

# Complex Patterns of Multiscale Human Mobility in United Kingdom

Defne Askar<sup>1,\*</sup>, Thomas House<sup>2</sup>,

1 Centre for Complexity Science, University of Warwick, Coventry, United Kingdom

2 Warwick Mathematics Institute and Department of Biological Sciences, University of Warwick, Coventry, United Kingdom

\* E-mail: D.Askar@warwick.ac.uk

## Abstract

Many of the real systems such as social, technological, biological and informational systems can be described as complex networks. In this study we are interested in two types of movement networks that occur in the United Kingdom, which are internal migration and commuting. United Kingdom 2001 Census data has been used in order to analyze movement networks through complex network tools at different spatial scales. Nodes represent locations in movement networks rather than individuals and edges store the total number of movements from one location to another. At larger scale, we observed high correlations between in- and out-degree of each location; however at smaller level degree correlations were not that informative so we analyzed the network motif structures, which are small patterns of interconnection that occur more frequently than one would expect from random networks. We observed different motif structures for internal migration and commuting networks. The results for reciprocity, which characterises the simplest non-trivial motif, are also presented in this study. A major motivation for studying movement networks is to understand the influence of network characteristics on epidemic spreading. The epidemic threshold for a network is closely related to the largest eigenvalue of its adjacency matrix, which we calculate.

## Introduction

Two claimed features of complex systems, the scale-free and small-world properties, increased the attention given to complex networks in the past 10-15 years. Specific models of these complex network properties were discovered in the late 1990s [1]. Small world networks have highly clustered connections, and the distance between any nodes are relatively short. This feature shows fast communication locally and globally [2]. On the other hand, scale free networks have many nodes with few connection and few nodes with many connections. The fraction of nodes having  $k$  edges,  $p(k)$ , decays with a power law degree distribution [3, 4].

Complex networks describe a wide range of real systems from different fields. Many social, technological, biological and informational systems can be described as complex networks [5]. Social networks include friendship networks, corporate network, email networks. Technological networks includes transportation, energy, communication and sensor networks. Biological networks include transcriptional gene regulation, ecological networks and protein interactions. Finally, information networks include World Wide Web, Twitter and peer-to-peer networks. In this paper, we are interested in two types of movement networks in the United Kingdom, which are internal migration and commuting.

We want to study commuting and migration systems through a complex network analysis. Quantitative analyses of complex networks are mostly based on graph theory. In graph theory, the graph  $G = (V, E)$  is a collection  $V$  of nodes that are linked to each other by set of edges  $E$ . In social networks, individuals are represented as nodes and the relationship between people is represented with edges. On the other hand, in movement networks nodes represent the locations rather than individuals and edges store the total number of movements from one location to another [6].

Better transportation systems have facilitated and increased commuter movements, so modeling and analyzing commuting behavior has been the focus of various studies [7]. Two related studies in the literature have been done by Patuelli *et al.* (2007) [8] and De Montis *et al.* (2010) [9]. Patuelli *et al.* offers a network analysis of commuting flows in Germany. They studied the spatial evolution of

the commuting network over time, and they found the commuting network becoming more dense and clustered. On the other hand, De Montis *et al.* analyzed the inter-municipal commuting systems of the Italian islands of Sardinia and Sicily by using a weighted network analysis technique. The second type of movement network, which is the internal migration also studied in the literature by Maier *et al.* [10]. They applied social network analysis to internal migration between US States. The paper offers the topological similarities of social and migration networks. In all of these studies, the movement data had transferred into network structure.

We are interested in finding the patterns of connections between components by looking at different features of movement networks from global to local scale. At global scale, we found the degree distribution, in-degree and out-degree correlation and sink and source dynamics for both of the migration and commuting network at different spatial scales. We concluded that it is more informative to observe the motif structures in networks at fine spatial scale, which are more sparse and have more nodes, rather than degree correlations. Network motifs are the small connected subnetworks in a network, with generally three or four nodes, that have significantly higher frequencies than would be expected in random networks. This is a new local feature that has been offered by R.Milo *et al.* to go beyond the global properties of networks [11]. We then analyzed the three-noded motif structures in commuting and migration networks, and observed how the results differ from each other. The simplest measure of non-trivial motif structure, which is called reciprocity, is also calculated for the movement networks.

