Robustness of Singapore Public Transportation Network

PROJECT MILESTONE REPORT

THE DAYLIGHT ANALYSTS	
Chew Tze Hoong	A0123694
Liang Nanxing	A0123702
Vishnu Gowthem Thangaraj	A0134525

EXECUTIVE SUMMARY

There were eleven media-documented cases of train breakdown events in Singapore in the first three months of 2015, triggering waves of commuters' dissatisfaction. The team recognises the need for the Singapore mass public transportation system to be improved, but given limited resources, the need to identify critical stretches of train routes is strong. This milestone report details the team's efforts-to-date in mapping the Singapore mass transportation network, and in identification of segments of train routes which when removed, will result in significant impacts on the infrastructure.

PROBLEM

There were eleven media-documented mass rapid transit (Includes train networks operated by both SMRT Corporation and SBS Transit) MRT breakdowns in Singapore in the first three months of 2015, sparking off waves of public criticism on the Singapore train network. While the intuitive step to resolve the issue should target at the root problem (i.e. maintenance issues), our team believes that a two-pronged approach should be considered, to include building an overall mass public transportation network that is resilient to failures.

PROJECT OBJECTIVES

The team would like to analyse the current mass public transport network, to understand the robustness of the network, through removing selected edges between train nodes iteratively. This mimics the Singapore train breakdown impacts, i.e. an entire stretch of train station is affected, rather than the mere local site where the train fault is located at. The team would then seek to identify crucial train paths which when removed, will bring about significant impacts to the mass public transport network. The team will also quantify the positive impact brought about by inclusion of new transportation nodes, by simulating the benefits brought about by newly-included network edges.

PROJECT SIGNIFICANCE

This project identifies stretches of MRT paths which are critical to the mass public transport network. Specifically, removing these paths will bring about significant impacts to the public due to a lack of convenient alternative paths. This project also evaluates the efficacy of proposed transportation nodes (trains stations and bus stops), to understand the benefits on the transportation network brought about by additions of new nodes/edges. The results of this study may be further disseminated to transportation companies so that special attention may be made to reduce chances of breakdown along these paths. The eventual goal would be improved public confidence on the mass public transport network, and better corporate image of train companies.

ASSUMPTIONS

1. Train networks by SMRT Corporation and SBS Transit form the core train network in Singapore.

2. Bus services by SMRT Corporation and SBS Transit form the core bus network in Singapore. All other smaller-scaled bus operators operating on flexible routes (These routes are predominantly in industrial estates, connecting factories to the nearest MRT station) have small-to-negligible impact on the overall transportation network resilience.

3. Bulk of mass public transport commuters will choose to complete their journey via mass public transport, as such, we do not factor impacts of cabs in the study of resilience of mass public transportation network.

4. The time taken by a bus / train to stop at a station for passengers to alight / drop off are not factored into the travelling time as they are not very significant.

5. During peak hours, buses travel on an average of 40km/h while during non-peak hours, buses travel on an average of 50km/h.

6. Commuters prefer to take the public transport to reach a station if the distance is greater than 100 metres. Short links (less than or equal 100m) between stations (bus, MRT, LRT) are represented by a 1 minute travelling time for walking.

7. The waiting time due to switching mode of transport is estimated at:

   A. change buses : 10 minutes

   B. change bus to MRT/LRT : 5 minutes

   C. change MRT/LRT to bus : 10 minutes

   D. change MRT/LRT at interchanges : 5 minutes

8. For MRT/LRT stations that are still under construction, we estimated their travelling time.

RELATED WORK

The analysis of the public transport system's resilience, reliability and robustness (or vulnerability) is a key concern in many countries and has been an important research topic for network analysis. Network vulnerability studies in general, have had contributions from various disciplines and from various countries, evidenced from its available literature. For example, Eduardo et al., 2014[1] outlines measuring the vulnerability and criticality in public transport networks.O Cats and E. Jenelius (2012)[2] considers vulnerability analysis of public transportation in the circumstance of disruption due to non-continuous availability. Murray (2013)[3] also suggested using multiple methodologies for network vulnerability approaches, namely, scenario specific, strategy specific, simulation and mathematical modeling.

