Cellular Automata, Spatial Interaction, and Network Models

Urban Simulation Coursework
MSc in Smart Cities and Urban Analytics
Center for Advanced Spatial Analytics
University College London
Word Count: 2993

Cellular Automata

Cellular Automata and Agent Based Models are both methods of simulating the evolution of a system over time. The two approaches differ in a fundamental way. Cellular automata define the system as a grid over some area, where each grid cell can take a property. The system evolves through local interactions between the grid cells. Agent Based Models, on the other hand, define the system as a set of agents that move and interact inside an environment (Batty, 2009). The key difference between CA models and ABMs is that in ABMs, the object of interest is not bound to one place, and is able to move throughout the geography of the model.

This paper examines a Cellular Automata model composed of a 400x400 cell grid, where each cell may represent one of four features: Earth, Forest, Water, or City. The features in each cell will change over time according to a set of transition rules based on the current feature in a cell and the features in a neighborhood of that cell. A cell's neighborhood is defined as the 8 cells that share a border or corner with the cell, known as the Moore neighborhood (Batty, 1997). At the start of the model a landscape is created by randomly generating rivers. The rest of the grid is assigned to Earth.

The transition rules provided incorporate both probabilistic and deterministic elements. If the current cell is Earth, there are a set of scenarios that lead to probabilistic assignment of the cell to either Forest, City, Earth, or Water. If the current cell is Forest, City, or Water, then there are a deterministic set of rules leading to the cell's next state. The three scenarios below are the result of changing the probabilities of an earth cell being assigned to either Forest or City, and of triggering the possibility of "Global Warming" flooding. Each scenario is run 10 times, for 500 time steps, using the same river set up for each run within a scenario.

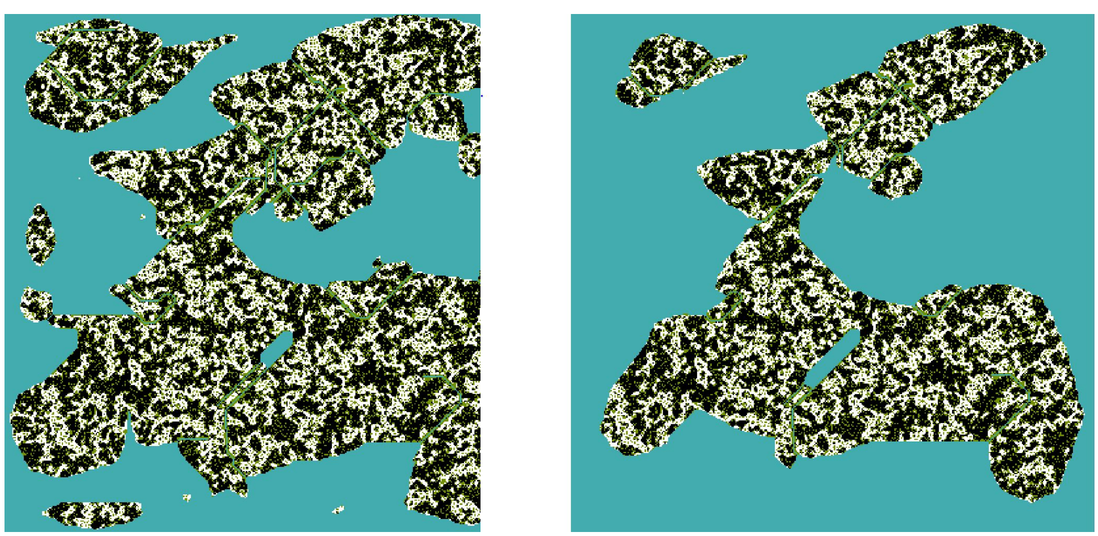
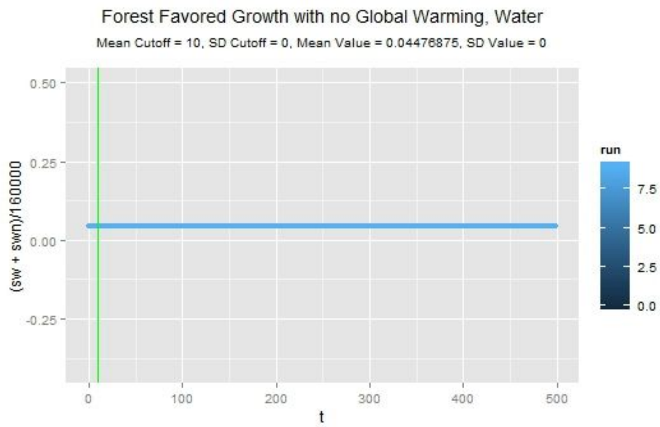
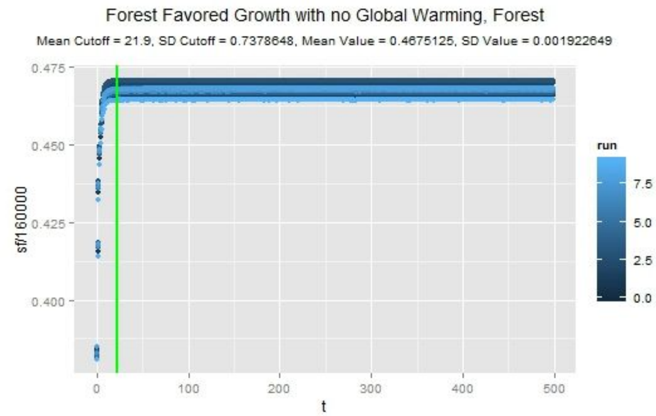
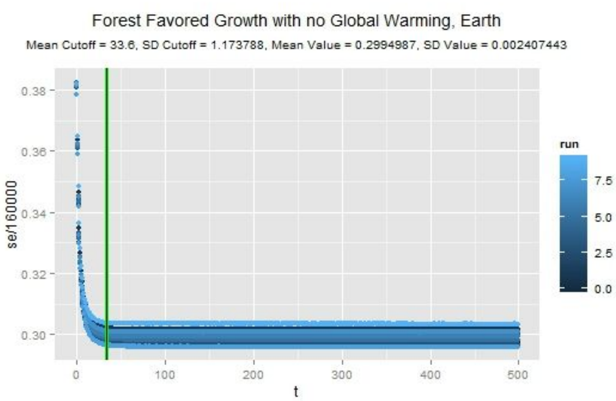