One motivation for analysis of movement networks is to understand the spread of infectious diseases. In the literature four main types of movement networks have played fundamental roles [6] : (1) the movement of individuals to and from work [12], (2) the airline transportation network [13,14], (3) the movement of livestock [15] and (4) the movement of dollar bills (activity as a prior to movement of people) [16]. For the disease spread modeling of our study, we just focused on the commuting behavior. We aim to understand the influence of commuting network characteristics on epidemic spreading. In recent studies, it has been proved that the epidemic threshold (whether an epidemic occurs in a population or the disease simply dies out) for a network is closely related to the largest eigenvalue of its adjacency matrix [17]. In general, the spectral property of a network might be a good start point on epidemic spread analysis; however a better model will be more complex and will require more information than the spectrum [18].

The paper is organized as follows: In the Methods section we introduce the construction of migration and commuting networks with the data obtained from United Kingdom 2001 Census, and the different spatial scales defined to analyze it. In the Results section we apply complex network analysis to observe different features of movement networks, and to understand the effect of network characteristics on epidemic spreading. Finally we discuss the results of the network analysis, and its relation to epidemic spreading in the Conclusion section.

## Methodology

### Data

We analyzed the movement networks in two different spatial scales of data according to the NUTS classification [19]. The NUTS Classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU. There are five different region classifications defined with NUTS. NUTS1 stands for Government office regions, NUTS2 stands for groups of unitary authorities, NUTS3 stands for upper tier authorities or groups of lower tier authorities. LAU1 classifies lower tier districts, whereas LAU2 classifies wards within the United Kingdom. At the time when the Census was done, NUTS4 and NUTS5 was standing for LAU1 and LAU2 respectively. An example of the hierarchy for the regions around the University of Warwick is shown in Figure 1 [20].

In this study we are interested in Local Authority Districts (LA) which corresponds to NUTS4 figure 1 in, and Output Area Levels (OA) which are small and local patches with a population of a few hundred