Knoop et al., 2012[4] outlined the methodologies for assessing robustness of networks. Scott et al., 2006[5] provides alternative measures of link importance in graphs and Derrible & Kennedy, 2010[6] details how the number of cyclic paths in a metro system seem to directly correlate with the robustness of the network. For Singapore, more specifically, Harold Soh et al., 2010[7] does a network analysis on the public transportation routes in Singapore with certain graph measures and simulation models such as MATSim (Medina et al. 2013)[8] have been developed to stress test the resilience of the Singapore Transport system.

DATA COLLECTION

There were 3 main sources of data that were used to determine the nodes and edges.

Transit Link Bus Routes data

The data were from http://www.transitlink.com.sg/eservice/eguide/service_route.php, where users could enter the bus number to obtain bus route information such as, bus direction, bus stop codes, distance between the bus stops and bus stop description. As the webpage was in php, we took a different approach to crawl the bus route information. All bus numbers were stored first and subsequently iterated by appending it to the link (e.g. http://www.transitlink.com.sg/eservice/eguide/service_route.php?service=10. Each of the bus route web pages were then crawled using Python.

OpenStreetMap Metro data

The data were from http://metro.teczno.com/#singapore , which stored the osm (OpenStreetMap) XML file for Singapore. Geospatial information for locations such as amenities, transport nodes, hospitals, schools etc were available in the XML file. To identify a MRT or LRT station, nodes with the tag k = 'railway' were scraped. For bus stops, nodes with the tags k='highway' and v='bus_stop' were scraped. Information obtained are the station name / bus stop codes and their, longitude, latitude and description. For the MRT stations, data on stations that were not constructed yet were also available and are included in our study.

Wikipedia (MRT/LRT)

The MRT/LRT station edges data were obtained from Wikipedia. http://en.wikipedia.org/wiki/List_of_Singapore_MRT_stations , http://en.wikipedia.org/wiki/Light_Rail_Transit_%28Singapore%29. The travelling time from station to station were obtained from other sources such as MRT mobile applications and http://www.sbstransit.com.sg/ .

DATA TRANSFORMATION

Obtaining Bus Nodes and Edges

After obtaining the bus routes, the next step was to convert the information into nodes and edges. For each bus number and each direction, the bus stops were ordered based on the distance travelled (km) from the beginning terminal in ascending order. For each row, the lag value for the bus stop code and distance travelled were extracted and merged together with the main data set. The lag value of the bus stop code indicate the start of the edge and the bus stop code indicate the end of the edge. To obtain the distance between the bus stops, we subtract the lag distance travelled from distance travelled.

From :

BUS	DIRECTION	DISTANCE_KM	BUSSTOP_CODE
3	DIRECTION 1	0	75009
3	DIRECTION 1	0.6	76059
3	DIRECTION 1	0.9	76131

To :

BUS	DIRECTION	DISTANCE_KM	BUSSTOP_TO (END)	BUSSTOP_LAG (START)	DISTANCE_LAG	DISTANCE_BETWEEN (EDGE)
3	DIRECTION 1	0.6	76059	75009	0	0.6
3	DIRECTION 1	0.9	76131	76059	0.6	0.3

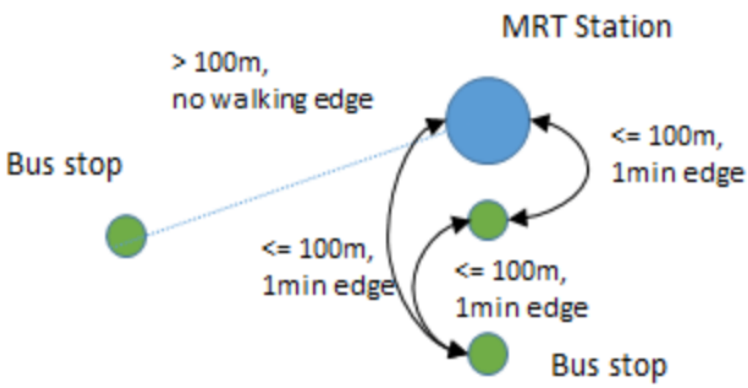
Building Links for Stations that are 100m apart

