An Exploration and Analysis of Su et al.'s Game Theory Model for Urban Public Traffic Networks

Coen Jones

January 31, 2020

Abstract

In 2007, a paper published by Su et al. proposed a game theory approach to the urban public transport network problem[1]. Little literature is available in the field, apart from a small array of conference papers on the topic [2][3]. As such, we are motivated to explore this model so as to critically analyse its relevance and its applicability to the problem. The original paper suggests that the basic idea of the model is that three network manipulators - passengers, the urban public traffic company and a government traffic management agency - play games in a network evolution process. Each manipulator tries to build traffic lines to magnify their own benefit. The authors suggest that simulation results show a good qualitative agreement with the empirical results, however our empirical study of simulated and local networks of varying types suggest that this may not be true.

1. Introduction

Urban Public Traffic Networks have historically been a popular topic of mathematical research. The challenge to optimise these presents the opportunity to optimise a large scale, centrally controlled system and have an impact on a large number of people[4], as well as the potential to make great savings on the large budgets at play, depending on how one looks at the problem. The Game Theory Model proposed in the 2007 paper by Su et al. presents an apparently original model, for which there is little background nor consequent investigations in the decade since. As such, the opportunity to evaluate and analyse a model which attacks the problem from a different angle using a simpler model is an interesting and important one.

The model proposed treats the stations of the network as 'nodes' and the lines connecting them as sets of these nodes. The simulations for the model operate in two stages - we shall refer to these as the 'set-up' phase and the 'game' phase. Each player has a 'selection principle' and a 'benefit'. The selection principles are used when a player is constructing a line to decide what nodes they would like to select. During the set-up phase, one player is put in charge of constructing a number of lines, with the expected outcome being that their benefit should be maximised. Once the game phase begins, these benefits are calculated at the beginning of each turn to decide who will make the next move. These decisions are made with inverse probability to benefits - the player with the most benefit at the beginning of a turn will be least likely to make the next move, and vice versa. These moves will consist of the player being able to construct a single line according to their selection principles. Supposedly, this process should lead to the players' benefits all converging to a similar level, thus suggesting that all players should benefit equally from the system once it evolves to some equilibrium.

Importantly, once the script for simulating is written, the only data that the model requires to be run is the number of passengers using a certain station. This flexibility is one of the key benefits of the model - a program can be written and, theoretically, used on any dataset of recorded or projected station use to find the optimal network. An additional benefit is that the expertise required to write and run a simulation for this model is on a lower level than that required to write what is currently considered state-of-the-art within the literature - the usage of genetic algorithms and related techniques.

Another feature that lends weight to this model is how it considers the benefit of all players, and uses their convergence as a way of suggesting that the network is optimal. This contrasts to many other papers, which consider how to minimise the cost (what would be considered the traffic company's benefit), whilst keeping things like the number of transfers within a certain bound by using constraints.

2. Formulation

To formulate the model, first we need to define what symbols and terminology will be used throughout, and then provide some semi-pseudocode to give an idea as to how the simulations are structured.

Sets

N	Set of Nodes - representing stations in the network
L	Set of Lines - each line being a set of nodes chosen from N

Inputs

$n_0 \in \mathbb{Z}^+$	Number of nodes at time $t = 0$
$a_i(t) \in \mathbb{Z}^+$	Number of passengers at node $i \in N$ at time t
$b_{i,j}(t) \in \mathbb{R}^+$	Congestion factor on the edge connecting i and $j \in N$ at time t
$g \in \mathbb{Z}^+$	Number of lines constructed during the 'set-up' phase
$t \in \mathbb{Z}+$	Number of time periods the game phase will run for
T	Poisson rate for length of lines

Calculated Values

$m_l \in \mathbb{Z}$	Length of line $l \in L$ - Drawn from a Pois(T) distribution.
$\delta_{i,l} \in \{0,1\}$	1, if node $i \in N$ is on line $l \in L$ - set during the line allocation process
$\bar{a} = \sum_{i \in N} \frac{a_i}{\lvert N \rvert}$	The average number of passengers waiting at stations
$\bar{h} = \sum_{i \in N} \frac{\sum_{j \in L} \delta_{i,j}}{\lvert N \rvert}$	The average number of lines passing through stations
$\bar{s}$	the average number of transfers required to get from point i to point j
$\bar{T} = \sum_{j \in L} \frac{m_l}{\lvert L \rvert}$	the average length of a line
$c_{i,j}(t)$	The number of edges connecting i and $j \in N$ at time t