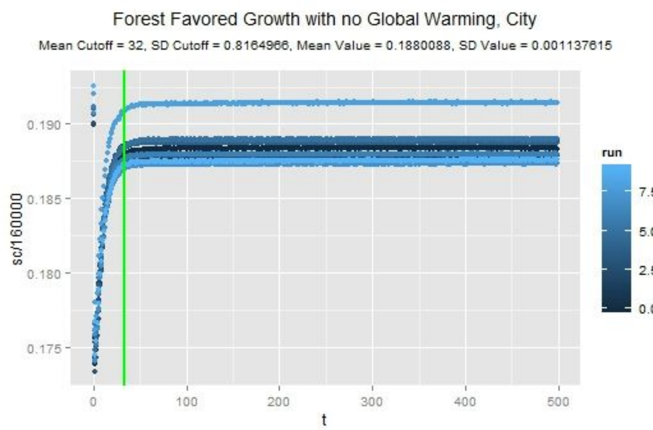
Image from Scenario 3: City Favored, Global Warming

Scenario One: Forest Favored, no flooding







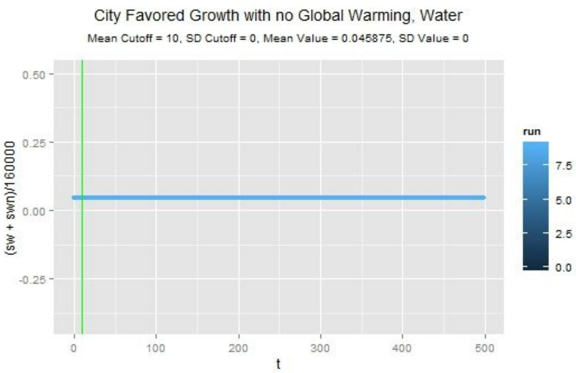


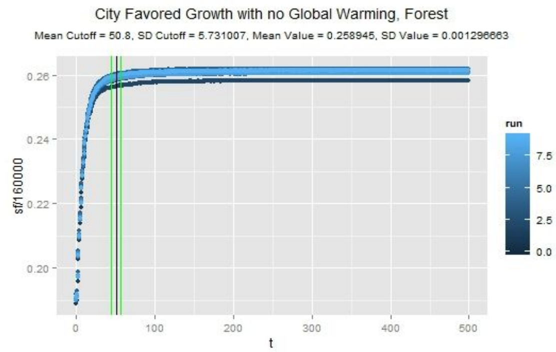
Using the code provided, whenever there was a probability of Earth to transition to Forest, this is set to .4, and whenever there was a probability of Earth to transition to City, this is set to .2. The four graphs above show the percentage of the environment that was assigned to each feature during each of the 10 runs.

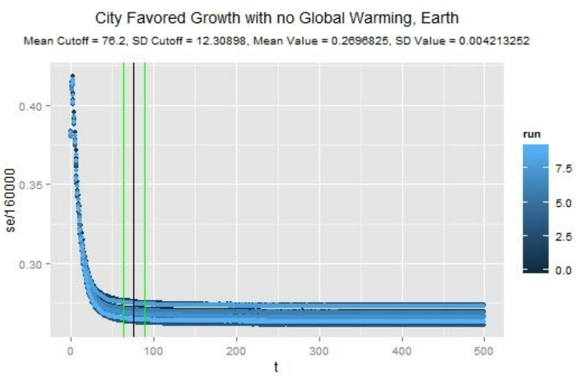
The cutoff indicated in the graphs represents the average number of time steps taken for the proportion of cells assigned to one of the four features to reach a stable state. The percentage change of each feature from one time step to the next was calculated, and a moving average over 10 time periods was calculated for the 10th through the final time steps of each run. The time step at which this moving average fell below .0001 was taken as the cutoff for that run. The cutoff shown on the graphs is the average of the 10 run’s cutoffs, and it is displayed on the graph along with green dashed lines to indicate +- one standard deviation.

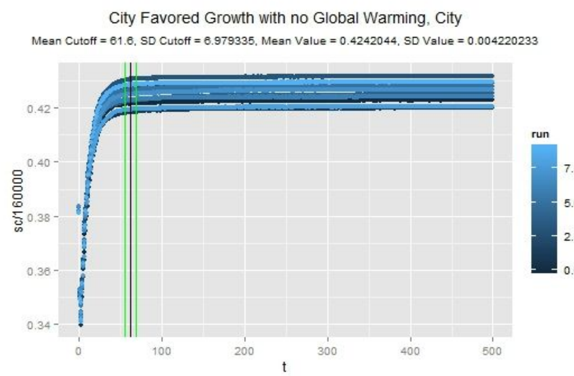
The number of steps needed to reach equilibrium is different for the different features. In this scenario, water is automatically resolved as it doesn’t change. Forest resolves next, at about 22 time steps. City and Earth resolve together, slightly later at the 32nd to 34th time steps.

Scenario Two: City Favored, no flooding









In this scenario, the probability of Earth transitioning to Forest is always set to .2, and the probability of transitioning to City is always .4. In this case, Forest once again resolves faster, at around 51 time steps. City then resolves at about 62 time steps, and finally Earth around 76 time steps. It is strange that Earth resolves so far after City and Forest in this case, as the only way for it to change is by conversion to either City or Forest. The standard deviations of the cutoff for Earth are high at about 12 time steps, indicating that there could be an outlier that is pulling the mean cutoff point higher. Inspection shows this to be the case:

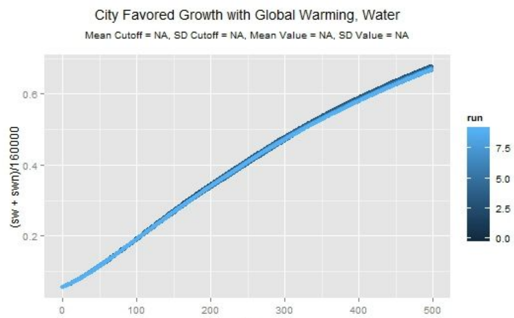
```
> cutOffsEarthCF
 [1]  71  65  71  75  78 108  65  71  79  79
```

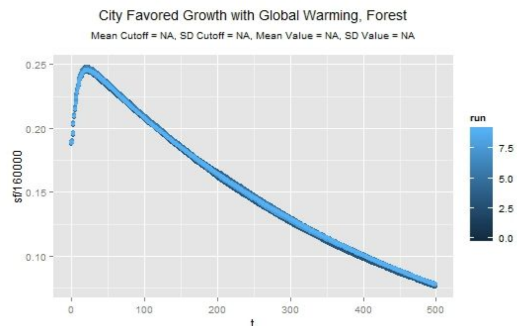
Removing the high value and recalculating the mean and standard deviation gives:

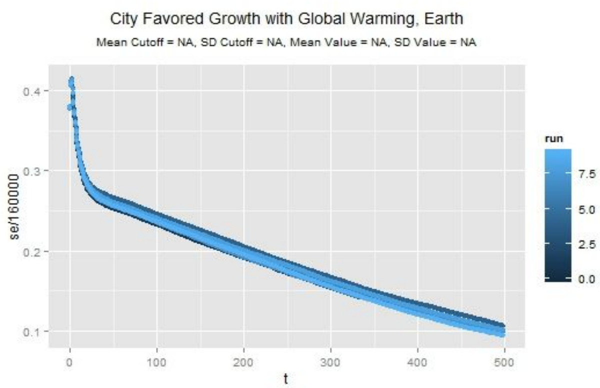
```
mean(cutOffsEarthCF[c(1:5, 7:10)])
[1] 72.66667
> sd(cutOffsEarthCF[c(1:5, 7:10)])
[1] 5.477226
```

