Urban Simulation Assessment

April 22, 2019

1 Part 1

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

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

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. Trains can travel longer distances fairly rapidly while traveling 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 train changes which are difficult and slow. Thus by using geodesic path, what is being maximized is the increase in stoppage time, and line change time for travelers in the network.

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 was not a clear method for comparing 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.

1.3 Node Removal Criteria

For each node, degree, betweeness, topological betweenness, closeness, topological closeness and eigenvector centrality were calculated. The correlations for these values across stations can be reviewed in figure 1. It was noted that correlations between weighted

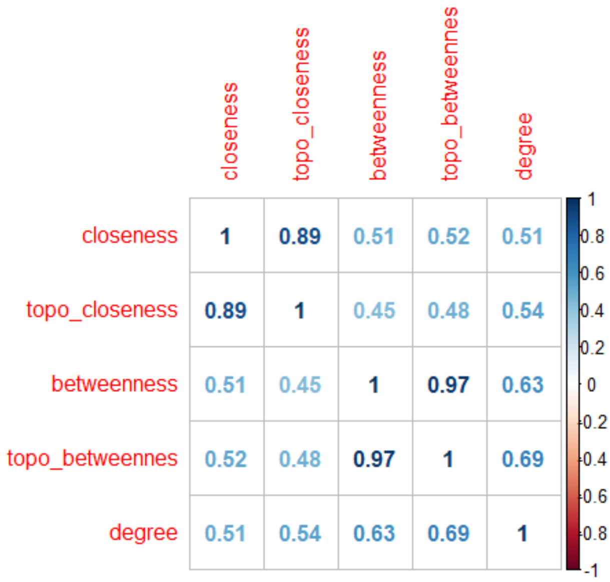

Figure 1: Correlation between station/node metrics

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. This supports the decision to use geodesic longest path. 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, it was not clear why correlation between weighted and topological eigenvector centrality was 0.

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. The function assumes that the distance between unconnected nodes is the longest possible distance on the graph. This will be discussed in greater depth below.

index	NodeDeleted	IncreaseGeodesic	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

Betweenness unconn is false			
	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 Betweeneness		
	Node Deleted	Increase in Avg Geodesic	Components
1	Green Park	0.464271515336074	1
2	King's Cross St. Pancras	0.344894786795857	2
3	Bank	1.14339690184063	2
4	Waterloo	1.28494742196984	2
5	Stockwell	0.563343124925417	4
6	Embankment	-6.16772972588596	5
7	Baker Street	2.32031653445306	6
8	Notting Hill Gate	-1.73581862937826	7
9	Ealing Common	-2.85804986154228	9
10	Stratford	-0.357026106908554	10
11	Canning Town	0.336399630312155	12
12	Hammersmith	-1.11269563661233	14
13	Shadwell	-1.44335210818291	16
14	Harrow-on-the-Hill	-0.489621630559977	18
15	Camden Town	-0.565504186081095	20
16	Canary Wharf	-0.335125486278896	23
17	Mile End	-0.846532555502438	24
18	Paddington	-0.328322619604598	28
19	Earl's Court	-0.196290601211155	31
20	Oxford Circus	0.016453621763357	33
21	Woodford	-0.056258058828067	34
22	Aldgate East	-0.000392838426682	35
23	Finsbury Park	-0.101735403043495	38
24	Northfields	-0.213720988377057	39
25	Wembley Park	0.008126154915853	41
26	North Acton	-0.075498287049283	43
27	Upney	-0.183419855551612	44
28	Rayners Lane	-0.120649861972213	46
29	Liverpool Street	-0.09859581775937	48
30	London Bridge	-0.079075958422429	50

	By Eigenvector Centrality		
	Node Deleted	Increase in Avg Geodesic	Components
1	Embankment	0.124797831125548	1
2	Cannon Street	-0.020785011623886	2
3	Moorgate	0.221126748710891	2
4	West India Quay	0.0217058916729	2
5	Great Portland Street	0.114088504753353	2
6	Farringdon	0.032692113174246	3
7	Paddington	-0.069066969738513	4
8	Leicester Square	0.031071992674539	4
9	Heron Quays	-0.213633720509144	5
10	Gloucester Road	0.404000167032718	5
11	Euston	-0.232471150322526	6
12	Aldgate	0.028208599848918	6
13	St. James's Park	0.034699837369548	6
14	Mile End	0.776660400757532	6
15	Oxford Circus	0.508933395700687	6
16	Notting Hill Gate	2.80231756571204	7
17	Rotherhithe	0.044124240677434	7
18	Blackfriars	0.001798634055188	8
19	Baker Street	-6.40337680481831	12
20	Barons Court	-0.217740790392554	13
21	Aldgate East	0.046883136212973	14
22	Ruislip Manor	-0.0251731067831	15
23	Blackwall	0.07893822023436	15
24	King's Cross St. Pancras	-0.928512463963177	18
25	Island Gardens	0.016339374551825	19
26	West Ham	0.366915223830574	21
27	Holborn	0.014000633210076	24
28	Waterloo	0.86009000361571	24
29	Victoria	-0.609084466226621	25
30	Custom House	-0.385377841072922	26