It should also be noted that a and b evolve over time in the following manner:

$$a_i(t+1) = a_i(t)\left[1 - \sum_{l \in L} \frac{\delta_{i,l}}{\sum_{i \in N} \delta_{i,l}}\right]$$
$$b_{i,j}(t+1) = b_{i,j}(t)\left[1 + \frac{c_{i,j}(t)}{\sum_{i,j \in N} c_{i,j}(t)}\right]$$

These evolutions occur each time a new line is built during the game phase due to the fact that a new line would decrease the number of waiting passengers, however an increase in transit between two nodes would increase the possibility of congestion upon a line. Additionally, the selection of a poisson distribution for line length is discussed both in the paper, and at length in Zhang, Chen et al. [5], a paper which shares a number of authors with our paper of discussion. The values from the 'Inputs' section are all up to the choice of whoever uses the model, however some of these may be considered data, such as a and n_0. The factor b is a unitless value, and whilst it may be a choice, it would make most sense to set this to 1 initially, and let the results scale around this, as $b \geq 1$ is required to allow it to increase naturally, while $b = 1$ means that without external influence, a pair of nodes without a line running between them will retain a congestion value of 1. However, it could be perhaps that the operator of the model would like to have custom b values to reflect the congestion of other vehicles on the network, i.e. Cars on the road in a city when trying to model a bus network. However, for the purposes of this paper, we have considered the case where $b_{i,j}(0) = 1$ for all $i, j \in N$.

As mentioned earlier in the paper, the model uses what we have referred to as 'selection principles' to help the three manipulators - the traffic *Company*, the *Passengers* and the *Governmental* traffic management agency - decide how they would like to construct their lines. These are as follows:

$$\begin{aligned}
\text{Company} &\rightarrow \max_{i\in N} \frac{a_i}{\sum_{j\in L}\delta_{i,j}m_l}\\
\text{Passenger} &\rightarrow \min_{i\in N} \frac{a_i}{\sum_{j\in L}\delta_{i,j}m_l}\\
\text{Government} &\rightarrow \min_{i\in N} a_i \sum_{j\in L}\delta_{i,j}m_l
\end{aligned}$$

These selection principles are chosen so as to maximise the benefit of the situation for each player. The traffic company would like to maximise profit, and thus wants to maximise the number of people using their lines (a_i), whilst minimising the amount of service they need to provide $\left(\sum_{j\in L}\delta_{i,j}m_j\right)$. This value for the amount of service is equal to the number of stations accessible from the station of choice - the δ acts as a binary value for whether the station currently has line j built upon it, and multiplies it by the length of line j.

In contrast to this, the passenegers want to select a station that minimises the number of passengers waiting there, so as to avoid overcrowding. They would also like the maximum possible service from that station - both number of lines and number of stops accessible directly, hence wanting to maximise $\sum_{j\in L}\delta_{i,j}m_j$. Finally, the Government would like to minimise both the number of people waiting, so as to minimise the risk of incident, and the number of lines running through and their length, so as to minimise congestion.

In addition, we also have 'benefits' which we use to measure the benefit of each player during the game phase, which then also aids us in deciding who should take the next turn. These are as follows:

$$\begin{aligned}
\text{Company} &\rightarrow \frac{\bar{a}\bar{s}}{\bar{T}}\\
\text{Passenger} &\rightarrow \frac{\bar{h}}{\bar{s}}\\
\text{Government} &\rightarrow \frac{1}{\bar{a}\bar{h}\bar{s}}
\end{aligned}$$

We choose these as our benefits due to the fact that, for the company, $\bar{a}\times\bar{s}$ represents the number of tickets sold, as each time a passenger transfers they need to buy a new ticket, and that $\bar{T}$ represents the amount of service the company then needs to provide. For the passengers, they want to maximise $\bar{h}$, as this represents the most convenience, being able to select many lines from their given station, and therefore get the most direct route to their destination, whilst minimising the number of transfers they're required to take due to the fact that this both costs time and money. Finally, the government would like to minimise all of $\bar{a}, \bar{h}$ and $\bar{s}$, as more passengers, more transfers and more lines lead to more incidencts and congestion, the things the government would like to avoid.