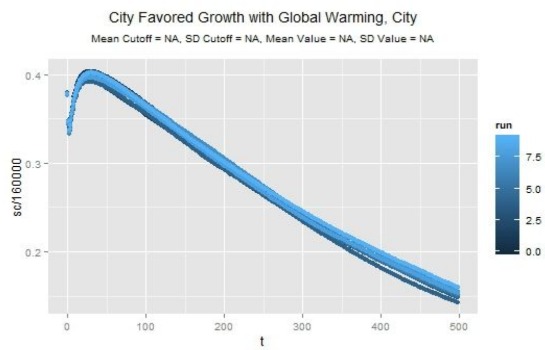
This measure is therefore still sensitive to outliers at this number of runs.

Scenario Three: City Favored, Flooding









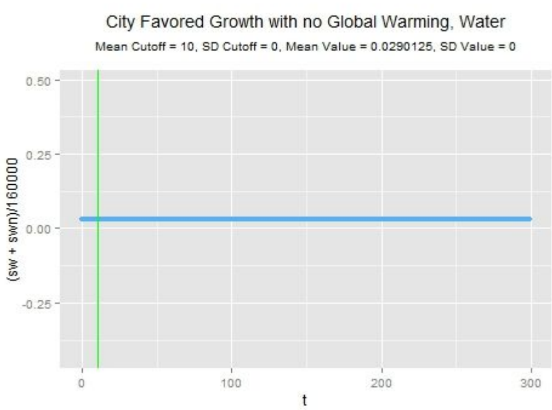
This final scenario is the same as the previous scenario, but with the introduction of the possibility of flooding. For every feature, there is introduced a .2 probability that the feature will turn to water if it is next to more than 3 water cells. This results in a rapid flooding of the landscape. There are no cutoffs recorded in this first trial, as the landscape is progressively covered by more and more water but does not have time to reach stable state. The Forest and City features are quick to cover the non-flooded surface of the area, to be slowly eaten away by the encroaching water. The Earth cells are rapidly converted to City and Forest, but the conversion slows as the conditions for turning Earth to City and Forest (namely - other neighboring City and Forest cells) are undone by these same waters.

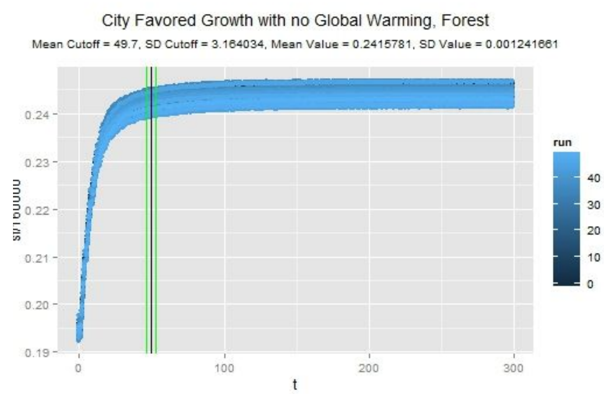
Conclusions: Search for Significance

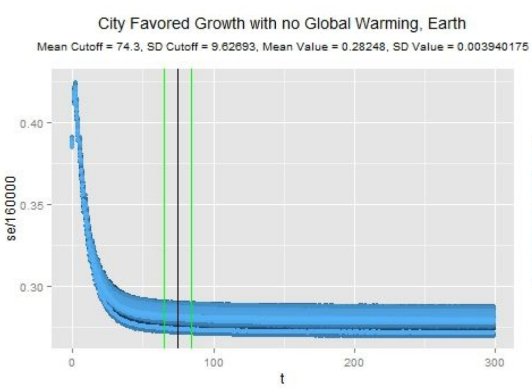
As illustrated in the three scenarios, the time taken to reach steady state is dependent on the transition rules of the Cellular Automata. For a single rule set, there seem to be fairly consistent dynamics. According to the results from Scenario One, the 95% confidence interval around the latest to resolve feature, Earth, predicts full resolution between time steps 32.87 and 34.33. In Scenario Two, the 95% confidence interval around the latest feature to resolve, again Earth, predicts full resolution between time steps 69.09 and 76.25. These values are specific to the initial conditions created by the river layout of a set of runs and the transition rules. After 10 runs of Scenarios One and Two we reach consistent final values for the proportions of each feature, with standard deviations under .5%. 10 runs seems therefore to be sufficient to make statements about the end dynamics to within 1% accuracy.

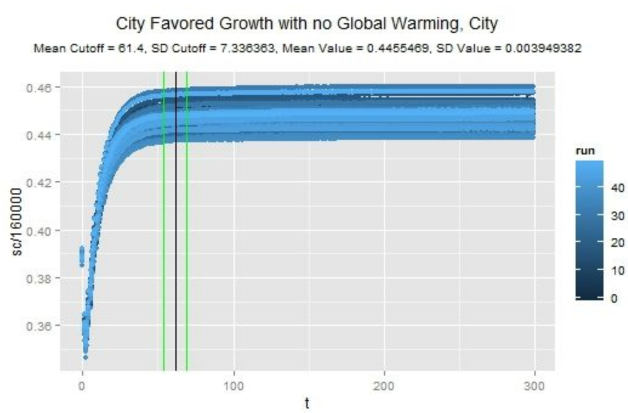
The figures below show two further experiments. The second Scenario Two illustrates the fact that increasing the number of runs from 10 to 50 reduces the standard deviation of cutoff time for Forest and Earth features, while it increases standard deviation of City feature resolution. More runs do not substantially reduce standard deviation of final proportions for any features. The second Scenario Three was run by allowing for the redrawing of the river system each run, and illustrates how the runs take substantially different amounts of time to converge to the stable state of total flooding.