	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

When Kings Cross is deleted, it creates a new unconnected component out of the 11 stations on the north east end of the Picadilly line. In igraph, the two ways to handle this for the mean_distance() function are either to exclude distances between those unconnected nodes and the rest of the network or to assume that the distance is one greater than the longest possible geodesic in the network, that is ,the number of nodes on the network. Excluding distances between unconnected nodes led to a decrease in average trip length because nodes disconnected tended to have higher than average distance to other nodes, lowering the average metric.

Looking at the effect data it seems reasonable to say that betweenness did a better job than eigenvector centrality of prioritizing nodes to remove. Using betweenness created more isolates. It's difficult to judge which method lengthened average shortest path the most because of the options for dealing with disconnected networks. This will be discussed in the conclusion.

1.5 Conclusions

To improve this work, it would be good to add data about transportation networks besides the underground. In particular information about bus routes connected nodes would be useful because it would allow for a better estimate of average shortest path when subway stations become disconnected as the shortest path could then go through a bus route instead. Similarly, it would be good to include more granular data about where a rider would have to change trains. The current network assumes there's no cost to switch trains relative to staying on the same train passing through a station. Anyone who has walked from the Picadilly line to the Northern line at Kings Cross knows that there is a big difference.

An improvement to the data generally would be to use travel time data instead of using distance as an approximation.

Lastly, it would be interesting to build an igraph function that can compute average shortest path using edge weights since the current function cannot. This could confirm or reject the thought that tube stations are spaced fairly regularly based on the high correlation between weighted and topological centrality measures.

929 words

2 Part 2

2.1 The Models

2.1.1 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. In particular, it is useful for studying long term effects of a change where residence and employment are more flexible.

2.1.2 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. In terms of the matrix, it implies that the sums of the rows of the matrix are constant.

2.1.3 Attraction Constrained

The source of flows into a region can change but the total flows into a region will remain constant. That is, any reduction in flows into a destination from another region will be fully replaced by flows into the destination from another region. This could be used to study a new housing development that pulls people into residence in a different part of an area or a natural disaster that forces residents out of an area. Employers outside the area still need workers but will not be able to draw them from the same places after a housing change or natural disaster. In terms of the matrix the sums of the columns are held constant. Additionally, it can be used to study the effect of a specific change to employment where the model can be constrained to the values that result from that change.

2.1.4 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. In this model, the sums of both the columns and rows are held constant.

2.2 The Parameters

Basically, the parameters are ratios of how much a change of 1, in the natural logarithm of one of the predictor variables affects the natural log of the estimate of flow. This is seen below, 1-3 for the unconstrained and 4-6 for the doubly constrained. The equation is log-linearized and then the flow estimate is solved for. In equation 6, A_i and B_i are vectors of parameters that allocate parts of the regions total flows across origins and destinations and k is an arbitrary intercept used for estimation.

$$T_{ij} = kV_i^{\mu} W_j^{\alpha} d_{ij}^{-\beta} \tag{1}$$

$$ln(T_{ij}) = ln(k) + \mu(ln(V_i)) + \alpha(ln(W_j)) - \beta(ln(d_{ij})) \tag{2}$$

$$T_{ij} = e^{\{ln(k)+\mu(ln(V_i))+\alpha(ln(W_j))-\beta(ln(d_{ij}))\}} \tag{3}$$

$$\lambda_{ij} = A_i O_i B_j D_j d_{ij}^{-\beta} \tag{4}$$

$$ln(\lambda_{ij}) = ln(A_i O_i B_j D_j) - \beta ln(d_{ij}) \tag{5}$$

$$\lambda_{ij} = A_i O_i B_j D_j - e^{\beta ln(d_{ij})} + k \tag{6}$$

2.3 A Scenario

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

The scenario used for this assessment will be: “What if teleportation was invented and dramatically reduced travel times in connected boroughs but could only be used in London’s outermost boroughs due to construction requirements?” Thus origin and destination attributes remain constant but the travel costs change dramatically in the most peripheral boroughs: Barnet, Bexley, Croydon, Enfield, Harrow, Havering, Hilingdon, Hounslow, Kingston, Redbridge, Richmond, Sutton, Waltham.

