

# Urban Simulation Assessment

April 2, 2019

## 1 Part 1

<https://saintgimp.org/2013/01/22/merging-two-git-repositories-into-one-repository-without-losing-file-history/>

### 1.1 Introduction

This is an analysis of resilience in the London tube network. The analysis uses a recursive function for node removal to calculate the marginal effect of a node's removal, pseudo-code for the function can be found in Appendix 1. Below, criteria for node removal and effect evaluation are discussed below.

### 1.2 Impact Evaluation

Because the node removal criteria was decided in the context of the network effect metric, the network effect metric will be discussed first.

**Calculate community metric for network to justify focus on distance?** *from igraph: cluster edge betweenness, cluster fast\_greedy, cluster\_label\_prop, cluster\_leading\_eigen, cluster\_louvain, cluster\_optimal, cluster\_spinglass, cluster\_walktrap.*

Because London's tube doesn't demonstrate strong sub clusters, breaking the network into isolates was investigated but not pursued. Instead, the focus will be on forcing tube users to travel further for longer on their journeys.

This is investigated using shortest path and shortest topological path. Given the spatial nature of the london tubes, where edge attributes represent actual distances, true shortest path might be an attractive option. In the context of the London tube though, total time and effort are more important than total distance though. Trains can travel longer distances fairly rapidly while travelling through a high number of stations increases time dramatically because of the need to stop. Further, it is assumed that traveling through a higher number of stops implies a higher number of time and effort consuming train changes to switch. Thus by using geodesic, what is being maximized is the increase in stoppage time, and line change time for travelers in the network.

The statistics calculated to understand the effect of node removal include: % of connected nodes in the graph, average of the node specific statistics above,

The igraph package's mean\_distance function was used to compute the average shortest path between nodes in the network. The unconnected parameter was used to specify that nodes that were not connected to the largest cluster were counted as 1 + the longest possible geodesic, the actual longest geodesic is much less than this. This demonstrates

the incompatibility of different network measures. There isn't really a way to compare the effect of a longer trip with the effect of removing a trip possibility entirely. To do that we would have to include alternate modes of transport like the bus network. The actual longest geo

### 1.3 Node Removal Criteria

The function was run for measures that include, degree, betweenness, topological betweenness, closeness, topological closeness and eigenvector centrality. The correlations for these values across stations can be reviewed in figure 1. It was noted that correlations between weighted and topological measures were high, indicating that the distances between tube stations are fairly consistent so that the number of stations between two stations is a decent approximation of the distance. The correlation of measures betweenness and degree is also fairly high, indicating that tube stations at the middle of a line, with higher betweenness, also tend to have multiple lines, high degree. Correlations between eigenvector centrality and the other measures was very low, indicating that this does not give the same information as other metrics. Lastly, correlation between weighted and topological eigenvector centrality was 0 **indicating a problem with how they were calculated!**.

In order to maximize the increase in travel time measured by the average length of geodesic paths, betweenness will be used to order node removals. This measure is the number of shortest paths between nodes that travel through a given node. Deleting the node with highest betweenness will force the highest number of trips to use an alternate, ostensibly longer, path through other stations.

One note about this process, deleting nodes in some places creates isolates. In this investigation **this was not an issue I hope because for the first 10 or so nodes by betweenness, they were not the only node connecting any other node to the network.**



Figure 1: Correlation between station/node metrics

	By Betweenness		
	Node Deleted	Increase in Avg Geodesic	Components
1	Green Park	0.464271515336074	1
2	King's Cross St. Pancras	30.9672711339296	2
3	Bank	1.01101854695752	2
4	Waterloo	1.13665048051966	2
5	Stockwell	17.7264111534033	4
6	Embankment	108.61606869706	5
7	Baker Street	2.77460558311449	6
8	Notting Hill Gate	30.6706717614217	7
9	Ealing Common	26.6151971068078	9
10	Stratford	10.5750849462714	10
11	Canning Town	5.34435028248589	12
12	Hammersmith	8.37777866974866	14
13	Shadwell	7.82942087022292	16
14	Harrow-on-the-Hill	2.89045216902605	18
15	Camden Town	2.37783237317933	20
16	Canary Wharf	1.95898061029061	23
17	Mile End	0.964285951026795	24
18	Paddington	0.182272933085244	28
19	Earl's Court	0.109301944667777	31
20	Oxford Circus	-0.440105209296462	33
21	Woodford	-0.425293708021684	34
22	Aldgate East	-0.627897731687995	35
23	Finsbury Park	-0.34940346537536	38
24	Northfields	-0.418389327214584	39
25	Wembley Park	-0.540097783642864	41
26	North Acton	-0.588994666477163	43
27	Upney	-0.5890868871723	44
28	Rayners Lane	-0.598069582065818	46
29	Liverpool Street	-0.601393755502556	48
30	London Bridge	-0.597733955470574	50