Scenario 2: v2

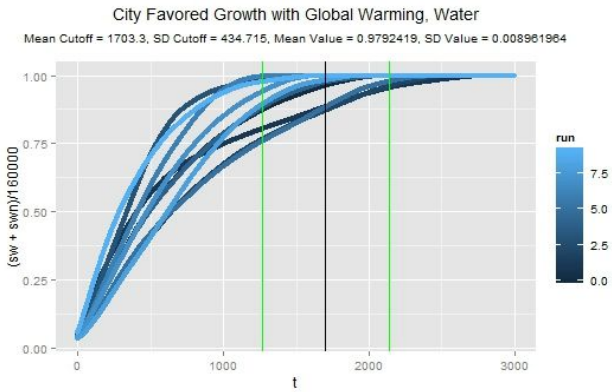


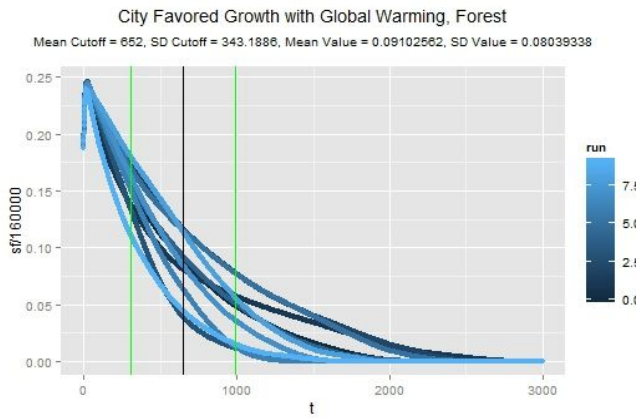


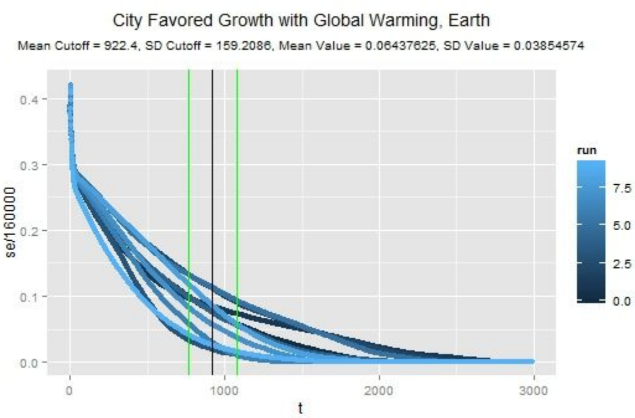


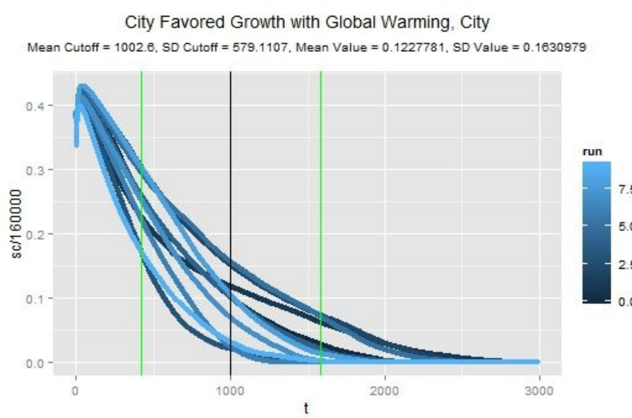


Scenario 3: v2









Spatial Interaction

Spatial interaction models can be used to simulate city's competition for people in an environment where locations for living are unequal in their intrinsic attractiveness and the ease by which you can move from one location to the others for employment (Fry and Wilson, 2012).

In this model, people want jobs and need to live somewhere. Cities are attractive in proportion to the size of the city's population. The assumption here is that people want to live in bigger cities, which may be logical if one assumes that amenities like arts, sports, community groups, etc, are more plentiful in cities with larger population. Ease of travel between cities is determined according to a distance matrix.

Distance Matrix

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	6	33	20	31	66	35
Manchester	33	4	13	6	34	5
Birmingham	18	13	4	19	46	15
Leeds	31	7	19	3	35	11
Glasgow	66	34	46	35	3	35
Liverpool	35	5	15	11	35	2

City Data

	Cities	Population	Employment	K	Wages
1	London	931	500	1950	2322
2	Manchester	260	121	780	1603
3	Birmingham	239	103	625	1566
4	Leeds	168	82	949	1603
5	Glasgow	122	62	504	1606
6	Liverpool	117	48	495	1603

The Spatial Interaction model will simulate people choosing where to live amongst these six cities, and how city populations grow. The equations that govern the model act in three stages:

1) Calculate Interactions Between Cities: $T_{ij} = B_j E_j P_i^{\alpha} exp(-\beta c_{ij})$ $\quad B_j = \frac{1}{\sum_i P_i^{\alpha} exp(-\beta c_{ij})}$

2) Calculate Revenues and Costs: $O_i = \sum_j T_{ij} w_j$ $\quad C_i = k_i P_i$

3) Update Population Sizes: $\Delta P_i(t,\ t+1) = \varepsilon[O_i(t) - C_i(t)]P_i(t)$

E_j	The employment available in city j
P_i	The population of city i
B_j	Proportionality factor for city j
T_{ij}	Population that works in city j that lives in city i
O_i	Revenues to city i
C_i	Cost of living in city i

ε	Population elasticity
β	Resistance to travel
c_{ij}	Cost of commute from city j to i
k_i	Cost of living for an individual in city i
α	Importance of city attractiveness
w_j	Wages in city j

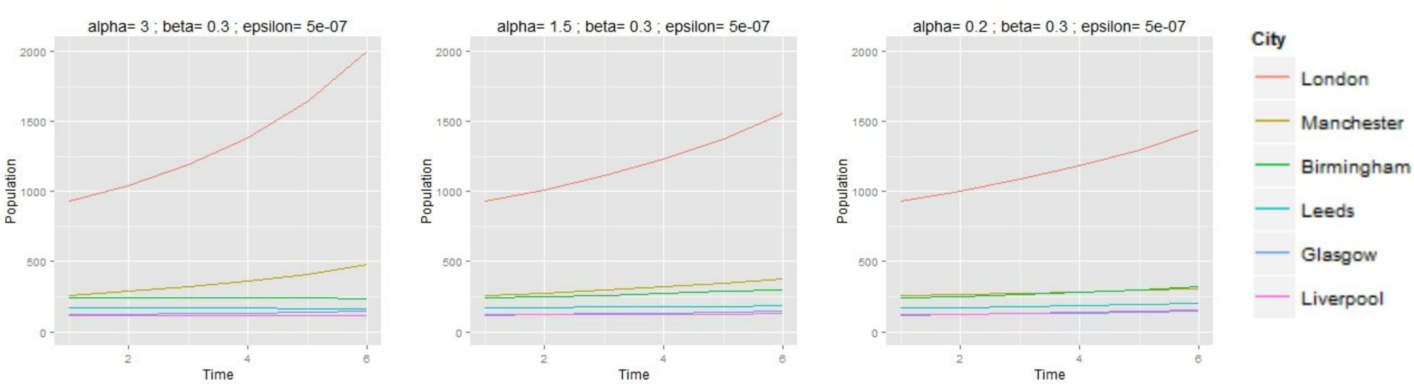
The Model: Explained

1) Trips (T_{ij}) in this model represent journeys from living site i to work site j. Employment (E_j) drives this model by setting the number of available trips. People respond by choosing to live to a city that is attractive (P_i^α) and close to their workplace ($exp(-\beta c_{ij})$). The choice is based off of the relative attractiveness of a city compared to all other options: ($\frac{P_i^\alpha exp(-\beta c_{ij})}{\sum_l P_l^\alpha exp(-\beta c_{lj})}$).
2) Revenues (O_i) are equal to the wages brought home to city i by workers in cities j. This is calculated by multiplying the number of trips from i to j (T_{ij}) by the wages in j (w_j). Costs to a city are represented by the product of the total population P_i with the cost of living in city i (k_i).
3) Cities that bring in more wages than the cost of living will expand:
$$O_i(t) - C_i(t) > 0 \quad \Rightarrow \quad \Delta P_i(t,\ t+1) > 0$$
and cities with fewer wages than the cost of living will contract:
$$O_i(t) - C_i(t) < 0 \quad \Rightarrow \quad \Delta P_i(t,\ t+1) < 0$$
Both expansion and contract occurs proportionately to the current city size (P_i) and an expansion factor (ε). The number of jobs available in a city will grow or shrink in proportion to the population size, and in this model is set to 47% of the population.