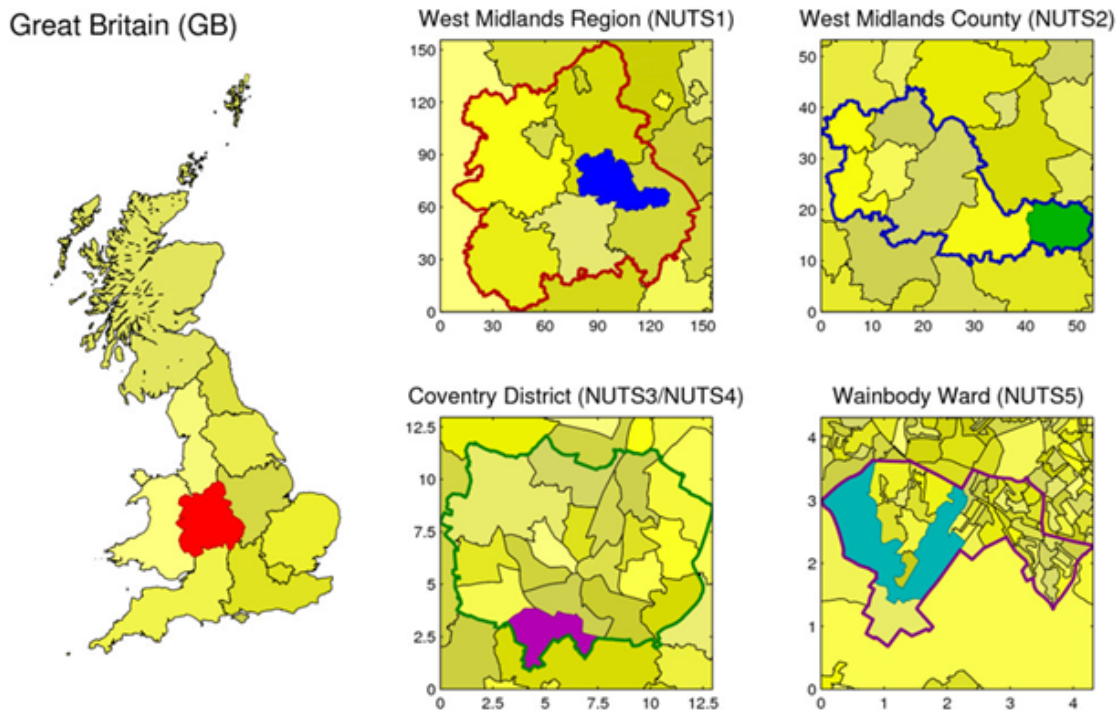
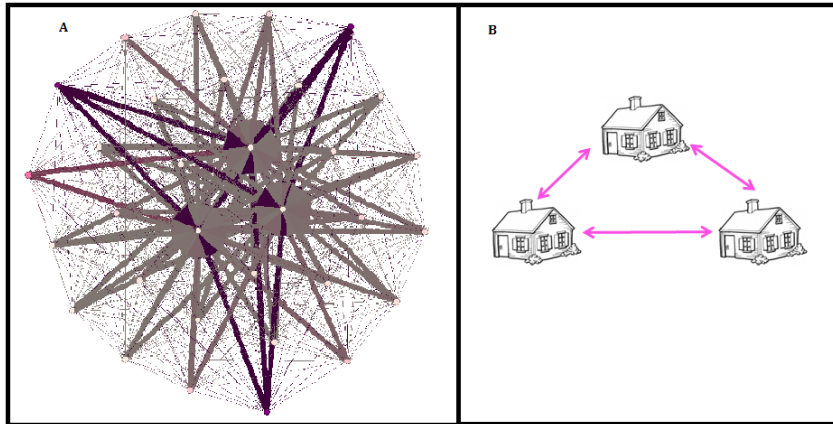


Figure 1. Hierarchy of regions around Output Area 00CQFP0009 [20].

people. When we divide the UK into local authority districts, there are 426 different areas; whereas when we divide the map into output area levels there are 218,038 different areas. We used the United Kingdom 2001 Census data in order to construct migration and commuting networks in the form of origin-destination matrix. The Census provides us a rich and high quality data available on migration and commuting. The data provide disaggregations by various topics such as gender, age, status of migrant, ethnic group, limiting long-term illness, economic activity, and method of travel to work or study [21]. Generally speaking, movement networks capture the information on the total number of people that travel between different nodes, where nodes represent different areas in the map. In the next two paragraphs, the stages of network mapping is explained with three stages: data collection, network graph construction and visualization [22].

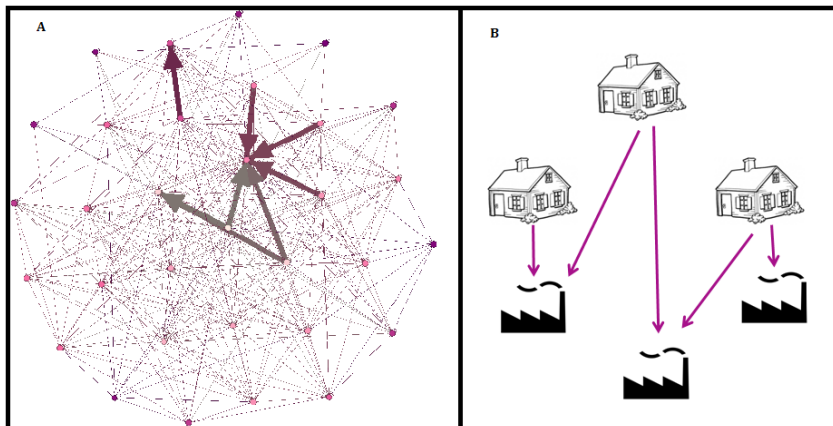
Migration data includes all people resident in the United Kingdom whose address on Census day was different from that one year before the Census [23]. The total resident population is assumed to be constant over two years, which means the residents coming to or going out from the United Kingdom is not considered. If we consider  $M$  as the migration network,  $M_{ij}$  give the number of migrants from district  $i$  to district  $j$ . The visualisation of internal migration network in Scotland at local authority level is given in Figure 2. There are 32 different districts represented as nodes, and 902 connections between each district represented as links. *Force Atlas 2* algorithm is run through *Gephi* in order to visualize the network. As the number of degrees increases, the colour of the nodes are changing through purple, pink and white.