This will be investigated using a doubly constrained model to estimate the short term effects where residences and businesses cannot relocate and a total constrained model to see the long term effects on business and residence locations as a result of the incredible new discovery.

To approximate the effect of near instantaneous transportation between two boroughs, the lowest distance between centerpoints of two London boroughs, 2080 meters, is substituted for the real distance between each of the 9 outermost boroughs.

2.4 Setup

The two models were estimated in R.

The parameters and goodness of fit statistics for the unconstrained model are seen in table 1.

For the doubly constrained model, the estimated intercept was 24.63 and distance parameter was -1.92. This model also produces a $32X32$ matrix of origin and destination specific parameters that will not be included here.

In both models the signs of the estimated parameters were consistent with expectations and the majority of parameters were significant at the 0.0001 level.

Parameter		Fit	
k :	-12.5	RMSE :	2330.9
μ :	1.62	R^2 :	0.386
α :	1.55		
β :	-1.5		

Table 1: Total constrained model results

3 Part 3

3.1 Overview

Cellular automata models are made of homogeneous cells that interact only with each other according to a defined set of rules. Expanding a CA model to include heterogeneous cells or a separate environment results in a full Agent Based Model. This CA could be thought of either as distinct from ABM or as a subset.

The simplicity of CA models make them very useful for studying mathematical processes whereas ABM is more useful for modeling 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 the combination of multiple processes. CA tends to focus on the effect of initialization states on the long term outcome of the model whereas 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.

3.2 Three Scenarios

The first is a baseline scenario used to explore a moderately recoverable disease to which immunity is low, for instance the seasonal flu without a vaccination program. The second scenario looks at the same disease with a vaccination program. The third looks at a disease like the common cold which is more recoverable but where immunity is close to non-existent.

Table 2: True Flows between the outer boroughs of London

	Borough	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Barnet	0	34	76	148	4098	2623	131	1023	611	68	305	229	44	555	74391
2	Bexley	132	0	4998	710	123	29	394	109	161	77	111	90	170	222	51231
3	Bromley	162	3199	0	6268	84	59	119	248	285	227	100	191	796	196	67450
4	Croydon	204	300	5152	0	120	97	82	457	581	827	64	480	6744	130	64539
5	Enfield	5642	52	76	136	0	328	233	457	339	38	538	98	47	1710	56955
6	Harrow	5008	26	47	103	325	0	29	6169	1141	107	44	246	34	95	49985
7	Havering	194	325	157	131	546	69	0	116	83	33	4844	45	50	1173	45621
8	Hillingdon	692	18	55	116	153	4688	21	0	5293	150	35	548	58	53	37054
9	Hounslow	253	27	69	215	92	396	28	12803	0	1006	31	7025	101	41	48403
10	Kingston	104	35	72	638	42	68	24	1070	1484	0	31	3788	1190	20	30687
11	Redbridge	567	141	132	173	1187	110	3004	212	231	46	0	88	48	5441	61005
12	Richmond	173	12	64	323	41	167	20	3365	6873	3549	12	0	260	46	45329
13	Sutton	81	68	514	7602	45	50	23	385	541	3122	26	750	0	41	39635
14	Waltham	715	79	149	253	2661	112	594	206	271	50	3736	101	41	0	57814
15	(all)	30744	11740	24934	31087	18456	18423	13275	48351	39292	17043	21033	23230	16652	17935	1800413