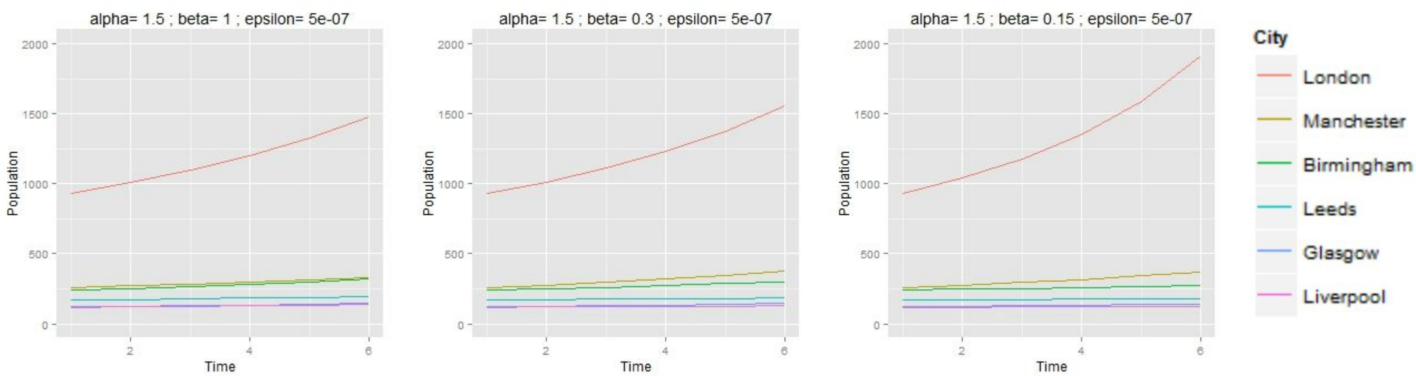
The output of the model will change depending on the values of α, β, and ε. The following experiments document the effect of changes in the these three parameters. Credit is given to Duccio Aiazzi for contributing an implementation of the Spatial Interaction model in R code.

Results

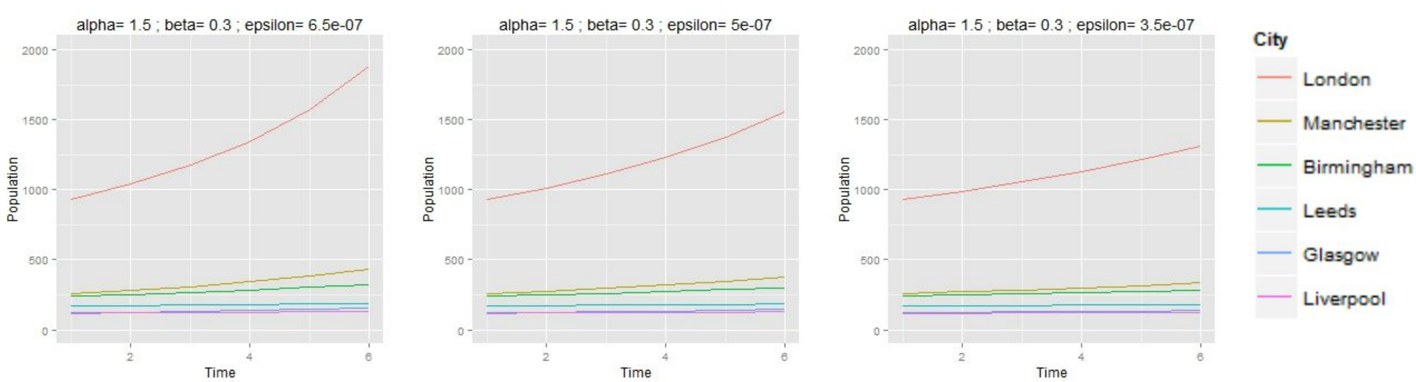
Higher values of α will lead to cities with bigger populations being preferenced more heavily. This is shown below, as when $\alpha = 3$ and we observe dramatic growth in the population size of London. London’s growth is less pronounced when α is decreased, as the city is no longer as attractive for people. Manchester also experiences significantly higher growth with $\alpha = 3$ compared to the other scenarios.



Higher values of β mean that people will be more sensitive to commute costs, and therefore less willing to travel for work. This is reflected in the faster growth of London when $\beta = .15$ compared to $\beta = 1$, as the attractive nature of the big city is less of a pull when people are averse to traveling long distances to work in other cities.



Higher values of epsilon correspond to larger responses in population growth or decline according to the balance of city revenues and costs.



Slow and Fast Dynamics

The slow dynamics of the model are the increase or decrease of each city's population between time steps. This expansion occurs between time steps in a model. The fast dynamic of the model are people's choice of where to live for a given job.

The α and β parameters determine the fast dynamics of the model, and the ε parameter determines the slow dynamics of the model. The following examples set each parameter to 0 while varying the others to see how the model reacts. In the matrices, the columns represent cities where employees live, and the rows represent the city in which the job is located. The figures have been normalized to represent a proportion of the jobs in a row city that are occupied by people living in the column cities.

If $\alpha = 0$, then there is no difference between cities based on their population. Competitiveness would be based only on distance, and fast dynamics do not change with growing population. This will still lead to population changes as some cities will be more profitable than others, and slow dynamic will still occur. Here step 1 shows the fast dynamics that lead to Iter2 of the city's population, and step 5 shows the dynamics that lead to Iter6 of the city's population.