In 2001, each individual in a household specified their employment information and destination of work if they are employed so that they could collect data on commuting. Commuting data includes all people aged between 16 and 74 that are employed and resident in the United Kingdom [24, 25]. Let us consider  $C$  as the commuting matrix, then  $C_{ij}$  gives the number of people that live in district  $i$  and work



**Figure 2. Internal migration network in Scotland.** Figure 2A shows the network generated by *Gephi*, and figure 2B shows that migration happens through districts where households are located.

in district  $j$ . The visualisation of commuting network in Scotland at local authority level is given in Figure 3. There are 32 different nodes like in the migration, however there are fewer connections which is 422. Again we used *Gephi* in order to visualize the network. As the number of degrees increases, the colour of the nodes are changing through purple, pink and white.



**Figure 3. Commuting network in Scotland.** Figure 3A shows the network generated by *Gephi*, and figure 3B shows that commuting occurs through districts where households and workplaces are located.

In LA level, we have weighted directed networks, where  $w_{i,j}$  captures the total number of people moves from area  $i$  to area  $j$ . In OA level, we have almost topological directed network and the matrix is very large, consisting of 218,038 nodes. In the output area level, it is more advantageous to use sparse matrix where we only denote the existing directed edges rather than including all possible edges between nodes. Using sparse matrices facilitates the calculations, since instead of storing  $4,75 \times 10^{10}$  points, most of which are zero, in commuting data we store 6,680,537 directed edges, and in migration data we store 1,760,957 directed edges with sparse matrices.

## Network Measures

We are interested in the patterns of connections in migration and commuting systems, which can be represented as a network as we explained in the previous part. The structure of the network has a big influence on the behaviour of the system. There are different properties of networks which we can measure such as degree, density, clustering coefficient, transitivity, modularity and betweenness. By applying tools from graph theory it is possible to characterize global and local features of the network. In our study, we are interested in the degree distribution, reciprocity and motif structures of the movement networks.

**Degree distribution** is one of the fundamental characteristics of the network structure, which is the frequency distribution of node degrees [5]. The degree distribution of the network is represented with  $p_k$ , which is the probability of a randomly chosen node having  $k$  edges. Directed networks have two different degrees for each node, the in- and out-degree, which are respectively the number of edges incoming into and outgoing from a node. Therefore, we are interested in two different degree distributions in directed networks. The nodes with high degree might play an important role in the system, because of this reason it is worth measuring degree properties for each node.

What will be more interesting to look at next is the in- and out-degree correlation for each location in the system based on the principle of source and sink dynamics (see Figure 4). When a location has more movements to outside it behaves as a source, and if a location has more movements inside it shows a sink behavior.

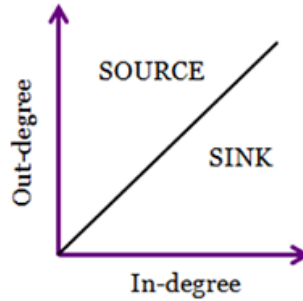


Figure 4. Sink source dynamics in the degree correlation plot.