Table 3: Scenario Flows between Outer Boroughs using Total Constrained Model

	Borough	1	2	3	4	5	6	7	8	9	10	11	12	13	14	(all)
1	Barnet	0	11988	12810	11988	11186	13018	11087	11285	10211	13228	10890	17547	11285	10307	173516
2	Bexley	7021	0	7256	6790	6336	7374	6280	6392	5784	7492	6168	9938	6392	5838	96224
3	Bromley	13035	12606	0	12606	11762	13689	11658	11866	10737	13909	11451	18451	11866	10838	175190
4	Croydon	11646	11263	12035	0	10509	12231	10416	10602	9593	12428	10231	16485	10602	9683	158777
5	Enfield	8720	8432	9010	8432	0	9157	7799	7938	7182	9304	7660	12343	7938	7250	120425
6	Harrow	5842	5650	6037	5650	5272	0	5225	5318	4812	6234	5132	8270	5318	4857	79325
7	Havering	7226	6988	7467	6988	6520	7588	0	6578	5952	7711	6348	10228	6578	6008	97189
8	Hillingdon	7123	6889	7361	6889	6427	7481	6371	0	5867	7601	6258	10083	6485	5922	95743
9	Hounslow	5937	5742	6135	5742	5357	6235	5310	5405	0	6336	5216	8404	5405	4937	82379
10	Kingston	3823	3697	3951	3697	3450	4015	3419	3481	3149	0	3359	5412	3481	3179	52009
11	Redbridge	7642	7390	7897	7390	6895	8025	6834	6956	6294	8154	0	10817	6956	6354	106850
12	Richmond	5284	5110	5460	5110	4768	5549	4726	4810	4353	5639	4642	0	4810	4394	71294
13	Sutton	5654	5468	5842	5468	5102	5937	5056	5147	4657	6033	4967	8003	0	4701	78340
14	Waltham	5748	5558	5939	5558	5186	6036	5140	5232	4734	6133	5049	8136	5232	0	83172
15	(all)	101282	101364	101381	96976	93013	112306	91949	94739	88280	115885	91885	153546	97636	89964	1800407

Table 4: Scenario Flows between Outer Boroughs using Double Constrained Model

	Borough	1	2	3	4	5	6	7	8	9	10	11	12	13	14	(all)
1	Barnet	0	7	14	23	14	12	5	23	29	13	15	22	13	19	74393
2	Bexley	10	0	5	7	4	4	2	7	9	4	5	7	4	6	51232
3	Bromley	13	3	0	9	6	5	2	9	12	5	6	9	5	8	67450
4	Croydon	15	3	7	0	7	6	2	11	14	6	7	11	6	9	64537
5	Enfield	17	4	8	12	0	7	3	12	16	7	8	12	7	10	56957
6	Harrow	12	3	6	9	5	0	2	9	11	5	6	9	5	8	49982
7	Havering	6	1	3	4	3	2	0	4	5	2	3	4	2	4	45622
8	Hillingdon	6	1	3	4	3	2	1	0	6	2	3	4	3	4	37055
9	Hounslow	12	3	6	9	5	5	2	9	0	5	6	9	5	8	48403
10	Kingston	7	2	3	5	3	3	1	5	7	0	4	5	3	5	30685
11	Redbridge	16	4	8	12	7	6	3	12	15	7	0	11	7	10	61005
12	Richmond	15	3	7	11	6	6	2	11	14	6	7	0	6	9	45331
13	Sutton	10	2	5	7	4	4	2	7	9	4	5	7	0	6	39633
14	Waltham	26	6	12	19	11	10	4	19	24	10	12	18	11	0	57814
15	(all)	30744	11741	24935	31087	18454	18421	13276	48352	39294	17043	21032	23232	16649	17936	1800406

Table 5: Actual flows beteen Inner and Outer Boroughs

	Borough	1	2	3	4	5	6	7	8	9	10	11	(all)
1	Barnet	0	12080	148	4098	2623	5775	68	305	229	555	16330	74391
2	Camden	1496	0	147	295	330	4987	89	84	195	204	18829	51652
3	Croydon	204	3248	0	120	97	1752	827	64	480	130	10583	64539
4	Enfield	5642	5588	136	0	328	5317	38	538	98	1710	9052	56955
5	Harrow	5008	3675	103	325	0	1395	107	44	246	95	7882	49985
6	Islington	1001	10188	157	619	111	0	50	117	118	393	12835	50391
7	Kingston upon Thames	104	1547	638	42	68	710	0	31	3788	20	5419	30687
8	Redbridge	567	3790	173	1187	110	3104	46	0	88	5441	8122	61005
9	Richmond upon Thames	173	2504	323	41	167	1002	3549	12	0	46	8336	45329
10	Waltham Forest	715	5554	253	2661	112	4310	50	3736	101	0	10314	57814
11	Westminster	514	6786	216	121	142	2442	90	60	258	121	0	39288
12	(all)	30744	147985	31087	18456	18423	86387	17043	21033	23230	17935	353405	1800413