Alpha = 0 , Beta = 0.15 , Epsilon = 5e-07

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	942.80708165	955.39135457888	968.831773178883	983.24875043935	998.800501059729
Manchester	260	270.1854168	281.398367760145	293.797989628583	307.587413990311	323.005938111719
Birmingham	239	259.303627185	282.057097742613	307.674625855652	336.670873101658	369.677100911018
Leeds	168	174.97329528	182.538285045276	190.779058861855	199.800646314572	209.723337593125
Glasgow	122	125.81899223	129.881436299073	134.212765646346	138.841632409923	143.798147767263
Liverpool	117	124.34421519	132.518769170945	141.669035026069	151.969097730482	163.636851137852

Alpha = 0 , Beta = 0.15 , Epsilon = 5e-07 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.85	0.01	0.1	0.02	0	0.01
Manchester	0	0.35	0.09	0.26	0	0.3
Birmingham	0.07	0.15	0.59	0.06	0	0.11
Leeds	0.01	0.28	0.05	0.51	0	0.15
Glasgow	0	0.01	0	0.01	0.97	0.01
Liverpool	0	0.31	0.07	0.13	0	0.49

Alpha = 0 , Beta = 0.15 , Epsilon = 5e-07 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.85	0.01	0.1	0.02	0	0.01
Manchester	0	0.35	0.09	0.26	0	0.3
Birmingham	0.07	0.15	0.59	0.06	0	0.11
Leeds	0.01	0.28	0.05	0.51	0	0.15
Glasgow	0	0.01	0	0.01	0.97	0.01
Liverpool	0	0.31	0.07	0.13	0	0.49

Alpha = 0 , Beta = 0.3 , Epsilon = 5e-07

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	1000.41519242	1080.54110435806	1173.99029590661	1284.26195633297	1416.16296478588
Manchester	260	266.4690782	273.431590911422	280.944853766157	289.082457502028	297.922964897818
Birmingham	239	251.898907675	266.246615518022	282.30016110138	300.378558639287	320.895295565073
Leeds	168	173.65225416	179.670721341594	186.091139414178	192.955899102504	200.319992523599
Glasgow	122	125.85450765	129.956260492134	134.329965145958	139.002409180182	144.006013524589
Liverpool	117	123.285396195	130.145394009864	137.65719835775	145.91575082258	155.031076446248

Alpha = 0 , Beta = 0.3 , Epsilon = 5e-07 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.98	0	0.01	0	0	0
Manchester	0	0.42	0.03	0.23	0	0.31
Birmingham	0.01	0.06	0.88	0.01	0	0.03
Leeds	0	0.22	0.01	0.71	0	0.06
Glasgow	0	0	0	0	1	0
Liverpool	0	0.27	0.01	0.04	0	0.67

Alpha = 0 , Beta = 0.3 , Epsilon = 5e-07 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.98	0	0.01	0	0	0
Manchester	0	0.42	0.03	0.23	0	0.31
Birmingham	0.01	0.06	0.88	0.01	0	0.03
Leeds	0	0.22	0.01	0.71	0	0.06
Glasgow	0	0	0	0	1	0
Liverpool	0	0.27	0.01	0.04	0	0.67

Alpha = 0 , Beta = 1 , Epsilon = 5e-07

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	1006.77224662	1095.37686053663	1200.27144525439	1326.21657640618	1479.97252150027
Manchester	260	265.3907906	271.026753654104	276.930303407601	283.122229707762	289.623685878887
Birmingham	239	251.62962518	265.63021292756	281.23212765968	298.720076805919	318.450849658809
Leeds	168	173.62501968	179.614395518494	186.002349269313	192.829567559722	200.14080743953
Glasgow	122	125.85450948	129.956264242315	134.329970913973	139.003495741674	144.007123394891
Liverpool	117	123.12628848	129.817197485895	137.149399956162	145.215158368845	154.121820717218

Alpha = 0 , Beta = 1 , Epsilon = 5e-07 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	1	0	0	0	0	0
Manchester	0	0.67	0	0.09	0	0.24
Birmingham	0	0	1	0	0	0
Leeds	0	0.02	0	0.98	0	0
Glasgow	0	0	0	0	1	0
Liverpool	0	0.05	0	0	0	0.95

Alpha = 0 , Beta = 1 , Epsilon = 5e-07 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	1	0	0	0	0	0
Manchester	0	0.67	0	0.09	0	0.24
Birmingham	0	0	1	0	0	0
Leeds	0	0.02	0	0.98	0	0
Glasgow	0	0	0	0	1	0
Liverpool	0	0.05	0	0	0	0.95

If $\beta = 0$, then there is no friction to distance, and people will live purely based off of the attractiveness of the city. This leads to an equal proportion of people from each city's employment base choosing to live in each potential home city. The value of α will determine how extreme the preference for large cities is. The slow dynamic will still occur, and will alter the fast dynamic as large cities grow and become more attractive.

Alpha = 0.2 , Beta = 0 , Epsilon = 5e-07

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	706.299281695	588.245692846286	514.163804587172	463.569467444298	427.371521130743
Manchester	260	284.9343276	307.242411034306	328.694025802605	350.135294265217	372.136115703921
Birmingham	239	264.34651404	288.075109856097	311.754341798017	336.185036588278	361.934355484126
Leeds	168	183.80706876	198.249366674463	212.391235945841	226.749389733591	241.683660043112
Glasgow	122	135.29167247	148.339742436704	161.910444442669	176.439435585842	192.291311718583
Liverpool	117	129.72376755	142.232770065484	155.258979528487	169.220660994148	184.473091407805

Alpha = 0.2 , Beta = 0 , Epsilon = 5e-07 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.22	0.17	0.17	0.16	0.15	0.14
Manchester	0.22	0.17	0.17	0.16	0.15	0.14
Birmingham	0.22	0.17	0.17	0.16	0.15	0.14
Leeds	0.22	0.17	0.17	0.16	0.15	0.14
Glasgow	0.22	0.17	0.17	0.16	0.15	0.14
Liverpool	0.22	0.17	0.17	0.16	0.15	0.14

Alpha = 0.2 , Beta = 0 , Epsilon = 5e-07 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.19	0.18	0.17	0.16	0.15	0.15
Manchester	0.19	0.18	0.17	0.16	0.15	0.15
Birmingham	0.19	0.18	0.17	0.16	0.15	0.15
Leeds	0.19	0.18	0.17	0.16	0.15	0.15
Glasgow	0.19	0.18	0.17	0.16	0.15	0.15
Liverpool	0.19	0.18	0.17	0.16	0.15	0.15

Alpha = 1.5 , Beta = 0 , Epsilon = 5e-07

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	1079.24313931	1282.16600278466	1571.04020754252	2002.05640013329	2682.82586765591
Manchester	260	270.0886734	279.797693054127	288.630695777699	295.955413600678	301.050671583082
Birmingham	239	248.823035035	258.528335646898	267.74018775336	275.939016116584	282.510570564816
Leeds	168	169.24881036	170.035490095342	170.26182776691	169.836151656062	168.67831163337
Glasgow	122	123.6267907	125.087451481955	126.332655438168	127.305180047937	127.952180551571
Liverpool	117	118.460928105	119.768570022046	120.880845891579	121.746778925874	122.31898829271