Moreover, **reciprocity** is another property of networks that might be interesting to calculate, especially because it is a popular subject in social psychology. In social network, it shows how likely you respond positively when you receive positive action. In general, reciprocity is the probability of two vertices in a directed network point to each other. We can calculate the reciprocity  $r$ , by using  $A_{ij}$ : the adjacency matrix and  $m$ : the total number of directed edges in the network by using equation (1). In migration and commuting systems, a higher reciprocity shows there are more movements in the network.

$$r = \frac{\sum_{i,j} A_{ij}A_{ji}}{m} \quad (1)$$

**Network motifs** are the major theme of this project. Motifs gathered attention recently to uncover structural design principles of complex networks. The analysis of motifs are done in directed networks. Motifs, the small patterns of interconnections, may indicate functional properties of network and differ according to different types of networks. For example, biochemical networks that are responsible for regulating the expression of genes in cells contain the three-node motif called *feed-forward loops* and the four-node connected motif called bi-fan far more than one would expect from a random network with similar properties [26]. There are two main tasks to detect network motifs:

1. Generate a small group of proper random networks. Randomized networks preserve the degree distribution of the nodes in the real network. Each node in the randomized networks has the same single node characteristics, which implies same number of incoming and outgoing edges as the corresponding node has in the real network. The randomized graph is obtained by applying a Markov chain algorithm based on starting the real network repeatedly swapping randomly chosen pairs of connections until the network is well randomized [27].
2. Scan each network for all possible  $n$ -node subgraphs. Record the number of occurrences of each subgraphs in the real network  $N_{\text{real}}$  and in randomized networks  $N_{\text{rand}}$ . If the input network is very large (e.g.  $\#of\text{edges} > 100,000$ ) or we want to analyze subgraphs of size  $> 5$ , then we sample subgraphs in the network and calculate their concentrations instead of occurrences to approximate motifs faster [28]. Concentration of a subgraph is the number of occurrences divided by the total number of subgraphs of the same size in the network.

Finally, we compare the real network to suitably randomized networks according to the two statistical measures:  $P_{\text{value}}$  and  $Z_{\text{score}}$ , and select the patterns in the real network occurring significantly higher than those in the randomized networks.  $Z_{\text{score}}$  is calculated with the equation (2) where SD stands for standard deviation. Note that the  $Z_{\text{score}}$  could be positive or negative. A higher value shows that the corresponding motif is more often in the network.

$$Z_{\text{score}} = \frac{N_{\text{real}} - N_{\text{rand}}}{SD} \quad (2)$$

An example of a comparison of real and randomized networks is given in Figure 5.

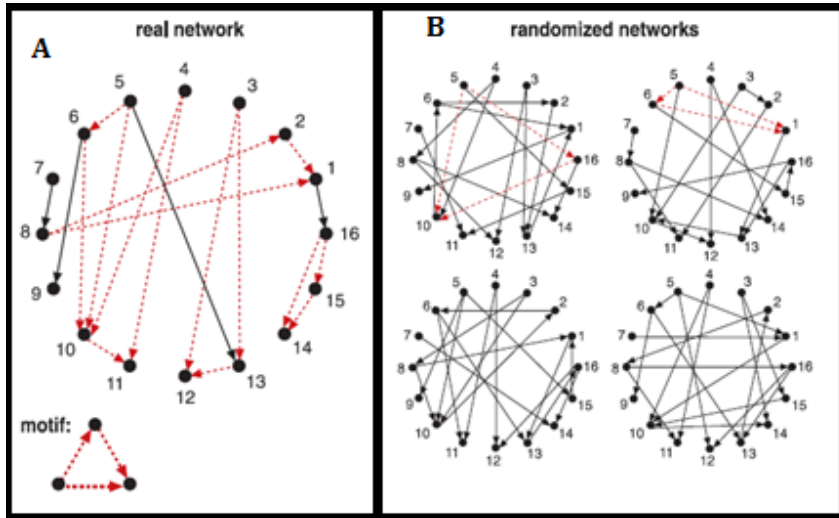


Figure 5. Comparison of motifs in the real network and randomized networks.