Table 6: Total Constrained Estimated Flows between Inner and Outer Boroughs

	Borough	1	2	3	4	5	6	7	8	9	10	11	(all)
1	Barnet	0	1626	11988	11186	13018	1144	13228	10890	17547	10307	1168	173516
2	Camden	713	0	170	360	360	3005	217	229	401	388	3300	24314
3	Croydon	11646	377	0	10509	12231	351	12428	10231	16485	9683	569	158777
4	Enfield	8720	640	8432	0	9157	672	9304	7660	12343	7250	520	120425
5	Harrow	5842	368	5650	5272	0	253	6234	5132	8270	4857	365	79325
6	Islington	391	2341	123	294	192	0	135	225	229	431	1359	19685
7	Kingston	3823	143	3697	3450	4015	114	0	3359	5412	3179	219	52009
8	Redbridge	7642	367	7390	6895	8025	461	8154	0	10817	6354	376	106850
9	Richmond	5284	276	5110	4768	5549	202	5639	4642	0	4394	423	71294
10	Waltham	5748	494	5558	5186	6036	703	6133	5049	8136	0	439	83172
11	Westminster	396	2556	199	226	276	1351	257	182	477	267	0	26914
12	(all)	101282	20319	96976	93013	112306	19356	115885	91885	153546	89964	26369	1800407

Table 7: Double Constrained Estimated Flows between Inner and Outer Boroughs

	Borough	1	2	3	4	5	6	7	8	9	10	11	(all)
1	Barnet	0	3192	23	14	12	2595	13	15	22	19	16212	74393
2	Camden	744	0	3274	690	752	455	1497	1306	1703	792	2587	51650
3	Croydon	15	9331	0	7	6	5286	6	7	11	9	18229	64537
4	Enfield	17	3612	12	0	7	1757	7	8	12	10	15676	56957
5	Harrow	12	3237	9	5	0	2715	5	6	9	8	10852	49982
6	Islington	974	733	2987	541	1016	0	1664	811	2116	419	4888	50388
7	Kingston	7	3762	5	3	3	2597	0	4	5	5	7209	30685
8	Redbridge	16	6090	12	7	6	2347	7	0	11	10	19567	61005
9	Richmond	15	4868	11	6	6	3756	6	7	0	9	9326	45331
10	Waltham Forest	26	4523	19	11	10	1488	10	12	18	0	17468	57814
11	Westminster	1296	887	2193	1028	865	1041	984	1439	1119	1049	0	39291
12	(all)	30744	147986	31087	18454	18421	86385	17043	21032	23232	17936	353401	1800406

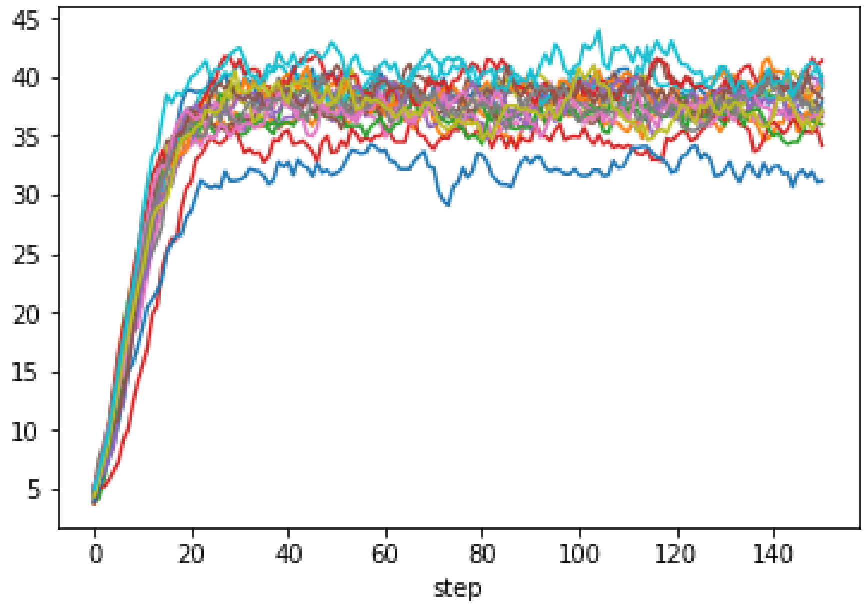

Figure 2: % of Population Infected Over Time
Scenario 1

All three scenarios will use a population of 1000 and infect 100 people initially.