Short travelling distances between stations need to be modeled to consider the walking commute from one station to another nearby station. The assumption made here is that commuters will walk if the distance is less than or equal to 100 metres, and take the public transport if it is more than that. To identify nodes that are of close proximity to each other, we obtained the longitude and latitude of all stations, cross-join on the stations and calculated the Haversine distance between all combinations. There were a total of 4907 stations in our data set, and the number of combinations was approximately 24 million. Since the longitude and latitude obtained had 7 decimals, the precision of the distance was very high. The Haversine distance is one of the more accurate and computationally faster calculation for short distances and is represented by the following formula:

$$d = 2r \arcsin\left(\sqrt{\operatorname{haversin}(\phi_2 - \phi_1) + \cos(\phi_1)\cos(\phi_2)\operatorname{haversin}(\lambda_2 - \lambda_1)}\right)$$

$$= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

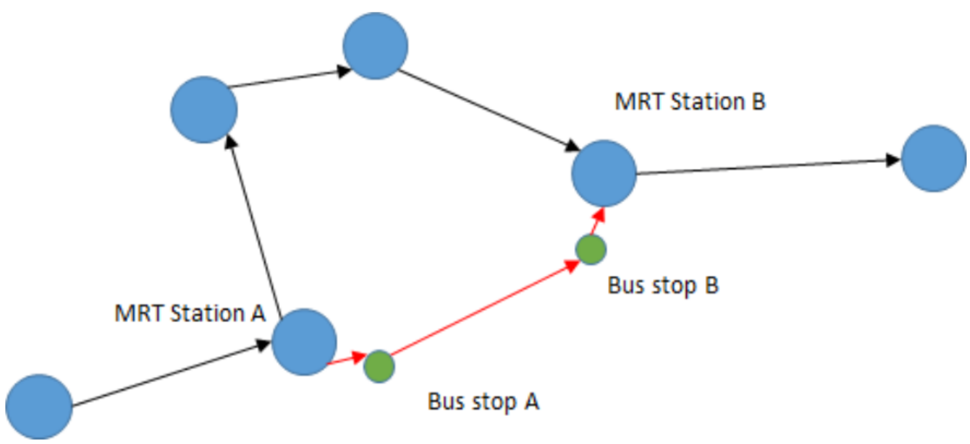
where r is the radius of the sphere (Earth) and approximated to 6378137 metres, ϕ is the latitude and λ is the longitude of the coordinates. Note that the decimal degree coordinates need to be converted to radians coordinates.



After the Haversine calculation is completed for all combinations, the next step is to filter out all node to node distances that are less than or equal 100 metres. A walking time of 1 minute is then defined for these edges.

Modelling Waiting Time for Switching Mode of Transport

If we do not penalize the travelling time made by commuters when they switch their mode of transport, there will be scenarios where commuters will switch transport because of the shortest path, but in real-life situations, this will not happen. E.g., the travelling time from MRT station A to station B is 5 minutes and the travelling time from bus stop beside MRT station A to station B is 3 minutes. If waiting time is not penalized, the simulated commuter travelling on an MRT passing by station A and B will proceed to switch to bus transport at station A, reach the bus stop at station B, and subsequently switch back to MRT to continue his travel. In normal circumstances, the commuter will continue his route on the MRT.



To model this scenario, we need to identify edges that are of the following:

A. MRT station to bus stop edge (identified previously by modelling less than 100m edges): 10 minutes waiting time.

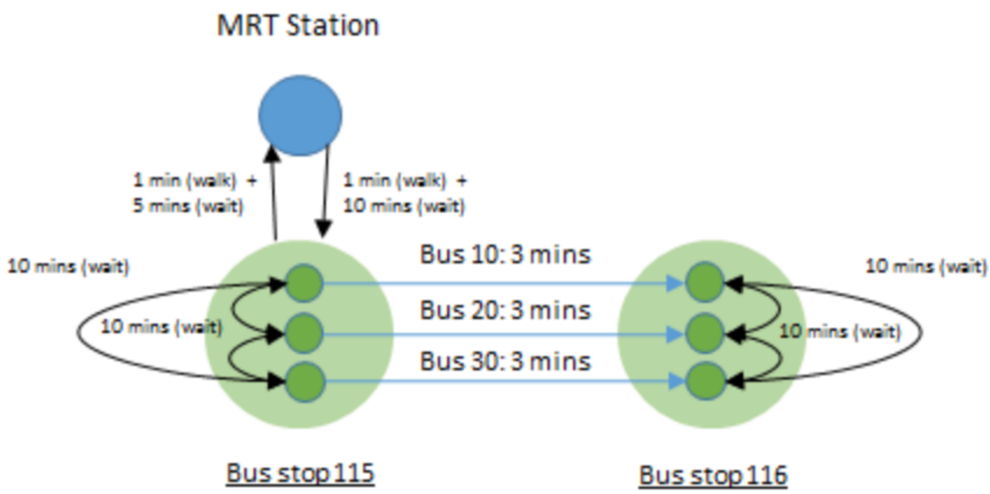
B. Bus stop to MRT station edge (identified previously by modelling less than 100m edges): 5 minutes waiting time.

C. MRT line switch (identify MRT interchanges): 8 minutes waiting time.

D. Bus to bus switch (identify stations that have more than one bus route): 10 minutes waiting time.

Scenarios a to c are relatively easy to identify, but d requires some drastic transformation to the nodes and edges. To achieve this, we split the bus stop nodes into individual bus-stop-route nodes.

E.g. bus stop 115 has bus routes 10, 20, 30. We transform bus stop 115 node to create 3 new nodes called 115_10, 115_20, 115_30. Similarly bus stop 116 has the same routes and is modeled as 116_10, 116_20, 115_30. We then add waiting time edges to all the combinations of bus transfer than can occur in that single node.



METHODOLOGY

ROBUSTNESS MEASURE

The team believes that an appropriate measure of robustness of the transportation network should possess the following traits:

1. An accurate quantification of the transport network's 'well-being';

2. Ease of translation to the physical domain, so that decision makers can easily relate to the figures produced, and the figures can easily be transformed to actual impacts on breakdown/changes to the transportation network; and

3. Aptness for application to different transportation networks of varying strategies (i.e. spoke-hub network vs point-to-point network), and varying components (e.g. ferry system).

In view of the above mentioned needs, total commute hours is adopted as this study's robustness measure. The team would like to define total commute hours as the sum of travel time for the entire population under study, for a defined set of (a) population, (b) starting and destination location-pair, and (c) travel period (e.g. peak hours between 0830hrs to 0930hrs).

Clearly, usage of total commute hours will allow scalability on the population growth, travel begin and end points evolution, and relevancy across different transport network types. The changes in commute hours itself may also be directly applied to understand productivity impacts brought about by network breakdowns and improvements. Last but not least, the total commute hours may be calculated for specific regions where the population is located, to understand the localized impacts brought about by network breakdown and improvements. A drawback of total commute hours is that the figures will be skewed by population growth/shrink, making direct comparisons difficult. In this case we should adapt the measure to compensate for population figures changes, for example, average commute hours should be used instead. For the purpose of this study, however, we shall use total commute hours we are studying a snapshot of the population (i.e. constant population), and the total commute hours can thus be translated to productivity impacts.

POPULATION SIMULATION

We assume 66.18%[1] of residents will commute to work in the morning during peak hours, and 75% of these commuters have destination within the Central Business District, while the remaining 25% will commute to the rest of the island. We also assume 90% of these commuters take public transport to work.

Population figures of residential districts are to be downloaded from SingStats, and the central points for each district will be identified. We obtain the number of commuters travelling from/to each transport node pair by making the further assumption that the residents are uniformly distributed within a 2km radius of the central point of each residential district, and that their commute destinations are uniformly distributed with the Central Business District, and out of the the Central Business District, in proportions as detailed in above paragraph.

With these assumptions, commuters between each start and end transportation node pair can be calculated. The shortest path (weighted by travel time) will be calculated for each node pair, and the total commute time can then be subsequently calculated. The team will then apply the following scenarios to understand the impacts on total commute time:

1. No breakdown (baseline robustness measurement);
2. Three actual peak hour MRT breakdown events, with the affected train segment/edges removed from the network;
3. Inclusion of proposed MRT lines that will be in service in the near future; and
4. Simulated removal of different train segment/edges to identify critical train segments.