In the literature, there are different motif detection tools such as Mfinder [28], MAVisto [29] and FANMOD. In this paper, we examined the network motifs for migration and commuting data by using FANMOD [30]. FANMOD implements two methods to detect network motifs: full enumeration of subgraphs [26], and sampling of subgraphs to estimate subgraph concentrations [11].

## Epidemic Spread

Networks are power tool for conceptualizing the potential interaction within a population. The SIR epidemic model divides the population into three compartments: susceptible, infected and recovered population [31]. The modeling of diseases is based on the law of mass action where the rate of epidemic increase is proportional to the product of the number of susceptible and infectious individuals [32]. Deterministic SIR dynamics on a metapopulation network is shown with equations (3) and (4).

$$\dot{S}_i = -S_i \sum_j \beta_{ij} I_j \quad (3)$$

$$\dot{I}_i = S_i \sum_j \beta_{ij} I_j - \gamma I_i \quad (4)$$

The asymptotic early growth regime corresponds to (see equations (5) and (6)).

$$S_i \approx 1 + O(\mathbf{I}) \quad (5)$$

$$\dot{I}_i \approx \left( \sum_j \beta_{ij} - \gamma \delta_{ij} \right) I_j + O(\mathbf{I}^2) \quad (6)$$

So, early in the epidemic infection  $I(t)$  is proportional to  $e^{\lambda t}$  (see equation (7)), where

$$\lambda = \|\underline{\mathbf{J}}\|, \quad \underline{\mathbf{J}} = (J_{ij}), \quad J_{ij} = \beta_{ij} - \gamma \delta_{ij}. \quad (7)$$

Finally, we relate equation (8) to commuting data  $C_{ij}$  using the equation (9) which ends up with a Jacobian.

$$\beta_{ij} = \alpha(C_{ij} + C_{ji}) + \nu \left( \sum_k C_{ik} C_{kj} \right) \quad (8)$$

$$\underline{\mathbf{J}} = \alpha(\underline{\mathbf{C}} + \underline{\mathbf{C}}^T) + \nu(\underline{\mathbf{C}}\underline{\mathbf{C}}^T) - \gamma \mathbf{I} \quad (9)$$

In previous studies, it has been proposed that the epidemic threshold ( $\tau$ ) is related with the dominant eigenvalue  $\lambda_{1,A}$  of the Jacobian (see equation (10)) [33]. We will argue this statement in the Discussion section.

$$\tau = \frac{1}{\lambda_{1,A}} \quad (10)$$

## Results

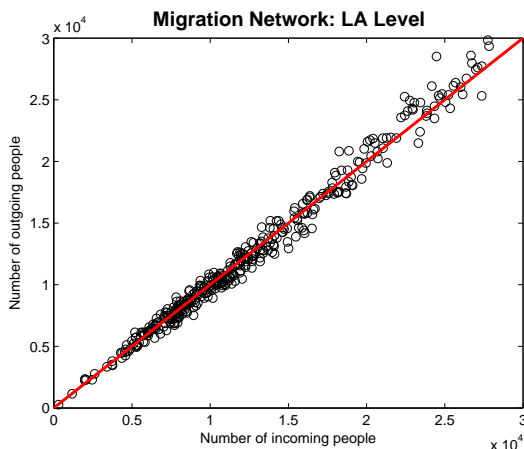
### Network Measures

The results of the network parameters at local authority level are given in Table 1. These basic parameters are offered by Newman in order to characterize the network [5]. The average degree and the density in the migration network are larger than the values in the commuting network. This shows there are more movements happening in the migration network.