3.2.1 Scenario 1

The first scenario, no vaccine flu, uses 50% chance of immunity, based on the idea that it's fairly common for someone with the flu to interaction with someone else without infecting them and a 10% recovery probability on the guess that flu symptoms last between 3 and 9 days for the majority of the population.

Looking at Figure 2, which plots the percentage of turtles infected over 150 steps of the model for 20 trials, a steady state becomes fairly clear. While this is not a perfect equilibrium, compared to the rapid increase in the first 20 ticks, the model becomes fairly stationary by tick 20 or 30.

The average percentage of turtles infected at 50 ticks was 38.79. The sample standard deviation was 1.799, which indicates a margin of error at 95% confidence of ± 0.842 using a t-value of 2.093. This implies that to get a margin of confidence of 0.5, 56.7 trial runs would be required. When 56 trials were run, the result was 37.93 ± 0.493, so the estimate for n was accurate.

The total infection time per non-immune turtle per tick was also recorded. The 20 trial mean at tick 40 was 0.6343 ± 0.00669. This adjusts for turtle immunity and runtime in order to compare results across scenarios later on.

Thus the model indicates that for these parameters, a fairly high number of turtles will contract the infection and the infection will be passed around continually rather than dying out at some point in the future. Considering the margin of errors for the results, if the model accurately reflects a real world scenario, we can have a good idea of how that scenario would play out.

3.2.2 Scenario 2

Scenario 2 looks at the flu in a society with high but imperfect immunization rates. In the US this is seen in cases where infants and the elderly cannot receive the vaccine and some people elect not to receive it or forget. This is accomplished by using the same recovery rate, 10% but increasing immunity probability to 80%.

Figure 3 shows the percentage of turtles infected over time in the trials. A steady state is more difficult to identify. One consideration is that the average percentage is lower than in scenario 1 because 80% of the population is immune. A second is that the lower mean scales the y-axis of the chart down from 0-14 instead of 0-45 in scenario 1. Thus while the standard deviation looks higher it is essentially the same, 1.869 in scenario 1 and 1.853 in scenario 2.

Because a steady state is less certain looking at the chart a t-test can be used to confirm that the mean of the set of trials is not changing between steps. At step 50 the mean percent infected was 5.92 with st. deviation 1.799. At step 100 these were 6.90 and 1.99. At step 250 7.935 and 1.696. At step 400 8.075 and 1.502. At step 500 7.956 and 2.113.

The p-value for a t-test with the null hypothesis that the means at steps 50 and 100 are different was 0.116. Indicating a 12% chance that they are different and that a steady state has not been reached with 95% confidence. The same test for the means at tick 250 and tick 500 had a p-value of 0.97. The test for steps 400 and 500 had a p-value of 0.84. Thus it is difficult to say with confidence that a steady state exists for this model as the effort to reject the null hypothesis that the means are different begins to resemble simply getting lucky enough to choose ticks with means that are very similar to each other by coincidence.

Sicktime per turtle per tick at 100 ticks was 0.1047 ± 0.01397.

At tick 250, the mean percentage of turtles infected was 7.94 ± 0.794 again demonstrating that the variability in results for this scenario is very similar to those of scenario 1 but that in light of the results of a t-test for a steady state, the true margin of error is likely to be higher since the process may not be stationary at tick 250, that is to say, the true mean may change across ticks.

3.2.3 Scenario 3

This scenario investigates how the common cold compares to the flu by using a higher rate of recovery 25% and lower rate of immunity 5% since there is no vaccine for the