To operate the model, we can write a program which either takes in our data or generates custom data, and then follows the following process:

Set all values up
Start 'Set-up' Loop:
    Generate line lengths from Pois(T) + 2
    For length of line 'k', select first station randomly, and then select stations using the player's selection principle until the line is the required length
    Update δ and c
Game Loop:
    Calculate transfer values for $\bar{s}$
    Complete evolutions of a and b
    Calaculate each player's benefit
    Select next player
    Follow same procedure as during the 'set-up', but using the selection principle of the chosen player

Note, that it should be of no consequence as to which player takes part in the 'set-up' phase, so long as they construct all lines. The selected player should begin the game with maximum benefit, whilst the other two begin with minimum, until the benefits eventually converge.

3. Exploration

So far we have discussed and formulated what appears to be the model described within the paper. The original paper is often vague in its description, and both from the outset and throughout the process, decisions on certain features of the model have had to be made and changed so as to try and reproduce the results both in a manner faithful to the original paper, but also promptly so that an empirical investigation and analysis could be completed within the time frame of the project. These design choices will be discussed here so as to clarify where they've appeared and what has motivated these decisions. Whilst some of these were made in advance, and some may not actually be changes to the model, it should be noted that these were all made during the process of using computer-generated data, and that the model which was built upon the completion of this chapter remained unchanged throughout the process of testing real data.

One design choice in fulfilling the experiment was the selection of data for which to run the model on. The original paper uses the Beijing bus network from 2003, however neither this data nor that of the current network was easily available. Given that such data was unavailable, it was thought that to try and test the rigours of this model that alternative data should be used. As such, Australian transit network data was used to simultaneously verify the conceptual model, and test its rigour outside of the Chinese setting. As all relevant literature has only involved empirical investigations upon other Chinese networks, this in fact had potential to be a key part of the project.

Another decision was the interpretation of generating from a Poisson distribution. Such a distribution has the potential to yield values of 0 and 1, which one may consider as being invalid lengths for a line. One option was to allow these 'trivial' lines, and consider them just to be lines which would not be realistically built. Instead, it was decided that we would shift the distribution up by 2, and shift the poisson rate T down by 2 to adjust. Thus, the minimum value we were now to get would be 2, whilst the mean would be $T + 2$.

A point of interest regarding the benefits is the units that go along with them. The most interesting of these is the power of $\bar{a}$ that each benefit has - for the Company, this is 1, for the passengers it is 0 and for the government it is -1. Upon doing simulations on small computer generated datasets, it could be noticed that if, say, the mean number of passengers were 1000, the company benefit would always be in the high hundreds, the passeneger benefit would sit between 0.2 and 0.8, while the government benefit was always smaller than 0.0001. These could be attributed to the placing of $\bar{a}$ in the benefits. To combat this, the benefits were changed so as not to use $\bar{a}$, but a slightly more complex value - the ratio of average passengers at stations currently served to average passengers at stations across the network. Thus, this value would be larger than 1 should the stations with the most passengers be currently serviced, and smaller than 1 should stations with the least number of passengers be serviced. Thus, the value still played the same part in trying to increase ticket sales or increase safety, but the difference in units was no longer a struggle. Upon this change, most benefits happily set between 0.2 and 0.8 at the start of each simulation.

The final design choice made was that of limiting the maximum number of transfers between two stations to be 2. This design choice was somewhat forced in that given the small period of time to complete the project, it was quite difficult to write an engine which could easily solve the number of transfers required between to stations for numbers greater than 2. This decision is supported by the fact that most of the available literature in solving public transit network design set a hard constraint on the number of transfers to be no more than 2, given the troubles this causes passengers [6],[7],[8].

Once these decisions were made, we were able to begin working with our finished engine on real data.

4. Analysis

Data was sourced from TransLink, Queensland's public transport provider, using what was the most recently available data, the Origin-Destination data for April 2019 [9]. These are collated such that we get the number of passengers originating from each station across the network. Below is a table describing a number of the sub-datasets within the collection.

One of the questions we wanted to answer was how the model performed computationally - not just qualitativly - as one of the drawcards we saw in its implementation was its speed.

Dataset	Number of Stops	Average Passengers
Queensland Rail	154	23008.870
Brisbane Bus	5708	1011.775
Sunshine Coast Bus	691	47.705
Logan Bus	953	334.227

Table 1: Dataset Characteristics

To do this, we need to first consider what the different algorithmic steps of the procedure are. Upon running simulations, it was decided that the main steps were:

- Line Selection
- Calculating Transfers
- Evolving passenger numbers
- Evolving congestion
- Calculating benefits

All of these are would most likely depend on N, except for line selection which will also depend on the selected value of T. Thus, we will first consider how this changes with respect to N and T, and then the others just with respect to N.