Alpha = 1.5 , Beta = 0 , Epsilon = 5e-07 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.69	0.1	0.09	0.05	0.03	0.03
Manchester	0.69	0.1	0.09	0.05	0.03	0.03
Birmingham	0.69	0.1	0.09	0.05	0.03	0.03
Leeds	0.69	0.1	0.09	0.05	0.03	0.03
Glasgow	0.69	0.1	0.09	0.05	0.03	0.03
Liverpool	0.69	0.1	0.09	0.05	0.03	0.03

Alpha = 1.5 , Beta = 0 , Epsilon = 5e-07 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.86	0.05	0.04	0.02	0.01	0.01
Manchester	0.86	0.05	0.04	0.02	0.01	0.01
Birmingham	0.86	0.05	0.04	0.02	0.01	0.01
Leeds	0.86	0.05	0.04	0.02	0.01	0.01
Glasgow	0.86	0.05	0.04	0.02	0.01	0.01
Liverpool	0.86	0.05	0.04	0.02	0.01	0.01

Alpha = 3 , Beta = 0 , Epsilon = 5e-07

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	1285.952794665	1834.16332718796	2705.83451892437	4183.85191557917	6978.40383783511
Manchester	260	252.1842518	242.459523096947	232.420950711588	222.781330068589	213.773887060731
Birmingham	239	233.87577308	227.267943070318	220.251288700523	213.338776223199	206.726902413494
Leeds	168	162.5020866	156.946812555775	151.583088000333	146.505943985737	141.732759365977
Glasgow	122	120.45874533	118.841306752288	117.21069773534	115.599932647317	114.02254659413
Liverpool	117	115.594956945	114.125012359012	112.646845562284	111.185219588017	109.752355575511

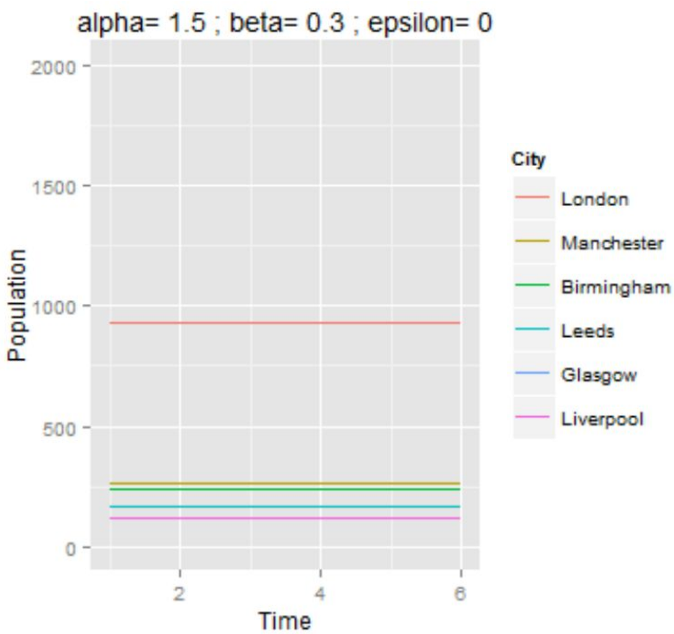
Alpha = 3 , Beta = 0 , Epsilon = 5e-07 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	0.95	0.02	0.02	0.01	0	0
Manchester	0.95	0.02	0.02	0.01	0	0
Birmingham	0.95	0.02	0.02	0.01	0	0
Leeds	0.95	0.02	0.02	0.01	0	0
Glasgow	0.95	0.02	0.02	0.01	0	0
Liverpool	0.95	0.02	0.02	0.01	0	0

Alpha = 3 , Beta = 0 , Epsilon = 5e-07 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	1	0	0	0	0	0
Manchester	1	0	0	0	0	0
Birmingham	1	0	0	0	0	0
Leeds	1	0	0	0	0	0
Glasgow	1	0	0	0	0	0
Liverpool	1	0	0	0	0	0

If $\varepsilon = 0$, then there will be no slow dynamics - no change in population - no matter what is happening in the fast dynamics of people's employment choices. There will still be fast dynamics, but these will be constant through all iterations as both the populations of cities and the distances between them remain constant.



Alpha = 1.5 , Beta = 0.3 , Epsilon = 0

	iter1	iter2	iter3	iter4	iter5	iter6
London	931	931	931	931	931	931
Manchester	260	260	260	260	260	260
Birmingham	239	239	239	239	239	239
Leeds	168	168	168	168	168	168
Glasgow	122	122	122	122	122	122
Liverpool	117	117	117	117	117	117

Alpha = 1.5 , Beta = 0.3 , Epsilon = 0 , Step = 1

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	1	0	0	0	0	0
Manchester	0	0.64	0.04	0.18	0	0.14
Birmingham	0.1	0.06	0.83	0.01	0	0.01
Leeds	0	0.35	0.01	0.61	0	0.03
Glasgow	0	0	0	0	1	0
Liverpool	0	0.53	0.02	0.05	0	0.4

Alpha = 1.5 , Beta = 0.3 , Epsilon = 0 , Step = 5

	London	Manchester	Birmingham	Leeds	Glasgow	Liverpool
London	1	0	0	0	0	0
Manchester	0	0.64	0.04	0.18	0	0.14
Birmingham	0.1	0.06	0.83	0.01	0	0.01
Leeds	0	0.35	0.01	0.61	0	0.03
Glasgow	0	0	0	0	1	0
Liverpool	0	0.53	0.02	0.05	0	0.4

Networks

Networks are composed of nodes and edges. Nodes usually represent an entity, and edges represent the relationships or connections between these entities. In this example there are 404 nodes representing stations on the rail network in London. Edges represent connections between stations. We will examine which nodes impact the network most by analyzing the betweenness, closeness, clusters, and communities of this network as nodes are removed. (“BBC - London - Travel - London Underground Map”)

Betweenness Centrality: Betweenness centrality measures the proportion of shortest paths on the network that go through a given node.

Closeness Centrality: Closeness centrality is an inverse sum of the distance from one node to all other nodes in the network.

The analysis below will assess which nodes have the most impact on the tube network. Impact will be defined as the degree to which the hierarchy of the network will be changed if that node is removed. The approach will analyze the changing rank of nodes in the network according to the Betweenness and Centrality measures as nodes are removed, as well as the emergence of **clusters** and **communities** in the network.

Clusters: Two nodes are considered part of the same cluster if there is some path of nodes and edges to get from one to the other. The London tube network begins as one cluster, and it is possible to travel from one node to any other. When nodes are deleted, this may result in splitting the network into multiple clusters that may not reach each other.

Communities: The Girvan-Newman algorithm for community detection (Fortunato and Lancichinetti, 2009) is used to identify communities. The algorithm works by first measuring the betweenness of all edges in the network. It proceeds to delete edges in the order of their betweenness, and assess the clusters that are created according to a measure called *modularity*. It stops splitting the network when *modularity* is maximized. A description of modularity may be found in the reference above.