Next we are interested in the degree-related properties such as degree distribution and degree correlation. We found the incoming and outgoing degree distribution for different spatial levels (see Appendix A). In movement networks, it is interesting to observe the in- and out-degree correlations. The degree correlation of migration network at local authority level is given in Figure 6. The in- and out-degree

**Table 1. Network Parameters in LA level**

Index	Migration	Commuting
Min in-degree( <i>id</i> );Min out-degree( <i>od</i> )	37;30	6;8
Max <i>id</i> ;Max <i>od</i>	422;418	370;392
Number of vertices with <i>od</i> > <i>id</i>	217	190
Number of vertices with <i>id</i> > <i>od</i>	203	233
Number of vertices with <i>od</i> = <i>id</i>	6	3
Average degree	516.81	366.64
Standard deviation degree	144.69	145.81
Density	0.60	0.43
Average clustering coefficient	9.98	25.99
Transitivity	10.78	25.00
Maximized modularity	0.73	0.83
Betweenness mean	367.89	410.79
Betweenness standard deviation	1008	471.57
Max betweenness mean	5521	4692



**Figure 6. Degree correlation of migration network at local authority level.**

are highly correlated in this network, which is what we would predict intuitively because the migration happens between home locations relatively symmetrically.

The degree correlation of commuting network at local authority level is given in Figure 7. The regions above the red line, that are shown with dots, behaves as a source whereas the regions below the red line behaves as a sink. Here, we observe a more distributed plot as we would predict before because there are different types of areas: sinks are likely to be commercial and industrial areas; sources are likely to be suburban areas.

Moreover, we continue our analysis on degree correlations at output area level (see Figure 8). However, in this case the degree correlation plots are not informative as the ones before due to the sparse, essentially biological, nature of the dataset. So, at this point we change our method to analysis of network motif structures. First we start our analysis with a measure of the simplest form of motif, i.e. at two node level, called reciprocity.



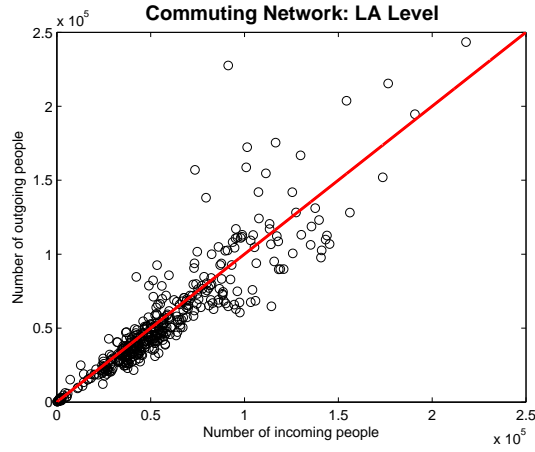


Figure 7. Degree correlation of commuting network at local authority level.

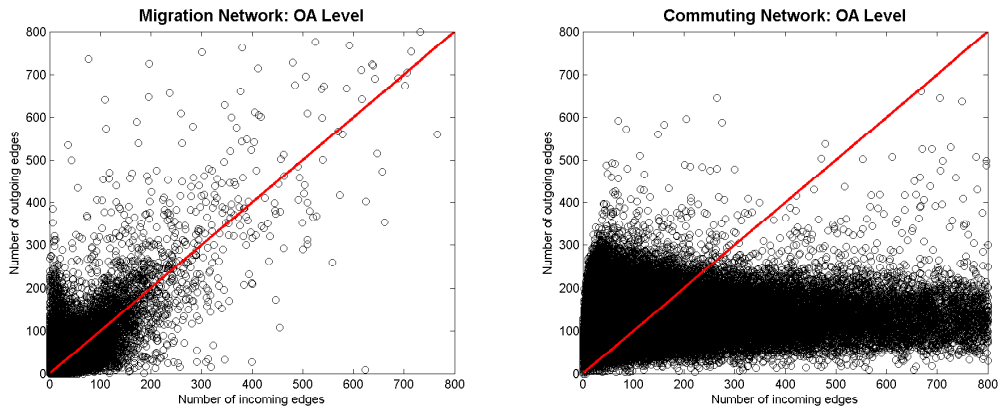


Figure 8. Degree correlation at output area level.