	Poisson Rate T				
No. of Nodes	3	8	13	18	23
154	0.016	0.012	0.017	0.027	0.024
1000	0.066	0.173	0.511	0.501	0.787
2000	0.486	1.828	1.739	1.876	1.783
3000	1.531	4.711	5.405	6.549	4.676
4000	1.495	1.723	7.025	8.18	9.296
5708	3.079	6.906	13.069	14.419	24.877

Table 2: Time taken to set up a single line (in seconds) with varying sizes of N and T

One can see that the increase with respect to N appears to be roughly $O(N^2)$, whilst it is difficult to say what with T may be, as it appears to vary - perhaps depending on the data used.

Now we consider the other processes. From the table, it would appear clear that the Transfers and Passenger algorithms have a complexity of $O(N^2)$ as well. Calculating benefits could also be at this level, however it appears to be negligable at any level of N. Given that

Process	Number of Nodes				
	154	1000	2000	3000	4000
Calculating Transfers	0.044	1.685	6.22	16.709	34.016
Evolving Passengers	0.018	3.036	17.838	65.375	141.293
Calculating Benefits	0.003	0.025	0.071	0.154	0.31

Table 3: Time taken (in seconds) for different processes with varying sizes of N

the congestion algorithm took over 2 minutes for size 154, it was ignored given that it is non-critical to the model - the current model can actually operate without calculating the congestion at all. What is worth noting is that due to the large run times on $|N| = 4000$, the times for $|N| = 5708$ were not calculated. As these processes need to happen on every turn, it would seem infeasible to run these simulations on any $|N|$ greater than around 2000. It should be noted that there is potential for the algorithms to be more efficient, and due to the fact that these simulations were run in Python, even switching languages could have some impact. However, it would seem that without a great change, the computational challenge of passenger evolution is a great one.

In addition to this computational analysis, it is equally critical, and perhaps even more so, to consider the qualitative results of the process. The graph below is from the paper by Su et al. [1], and can be considered to be their quality assurance for this model. They have used the Government as the setting up player, and as one can see in the graph, this leads to a convergence of the benefits:

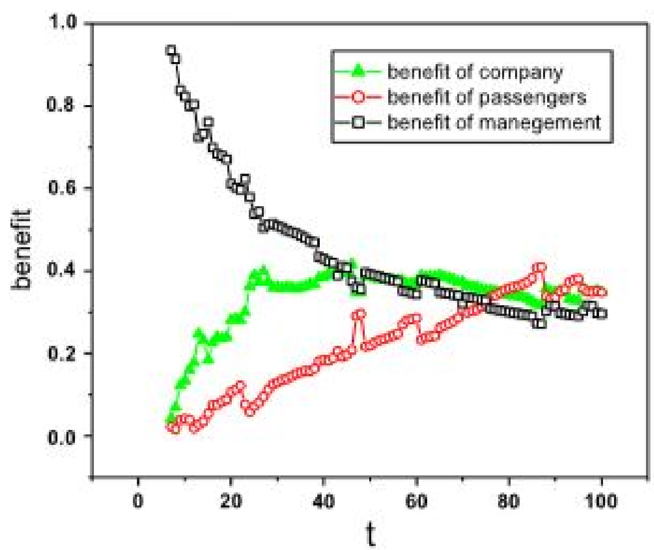

Fig. 6. The benefit evolution of the three network manipulators (color online).

However, in our simulations (completed on the QR dataset with $T = 13$), we have achieved quite different results, as can be seen in the graphs below:

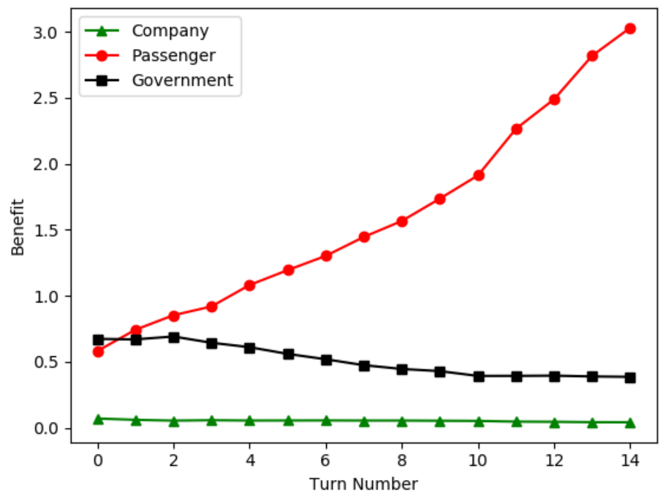

Figure 1: Simulated using Company as setup player

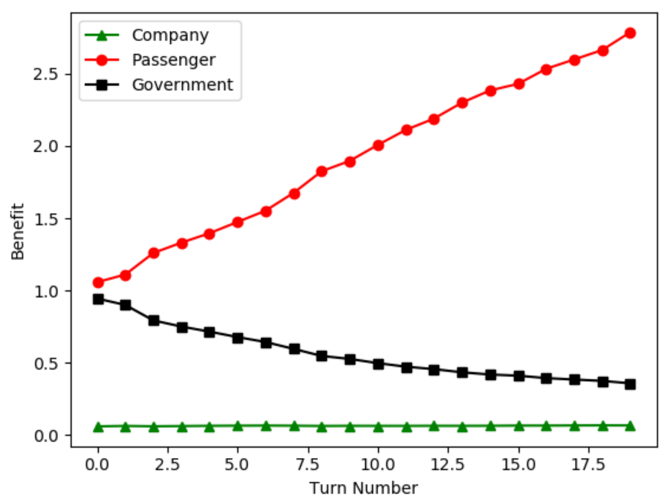

Figure 2: Simulated using Government as setup player

From these, it is quite clear that there are a number of interesting discrepancies. First and foremost, there is no convergence whatsoever. Secondly, even when the Company are allowed to set-up lines, their benefit remains the lowest of the three players. Finally, the Passenger benefit is strictly increasing at all times. These outcomes are consistent across all datasets, and in fact can be explained by the following principles:

Consider the benefits described in the paper. $\bar{a}$ may change, however it is likely to converge to 1 over time, as the ratio will be 1 if all stations are selected. Even if a different interpretation of $\bar{a}$ were made, due to the fact that $a_i(t)$ is strictly decreasing for all $i \in N$ due to the evolution function, it would seem that at the very least $\bar{a}$ is non-increasing with time. $\bar{T}$ will clearly converge to $T + 2$ over time, as it is a mean of a group of sampled values from $\text{Poi}(T) + 2$. $\bar{h}$ must be strictly increasing, as at every time period, a new line will be built. Thus, over time, the average number of lines through a station must increase. Similarly, $\bar{s}$ is non-increasing, as it is impossible to make it such that more transfers are required as a result of new lines being built. It is only possible for the same number of transfers to be required, or less, without the destruction of lines. From this, we can clearly see that the passenger benefit, $\frac{\bar{h}}{\bar{s}}$ must be strictly increasing, whilst the company benefit, $\frac{\bar{a}\bar{s}}{\bar{T}}$ and the government benefit, $\frac{1}{\bar{a}\bar{h}\bar{s}}$ will be likely to decrease over time, as seen in the graphs from our simulations.

Interestingly, this is not the only qualitative problem that the model appears to have. Due to the nature of the selection principles each player has, the number of passengers served by their lines varys drastically. Whilst the Company may build lines which one would consider feasable, often similar to real world routes in some ways, the ones constructed by passengers often have less than 1000 passengers served across all stations along the line (per month).

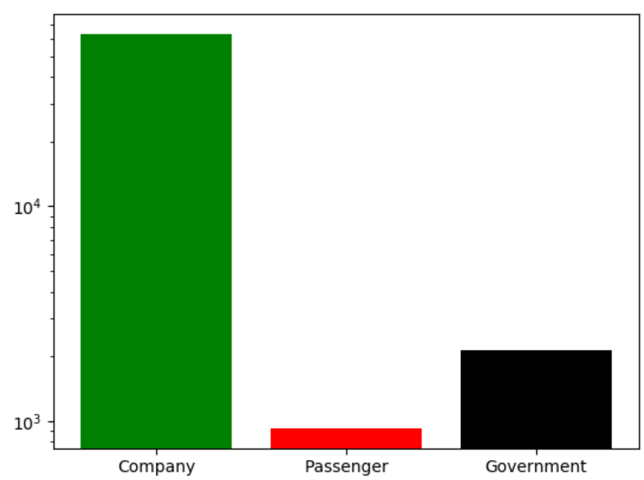

Figure 3: Average Passengers using lines built by player

These lines quite often include stations ridiculous distances away from each other - another interesting issue with the model. Given its inability to consider distance, some of the lines created ended up being not just infeasible for the sake of use, but the long distances which the lines covered. Due to the density of Chinese cities compared to Australian ones, this could be justified, however in Brisbane, this lead to lines which would have physical distances of over 400km. Thus, while such a model may have computational benefits, it appears the model lacks in its qualitative applicability, thus we must reject the model in its current state.