	By Eigenvector Centrality		
	Node Deleted	Increase in Avg geodesic	Components
1	Embankment	0.124797831125548	1
2	Cannon Street	5.66059665069182	2
3	Moorgate	0.215731180123939	2
4	West India Quay	0.02021050051145	2
5	Great Portland Street	0.110753998811909	2
6	Farringdon	1.91439512290877	3
7	Paddington	23.3379220818089	4
8	Leicester Square	0.017700366401854	4
9	Heron Quays	15.4821648982723	5
10	Gloucester Road	0.319473942355302	5
11	Euston	34.9321449605953	6
12	Aldgate	-0.027020815846441	6
13	St. James's Park	-0.022791930367916	6
14	Mile End	0.500638234735106	6
15	Oxford Circus	0.310169191636348	6
16	Notting Hill Gate	3.48423397547472	7
17	Rotherhithe	-0.022978065465537	7
18	Blackfriars	-1.59050475761802	8
19	Baker Street	90.790225864921	12
20	Barons Court	5.840627408535	13
21	Aldgate East	1.495445128136	14
22	Ruislip Manor	1.32642824899307	15
23	Blackwall	-0.342224789480696	15
24	King's Cross St. Pancras	15.4170099145231	18
25	Island Gardens	-1.07286413770476	19
26	West Ham	9.42385307392772	21
27	Holborn	2.77304884203235	24
28	Waterloo	-0.370341481560303	24
29	Victoria	5.9845029956258	25
30	Custom House	2.3713875848423	26

#### 1.4 Analysis

Wow, look at that effect!

#### 1.5 Conclusions

The most important stations in the London tube network vary depending on the measurement used.

## 2 Part 2

*explain different models briefly: specifying which situations are appropriate for each*

- unconstrained,

The model is constrained to the total flows of the system but flows out of an origin and into a destination can be any value between 0 and total system flows.

This is useful for studying the change in connectivity between regions, for instance if a new transportation link was built.

- production

The direction of flows can change but the total flows from each origin will remain constant. This is useful for studying the effect of a new employment or consumption location that changes the destinations of people going to work or to spend money.

Sums of the rows are constant

- attraction constrained

The source of flows into a region can change but the total flows into a region will remain constant. This could be used to study a new housing development or

Sums of the columns are constant

- doubly constrained,

doubly constrained models could be used to test the short term effects of a change to transportation networks given that homes and businesses won't relocate but flexible behavior patterns like shopping could change almost immediately due to the change in accessibility or travel times between locations.

rows and columns sum constrained

*explain the role of the parameters*

Sensitivities to origin attributes, destination attributes, and linkage attributes

*select a scenario and explore the consequences of varying model parameters and inputs on interaction flows and the origin/destination estimates*

What if Amazon built a big office in Long Island City?

## 3 Part 3

*define, compare and contrast Cellular Automata and Agent Based models* Compare CA and ABM

While Cellular automata models are often viewed as separate modeling techniques, a more modern view of these methods considers them to be cousins or related in the sense that all CA are a subset of ABM.

Expanding a CA model usually results in the creation of an ABM. Cellular automata models are defined by a set of homogeneous cells that interact only with each other according to a defined set of input/output functions, e.g., if a cell with value 1 is surrounded by other cells f value 1 it's value becomes 0. The models can be extended to "n" states but all cells should be capable of reaching all n states to maintain the homogeneity of the cells. This can be contrasted with ABM where cells can have infinite heterogeneity and future states can be functions of the "environment" other cells or agents" and the cells own state.

The simplicity of CA models make them very useful for studying mathematical processes whereas Agent Based Modeling flexibility make them useful for modeling more complex "real world" phenomena and make them more accessible to non-technical audiences. Often the value of ABM comes from the ability to conduct parameter sweeps, to study combination rules that apply to multiple conceptual processes with different parameter values. The value of cellular automata models tends to be focused on the effect of initialization states on the long term outcomes and equilibria of the model e.g. for the same model, outcomes are a function of the rule set and initialization values, where as agent based models tend to be calculated for a large number of initialization values in order to study the effect of model dynamics independent of initialization values that may not accurately reflect the real world.

*vary model parameters to construct 3 scenarios, describe them, find time required to reach steady state, minimum runs required for statistically meaningful results.*

Scenarios

- Baseline
- high immunity
- high recovery

Record the run by using Behavior space ? Change script so it infects people automatically?

How is this going to be 1000 words?

use pseudocode

XXXX words excluding headings, figures, and references.

## References

- Grimm, Volker et al. (2010). “The ODD protocol: A review and first update”. eng. In: *Ecological Modelling* 221.23, pp. 2760–2768. ISSN: 0304-3800.
- Tasseron, G. and K. Martens (2017). “Urban parking space reservation through bottom-up information provision: An agent-based analysis”. In: *Computers, Environment and Urban Systems* 64, pp. 30–41. ISSN: 01989715.