TECHNICAL ANALYSIS

The nodes will be the bus stops, MRT and LRT stations. For the edges, it will comprise of both the travelling routes of buses, MRTs and nodes and the walking distance (<= 100m) between the nodes. Factors that will contribute to the edge weights include:

Buses to Buses:

i. Travelling Time (Travel Speed / Distance)

ii. Waiting Time: 10 minutes

iii. Walking Time between Nearby Nodes: 1 minute

Buses to MRT/LRT:

i. Waiting Time: 5 minutes

ii. Walking Time between Nearby Stops: 1 minute

[1] 3.62M labor population, with 5.47M population. 66.18% of population is employed [source: http://www.singstat.gov.sg/]

MRT/LRT to Buses:

i. Waiting Time: 10 minutes

ii. Walking Time between Nearby Stops: 1 minute

MRT/LRT to MRT/LRT:

i. Traveling Time: Based on usual travelling time. Information can be found in train stations, mobile applications and transportation websites.

ii. Waiting Time (Interchange): 5 minutes

For each edge between nodes, the total weight will be the summation of all the respective times mentioned above.

Dijkstra's algorithm will be used for calculation of shortest path between transportation nodes. Subsequent geo-spatial visualization will be performed on ESRI.

WORK COMPLETED

Data Acquisition

All bus routes have been crawled using Python BeautifulSoup package with XPath language. It was also realised that the web pages contained Javascript element which hindered the data retrieval, thus an additional package called Selenium with PhantomJS was utilised to "load" the web pages in a browser for scraping.

For MRT/LRT routes, since the numbers are significantly lesser, we were able to manually copy, paste and enter the information from the internet into a spreadsheet.

For the Longitude and Latitude data of all bus stops and MRT/LRT station, the osm (OpenStreetMap) XML file was parsed using BeautifulSoup and only relevantly information were extracted. Note that about 200 bus stop codes obtained from TransitLink could not be mapped to the nodes in the osm file.

Data Transformation

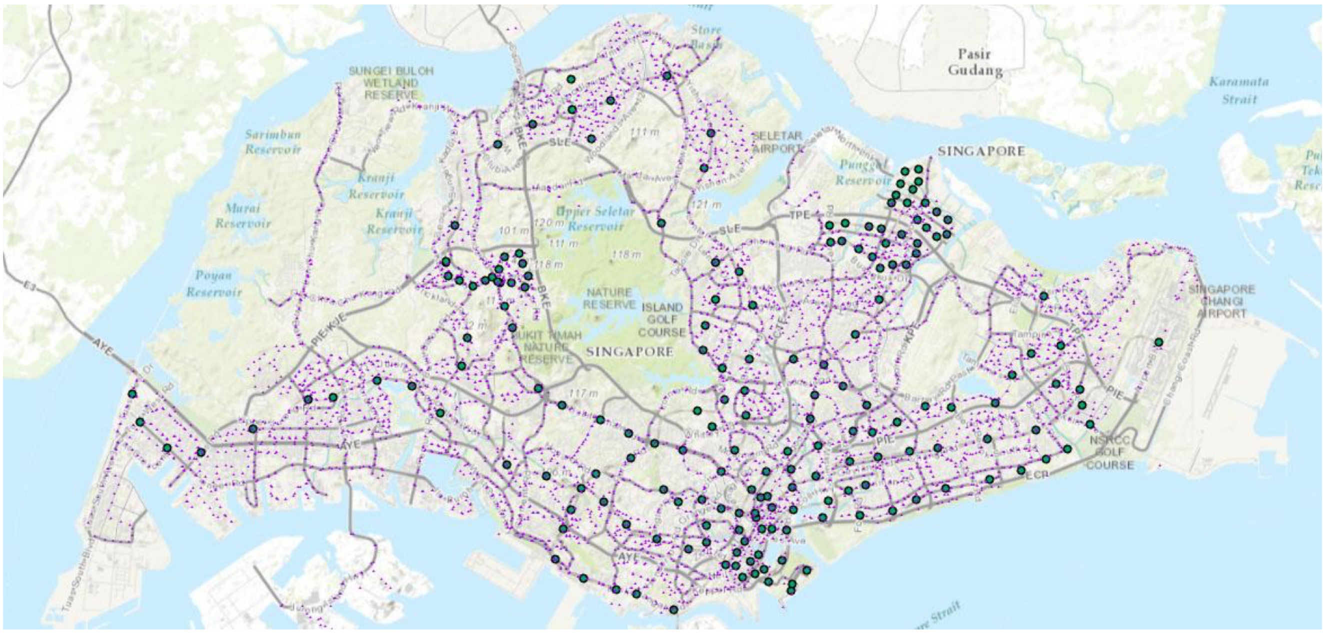
By making use of lag functions, we transform the route data into proper nodes and edges with weights. This was achieved using Python pandas package. The following data were collected:

Buses: Bus Number, Bus Direction, Bus Stop Codes (From), Bus Stop Codes (To), Distance between Bus Stops for each Bus Route, Longitude and Latitude of Bus Stops. Travelling time is estimated using Distance between Bus Stops divided by Speed (40km/h for peak, 50km/h for non-peak).

MRT/LRT: Train Station Code, Train Station Name (From), Train Station Name (To), Travelling Time, Train Line, Completed? (Some MRT stations are still in construction), Longitude and Latitude of MRT/LRT Stations.

Tool Familiarization

ArcGis and ArcMap from ESRI is the choice of geospatial visualization tool for this project. A 60-day trial copy of ArcGis and ArcMap is to fulfill the visualization need. The above-described crawled data is plotted on ArcGis, overlaying Singapore island's outline. Visual inspection for data anomalies (e.g. transportation nodes on water bodies and out-of-island) were carried out to ensure that the longitude and latitudes crawled are correctly mapped to the WGS84 coordinate system. Below figure shows all transport nodes mapped to Singapore map.



WORK-IN-PROGRESS

Population across Singapore

To identify the most important and least important nodes, the most travelled routes, nodes that have a very high impact in the event of a breakdown, we used the population data across various towns. Also, Singapore has a hub-and-spoke transportation model, with the town centres acting as hubs for each town, typically being an MRT station as well and are connected by the buses outwards to other nodes in the town. Singapore is subdivided into 55 areas(based on urban planning), organized into 5 regions (based on land area) and 15 towns (based on constituencies). Bedok, Jurong West, Tampines, Woodlands and Hougang were found to be the most populous areas based on resident population[2].

Further Data Transformation

Calculating Short Distances between Nodes: R is being used for calculating the distances between the nodes. A package called geosphere was used and it contained a Haversine distance calculation function to determine the shortest distance between two longitude and latitude coordinates. As there were 24 million combinations of coordinates to calculate, it is extremely computationally expensive.

Modeling Waiting Times when Switching Transport: To model this, bus stop nodes need to be split into bus-stop-route nodes. This may increase the number of bus nodes and edges by 5 to 6 times. Once completed, we can analyse all existing edges that involve a switch in transportation and add the waiting time defined earlier onto the edges.

Aggregating Total Traveling TIme: Once all edges traveling time have been calculated, the next step is to aggregate traveling time (weights) to remove duplicate edges, and get the final output for modeling.

Major Train Breakdowns

When a particular MRT train breakdown occurs, the stations that are connected closely to it will be impacted due to blockage of the route. The main idea here is to identify within the MRT networks, what are the clusters of stations

[2] http://www.citypopulation.de/php/singapore-admin.php

that will be affected if any of the stations suffer a breakdown in the connected route. For example, if a breakdown occurs at Clementi MRT station, the edge between Jurong East, Dover and Clementi will be removed. We will also be considering portions of the routes where a train depot is situated that can alleviate the breakdown situation. For example, if the train station at Bedok MRT breaks down on the west to east line, the train service between Tanah Merah to Pasir Ris can still continue as a train depot resides at Tanah Merah MRT station. This is part of the interdiction method of using pre-selected vulnerable links to assess robustness.

