delay*, the two transportation relevant factors. The objective is to determine critical nodes so as to improve the design of a transportation network as well as to devise plans for coping with the network failures. Unlike network topology, commuter flow and travel time delay may change dynamically with time, allowing us to study the evolving node importance. This time-dependent approach to measure node importance permits us to study a transportation network at a much finer time granularity. Due to space constraint, we unfortunately will only focus on time-independent analysis in this paper.

To compute *DelayFlow* measure, commuter flow and travel time delay are derived from commuter trip data and a public web service that offers travel time information respectively. We apply the proposed measure on the Singapore's subway network involving 89 stations and more than 2 million commuter trips per day, and compare them with traditional centrality measures.

II. Existing and Proposed Centrality Measures

We model a *transportation network* as an undirected graph $\langle V, E\rangle$ with node set V representing stations and edge set E representing the connections between stations. Every node $i \in V$ is associated with two numbers, $in(i)$ and $out(i)$, that refers to the number of commuters entering and exiting the station of node i respectively. Given that the total numbers of commuters entering and exiting the stations of the network are the same, the equality $\sum_{i\in V} in(i) = \sum_{i\in V} out(i)$ holds.

We now review the definitions of some existing network centrality measures used in social network analysis.

Degree centrality. The degree centrality of a node i is defined as $C_{deg}(i) = |\{(i,j)|(i,j) \in E\}|$

Closeness centrality. We denote the shortest path distance between two nodes i and j by $d(i,j)$. The closeness centrality of a node i is defined as $C_{cls}(i) = \frac{1}{d(i)}$ where $d(i) = \sum_{j \in V, j \neq i} d(i,j)$.

Betweenness centrality. Let g_{jk} denote the number of shortest paths between nodes j and k, $g_{jk}(i)$ denote the number of shortest paths between nodes j and k through node i. The betweenness centrality of a node i is defined as $C_{btw}(i) = \sum_{j \in V} \sum_{k \in V, k > j} \frac{g_{jk}(i)}{g_{jk}}$

Time delay is incurred for commuters to find alternative means to reach destinations when a node is down. To determine the extent of delay caused to the people affected, we compute $t_{exp}(i,j)$, the expected time taken to travel from node i to node j, and $t_{alt}(i,j)$, the time taken to travel from node i to node j using an alternative means of transportation (e.g. bus). The time delay factor l_{ij} is then defined to be $\frac{t_{alt}(i,j)}{t_{exp}(i,j)}$. We assume that t_{exp} and t_{alt} are static throughout the day. The time delay factor l_{ij} is asymmetric as both $t_{exp}(i,j)$ and $t_{alt}(i,j)$ are asymmetric. The larger the l_{ij} value (> 1), the greater the time delay.

Let h_{ij} denote the number of commuters from node i to node j per hour, and $h_{jk}(i)$ denote the number of commuters from node j to node k through node i per hour.

DelayFlow centrality. The DelayFlow centrality of node i is defined as $C_{dflow}(i) = \frac{\sum_{j \in V, j \neq i} h_{ij} l_{ij} + h_{ji} l_{ji} + \sum_{j \in V, j, k \neq i, j \neq k} h_{jk}(i) l_{jk}(i)}{\sum_{j \in V} in(j)}$ where $\sum_{j \in V} in(j)$ is the total commuter flow of the transportation network and $l_{jk}(i) = \frac{t_{alt}(j,k)}{t_{exp}(j,i) + t_{exp}(i,k)}$.

III. Comparison of Centrality Measures

Dataset. We compare the above measures using the Singapore's Mass Rapid Transit (MRT) network which consists of 89 stations in four train lines. In our experiment, we used three days worth of MRT trip transaction data from November 26 to 28, 2011 to derive the commuter flow information. Each trip transaction consists of the origin station, destination station and the timestamps at the two stations. We use the trip transactions to derive commuter flow h_{ij}'s. We compute the overall h_{ij} by dividing the total number of trips between stations i and j (not necessarily the origin and destination stations) by the number of MRT operating hours, i.e., 19 hours (from 0500 to 0000 hours). In a similar way, we compute $h_{jk}(i)$ from trips between j and k through i.

To determine the expected travel time and travel time using alternative routes, we made use of a third-party route suggestion service known as "*gothere.sg*"[1]. The *gothere.sg* API's allow us to determine the travel time by MRT train (expected) or bus (alternative) for each origin and destination station pair.

[1] http://gothere.sg

TABLE I
PEARSON CORRELATION BETWEEN CENTRALITY MEASURES

	C_{deg}	C_{cls}	C_{btw}	C_{dflow}
C_{deg}	1.0	0.42	0.67	0.64
C_{cls}	-	1.0	0.62	0.39
C_{btw}	-	-	1.0	0.52
C_{dflow}	-	-	-	1.0

TABLE II
HIGHEST CENTRALITY STATIONS

Rank	C_{deg}	C_{cls}	C_{btw}	C_{dflow}
1	Dhoby Ghaut	Bishan	Bishan	Dhoby Ghaut
2	Bishan	Serangoon	Serangoon	City Hall
3	City Hall	Lorong Chuan*	Buona Vista	Jurong East
4	Raffles Place	Dhoby Ghaut	Paya Lebar	Bishan
5	Outram Park	Marymount*	Lorong Chuan*	Outram Park
6	Buona Vista	City Hall	Dhoby Ghaut	Raffles Place
7	Paya Lebar	Braddell*	Outram Park	Serangoon
8	Serangoon	Woodleigh*	Dover*	Ang Mo Kio*
9	Jurong East	Clarke Quay*	Bartley*	Buona Vista
10	Tanah Merah	Bartley*	Clementi*	Toa Payoh*

Correlation between centrality measures. Table I show the Pearson Correlation scores between the different centrality measures. The table shows that among the traditional centrality measures based on network topology, degree centrality and betweenness centrality are more similar with each other than with closeness centrality. The nodes with high closeness are more likely be near the center of the network while high degree and betweenness nodes may be located away from the center.

DelayFlow centrality are more similar with degree centrality and betweenness centrality with correlation scores above 0.5. This is later verified by the top 10 stations of each centrality measure in Table II. The top stations ranked by most centrality measures (except closeness) are usually the interchange stations. The non-interchange stations are annotated with "*".

IV. Conclusion

In our paper, we have demonstrated the importance of considering transportation factors such as commuter flow and time delay in measuring network centrality of transportation network. Compared with the network topology based centrality measures, our new *DelayFlow* centrality measures are more relevant to the transportation domain in identifying critical nodes.

Acknowledgement

We would like to thank the Land Transport Authority (LTA) of Singapore for sharing with us the MRT dataset. This work is supported by the National Research Foundation under its International Research Centre@Singapore Funding Initiative and administered by the IDM Programme Office, and National Research Foundation (NRF).

References

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW*, pages 107–117, 1998.

[2] S. Derrible. Network centrality of metro systems. *PLoS One*, 7(7), 2012.

[3] L. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[4] L. Freeman. Centrality in social networks - conceptual clarification. *Social Networks*, pages 215–239, 1978/79.

[5] J. Scheurer, C. Curtis, and S. Porta. Spatial network analysis of multimodal transport systems: Developing a strategic planning tool to assess the congruence of movement and urban structure. Technical Report GAMUT2008/JUN/02, University of Melbourne, 2008.