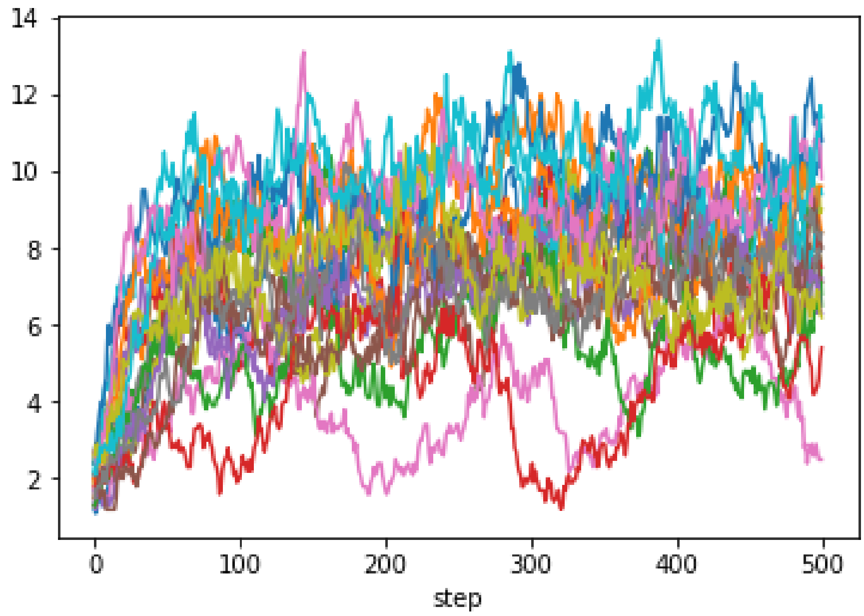

Figure 3: % of Population Infected Over Time
Scenario 2

common cold and many members of society continue their daily routines with a cold exposing relatively more people than the flu.

Figure 4 illustrates the percentage of turtles infected over time. Here a steady state seems more obvious. At 50 ticks the mean was 66.19 and st. deviation was 1.597. At 100 ticks this was 65.84 and 1.648. The 0.04994 p-value for a two tailed t-test for difference in means indicates only a 5% probability of the null hypothesis that the means are different.

At 50 ticks the mean percentage of turtles infected was 66.08 ± 0.902 . Thus the low variability of outcomes for different outcome values as a result of different parameter values continues.

Results for sick time per turtle per tick at tick 50 were 0.311 ± 0.0013. Less than the no vaccine flu scenario. which was less recoverable but more than the vaccinated scenario.

3.3 Conclusion

Comparing the three scenarios, The unvaccinated flu, scenario 1, resulted in a 38% average infection rate while the vaccinated flu, scenario 2, resulted in 7.93% of turtles infected on average. The common cold, scenario 3, gave 66% of turtles infected. At tick 50, non-immune turtles in scenario 1 had spent 63% of ticks infected on average while

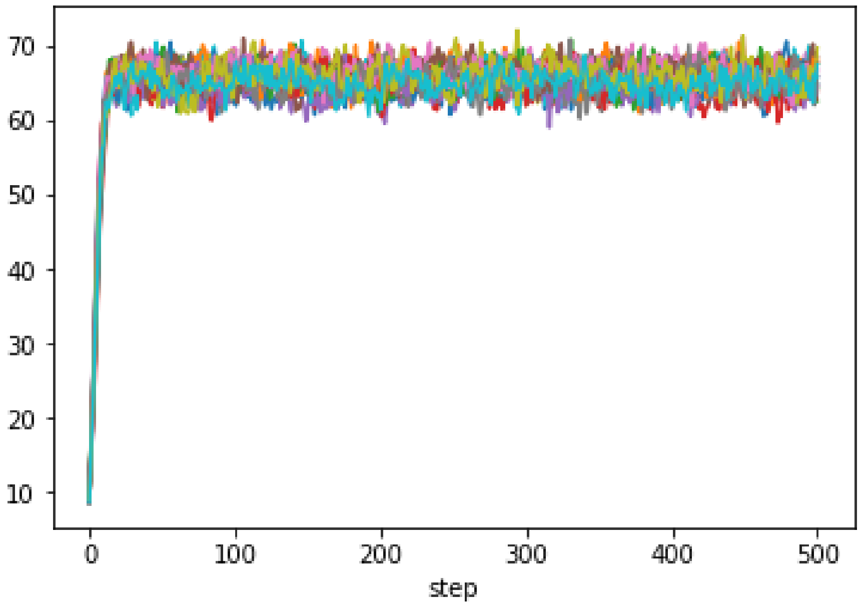

Figure 4: % of Population Infected Over Time
Scenario 3

this number was 10% in scenario 2 and 31% in scenario 3, demonstrating the value of population immunity to non-immune members. At a high level it can then be said that this model replicates the empirical idea of "herd immunity" where general immunity of a group protects non-immune members of the group.

To continue investigating this, one could test whether there is a constant increase in sick time per non-immune turtle or whether as population immunity decreases sick time accelerates. It would also be useful to add in a parameter for the contagiousness of the infection, the radius within which turtles affect each other. Lastly, adding in death or post-infection immunity would be an interesting extension.

1285 words

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.