WORK PENDING

Network Robustness Measures

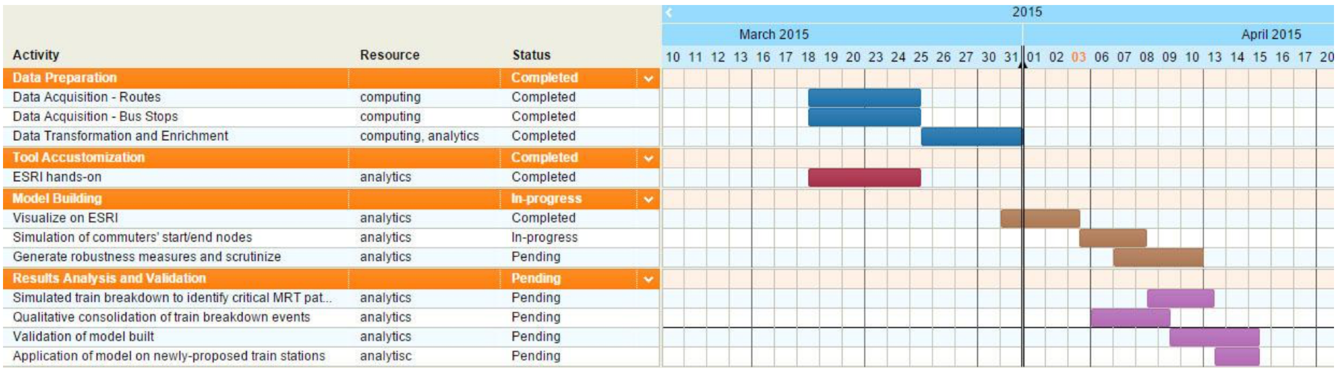
The global network robustness will first be analysed, using measures such as betweenness centrality, global clustering coefficient, average degree and degree-degree correlation (assortativity). We can then focus on the MRT train stations to determine which are the critical points within the MRT networks using robustness measures such as betweenness centrality and local clustering coefficient. Using information of MRT lines that has not been constructed yet, we can further analyse how the addition of these new lines will improve the network robustness of Singapore's public transport network. NetworkX, ARCGis are possible tools for calculating the measures.

Network Simulation of Commuters

The best way of simulating commuters is to use actual commuter movement transportation data, but due to inaccessibility of it, the best we could do was to make use of what's available. To further test the resilience of the public transport network, population simulation of commuters travelling from randomly selected nodes that are uniformly distributed within each district area to the Central Business District will be carried out. By using the area population data, we can do a good approximation of the commuter flow. Scenarios such as MRT breakdowns, addition of MRT lines will be simulated to determine their impact on the public transport network. Python's NetworkX has network simulation capabilities that we will be leveraging on.

Gantt Chart

The below Gantt Chart summarizes the timeline of delivery for each project milestone, and their inter-dependencies.



REFERENCES

[1] E. Rodríguez-Núñez, J.C. García-Palomares / Journal of Transport Geography 35 (2014) 50–63 : Measuring the vulnerability of Public Transport Networks.

[2] O. Cats and E. Jenelius, 2012 : Vulnerability Analysis of Public Transport Networks: A Dynamic Approach and Case Study for Stockholm.

[3] Murray, A.T., 2013. An overview of network vulnerability modeling approaches. GeoJournal, 78(2), pp.209–221.

[4] Knoop, V.L., Snelder, M., Van Zuylen, H.J. & Hoogendoorn, S.P. (2012). Link-level vulnerability indicators for real-world networks. Transportation Research Part A, 46, 843-854.

[5] Scott, D.M., Novak, D.C., Aultman-Hall, L., Feng, G., 2006. Network Robustness Index: a new method for identifying critical links and evaluating the performance of transportation networks. J. Transport Geogr. 14 (3), 215–227.

[6] H. Soh et al. / Physica A 389 (2010) 5852–5863 : Weighted complex network Analysis of travel routes on the Singapore Public Transportation System.

[7] Derrible, S., Kennedy, C., 2010. Characterizing metro networks: state, form, and structure. Transportation 37, 275–297.

[8] Ordóñez Medina, S.A., Shah, M. & Erath, A., 2012. Transport supply modelling of Singapore.