5. Potential Solutions

Whilst the time limit on this project has prohibited the further exploration of finding solutions to these issues which would have been potentially rewarding, this chapter aims to suggest some potential avenues for further research on the topic.

The first suggestion is to offer greater negative payoffs for other players, to prevent players building lines which are too negative on their fellow players. A mechanism which would prevent 'Passengers' from constructing lines so sparsely populated that it would be impossible for the company to turn a profit would be paramount of these, however it would appear that other problems could perhaps be solved by this too.

Following in the same line of thinking is the potential to tweak the benefits themselves slightly - in particular, rather than using $\bar{T}$ on the denominator of the Company's benefit, it may be more realistic to have the average physical distance of lines, so as to bring the idea of using physical distance more into the model.

Altering the setup principles of each player may also have an effect too. In particular, finding a way to weight the selections of passengers and government so that the benefit were magnified by the number of individual passengers who would benefit could perhaps balance out their complete avoidance of stations with any number of passengers using them.

Finally, potentially allowing players to destruct lines, rather than create, could help create different outcomes. Alternatively, perhaps once the system reaches a certain number of lines, turns could consist of both the destruction of a line of choice, to be replaced by a line constructed of the players choosing.

6. Conclusion

Whilst the initial paper seemed promising in its output, further investigation and extrapolation to local datasets have introduced doubt to its applicability and relevance. While there is the potential for this model to be extended upon by further works, perhaps into something more fulfilling and capable of solving the urban public transit network problem, it would seem that the current school of thought of employing metaheuristics is most likely the best way forward. Whilst the idea is interesting - the conclusion of Su et al. - and is simple and straightforward, also as concluded by the authors, it appears to be of no particular use.

Acknowledgements
This paper was written on partial fulfillment of the course SCIE3251 at the University of Queensland, Australia, under the supervision of Dr. Thomas Taimre.

Further Resources
The Python code and dataset used in the empirical study can be found as github project here: INSERT LINK
For any queries regarding its use and implementation, feel free to email me at
coen.jones@uq.net.au

References

[1] B.b. Su, H. Chang, Y.-Z. Chen, and D.r. He. A game theory model of urban public traffic networks. *Physica A: Statistical Mechanics and its Applications*, 379(1):291–297, 2007.

[2] Y. Ding and Z. Ding. Empirical study on some urban public traffic networks in south china. *2009 International Conference on Management and Service Science*, September 2009.

[3] Y. Ding and Z. Ding. A study on some urban bus transport networks in middle china. *2009 International Conference on Management and Service Science*, September 2009.

[4] J. Cooper and J. Corcoran. Journey to work in australia. *Australian Bureau of Statistics 2016 Census*, May 2018.

[5] P.P. Zhang, K. Chen, He. Y., T. Zhou, B.B. Su, Y. Jin, H. Chang, Y-P. Zhou, L-C. Sun, B-H. Wang, and D-R He. Model and empirical study on some collaboration networks. *Physica A*, 360(2):599–616, 2006.

[6] S. M. Hassan Mahdavi Moghaddam, K. Ramachandra Rao, and P. Biyana. Simultaneous bus transit route network and frequency setting search algorithm. *Journal of Transportation Engineering*, 145(4), April 2019.

[7] H. Zhao, W. Xu, and R. Jiang. The memetic algorithm for the optimization of urban transit network. *Expert Systems with Applications*, 42(7):3760–3773, May 2015.

[8] L. Liu, P. Olszewski, and P-C. Goh. Combined simulated annealing and genetic algorithm approach to bus network design. *International Conference on Transport Systems Telematics*, pages 335–346, 2010.

[9] TransLink. April 2019 - translink public transport origin-destination trips, October 2019. Data retrieved from https://www.data.qld.gov.au/dataset/go-card-transaction-data/resource/377af3d9-d90d-43c0-aee8-00caa74cc32f.