Results

The table below shows the highest ranking nodes in the **betweenness** and **closeness** measures. Kings Cross St. Pancras is the highest rank node in both lists, and obviously plays an important role in the network for that reason. Euston and Baker Street also appear in the top ten for both measures, but all others are different. There is only one **cluster** in the network, as all nodes are initially connected. The Girvan-Newman algorithm groups the 404 nodes into 12 **communities**, with the following number of nodes in each: 69, 50, 32, 23, 31, 41, 41, 25, 42, 17, 20, 13.

Closeness and Betweenness, Original Network

	top10Clo.names	top10Clo	top10Btw.names	top10Btw
1	Kings Cross St. Pancras	1.0000000	Kings Cross St. Pancras	23633
2	Oxford Circus	0.9939745	Surrey Quays	19683
3	Euston	0.9868959	Canada Water	18506
4	Baker Street	0.9852937	Baker Street	17817
5	Bond Street	0.9787803	Euston	17018
6	Euston Square	0.9750021	Stratford	14230
7	Angel	0.9739506	Finchley Road	14037
8	Warren Street	0.9718837	Earls Court	14009
9	Great Portland Street	0.9640228	West Brompton	13882
10	South Hampstead	0.9626777	Canary Wharf	13536

There are therefore 100% of the nodes in the one cluster, and the following percentages of nodes in each of the identified communities: 0.17, 0.12, 0.08, 0.06, 0.08, 0.10, 0.10, 0.06, 0.10, 0.04, 0.05, 0.03.

The analysis that follows describes the results of sequentially deleting nodes from the network according to the highest ranking betweenness or closeness. For both betweenness and closeness, the first node to delete is Kings Cross St. Pancras. The betweenness and closeness will then be recalculated for the new network, and the new highest ranking nodes for each measure will be deleted. This process is repeated 10 times.

As nodes are removed, the hierarchy in the network changes. This change will be measured in the following way. At each step, the network's closeness and betweenness measures are recorded and the nodes are ranked according to each. Every node is then given a score according to the percent of total nodes that are ranked below them. As nodes are removed from the network, the new measures, rankings, and percentiles are recalculated.

The **impact** of removing a node η on a measure is defined as: The sum of the absolute value of the differences between all nodes' percentile scores before and after removal of node η.

Impact on Betweenness and Closeness

	RemovedStation.Btw	Impact.Btw	RemovedStation.Clo	Impact.Clo
1	Kings Cross St. Pancras	22.2906	Kings Cross St. Pancras	23.6652
2	Canada Water	19.6778	Oxford Circus	13.9328
3	Waterloo	18.127	Embankment	13.7096
4	Bond Street	5.0018	Earls Court	15.927
5	Earls Court	16.8616	Shepherds Bush	40.6614
6	Baker Street	14.9664	Bank	28.9226
7	Willesden Junction	33.5834	West Hampstead	38.5028
8	Shepherds Bush	49.0302	Canada Water	99.606
9	Highbury and Islington	17.496	West Croydon	16.4104
10	West Ham	27.4278	Stratford	18.6798

Removing King's Cross St. Pancras results in an impact of 22.2906. The removal resulted in some nodes increasing in betweenness relative to the rest of the network, and others decreasing. For example, Canada Water moved from having the 3rd ranked betweenness to having the 1st ranked. Canada Water therefore contributed $abs(\frac{402}{404} - \frac{403}{403}) \approx 0.0050$ to the score.

Removal of Bond Street resulted in a much lower impact. Bond street may therefore be considered less impactful than all other stations listed above in terms of betweenness. It must be noted that the impact of Bond Street is being measured after Kings Cross St. Pancras, Canada Water, and Waterloo stations have all been removed. At this point it has become the highest ranked station regarding betweenness, but removing it does not result in comparatively significant changes in the hierarchy of the rest of the network.

Shepherds Bush is notable in that it has high impact in both the betweenness and closeness structure of the network. The earlier removal of Willesden Junction, Earls Court, and Baker Street made Shepherds Bush important to connecting West London with the rest of the network.

The same sequence of deletions was repeated, and impact was measured on the clusters and communities of the network. The **impact** of removing a node η on is defined as: The sum of the absolute value of the differences between the ordered percentages of nodes belonging to the set of clusters or communities before and after removal of node η.

Impact on Clusters and Communities

	RemovedStation.Btw	Impact.Clusts.Btw	Impact.Comms.Btw	RemovedStation.Clo	Impact.Clusts.Clo	Impact.Comms.Clo
1	Kings Cross St. Pancras	0	0.0924379038400118	Kings Cross St. Pancras	0	0.0924379038400118
2	Canada Water	0	0.408367591323778	Oxford Circus	0	0.304173919484464
3	Waterloo	0	0.225902904430466	Embankment	0	0.195878463046364
4	Bond Street	0	0.0738778054862843	Earls Court	0	0.178628428927681
5	Earls Court	0	0.168496240601504	Shepherds Bush	0	0.1031328320802
6	Baker Street	0.0100502512562814	0.177390712963313	Bank	0	0.352426291860304
7	Willesden Junction	0.0655164993734415	0.115058921813096	West Hampstead	0	0.371682088021974
8	Shepherds Bush	0.328473653410681	0.0893952115614584	Canada Water	1.03535353535354	0.0045289163677073
9	Highbury and Islington	0.00102288709883645	0.019549929676512	West Croydon	0.0042321953714359	0.0490857946554149
10	West Ham	0.635545845916597	0.145935873546231	Stratford	0.368103836021333	0.0436419713422862

Removing by betweenness, there remained only one cluster through the removal of Earls Court, after which small clusters were formed by the removal of Baker Street and Willesden Junction. The removal of Shepherds Bush led to a large cluster breaking off from the network.

Removing by closeness, there was only one cluster through the removal of Canada Water, at which point the network split into three large components. This can be known from the fact that a value greater than 1 can only be obtained when the graph splits into more than two components, and high values will be obtained from relatively even splits.

Conclusion

Based on the measures above, the nodes that have the most impact on the network are Kings Cross St. Pancras, Canada Water, Earls Court, Willesden Junction, Baker Street, and Shepherds Bush.

References

Batty, M., 2009. Urban Modeling. Int. Encycl. Hum. Geogr. 51–58.

Batty, M., 1997. Cellular Automata and Urban Form: A Primer Journal of the American Planning Association 266–274.

BBC - London - Travel - London Underground Map [WWW Document], n.d. URL http://www.bbc.co.uk/london/travel/downloads/tube_map.html (accessed 4.25.16).

Fortunato, S., Lancichinetti, A., 2009. Community detection algorithms: A comparative analysis. Phys. Rev.

Fry, H., Wilson, A., 2012. A dynamic global trade model with four sectors: food, natural resources, manufactured goods and labour. UCL Work. Pap. Ser.