We calculated the reciprocity factor for migration and commuting data in both spatial scales. The results are shown in Table 2. As we can conclude, there is less reciprocity in commuting. This is mainly because the work places are located in different districts than the residential places, whereas in migration we observe more reciprocity because people tend to migrate over the residential places. Finally, in the Output Area level, we observe very small reciprocity, also they are equal to each other, this is mainly because of the matrix being very large and sparse. The reciprocity in output area level is not therefore as informative as at the local area level. At this point, question of interest might be the relation of reciprocity  $r$  with two-motif  $Z_{score}$ . Probability of having a directed edge from  $j$  to  $i$  given the directed edge from  $i$  to  $j$  in the real network will be relatively higher than one would expect to see in a null model.

Moreover, we also calculated the reciprocity in LA level for male and female differently at local authority level (see Table 3). In migration network, male and female have similar results, as people tend to move with their families. In commuting network, we observe a higher movement in male population rather than the females. This could be related with the employment percentage of male to be higher than the female.

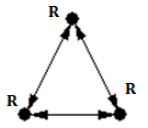

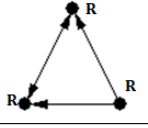
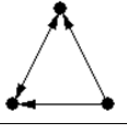
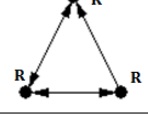
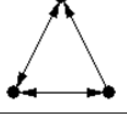
**Table 2. Results of Reciprocity: LA Level and OA Level**

	Migration	Commuting
LA Level	$r = 0.7456$	$r = 0.6151$
OA Level	$r = 0.0582$	$r = 0.0579$

**Table 3. Results of Reciprocity: Female and Male**

	Migration	Commuting
Female	$r = 0.6593$	$r = 0.5383$
Male	$r = 0.6691$	$r = 0.6141$

Then, we examined the motif structures in migration and commuting data at different scales. The results of the data at local authority level is given in Appendix B. What we are more interested is the motifs at output area level, as we have almost a topological network. The results of the migration network is given in Figure 9. The motif results are obtained with full enumeration, which is computationally expensive and time consuming. The motif with the highest  $Z_{score}$  is a clique, this implies a stronger symmetry between patches than reciprocity alone. It shows there are lots of movements in the network that are equally likely in terms of direction, at the level of triples of nodes, which we would heuristically predict from our picture.

MIGRATION OA LEVEL		Frequency (Original)	Mean- Frequency (Random)	Standard- Deviation (Random)	Z-Score
		0.00284%	2.056e-009%	6.502e-010	43784
		0.37165%	0.0056823%	1.2527e-006	2921.5
		0.01962%	4.090e-005%	9.7332e-008	2011.6

**Figure 9. Migration network motifs at output area level.**  $R$  represents the area of residential buildings, the donation added after the results have been obtained from FANMOD.

There are two possible ways to find motif structures, the first one is with full enumeration and the second is an edge sampling algorithm. In this part, we compared the results of the migration network motifs when we use different algorithms 10. FANMOD also supports several models for randomized network generation such as local constant model, global constant model and no constant model. In the local constant model, directed edges are exchanged with one another and the number of edges connected to each node remains constant. The comparison is done with the result from full enumeration with local constant model, sampling algorithm with global constant model and sampling algorithm with no constant

model. We observe the similar results in sampling algorithm, however the motif with the highest  $Z_{score}$  could not be found with the sampling algorithm.

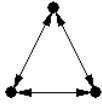
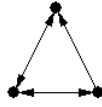
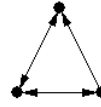

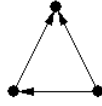
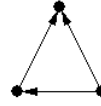
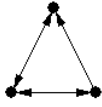
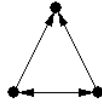
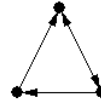
Full Enumeration; Local Constant	Z-Score	Sampling; Global Constant	Z-Score	Sampling; No Constant	Z-Score
	43784		2809		20016
	2921.5		1097.4		3550.8
	2011.6		991.55		2191.2

Figure 10. Comparison of migration network motifs generated with different algorithms.

Finally, we found the commuting network motifs. Because of the commuting network is very large we could not find the results of the whole network. However, we divided the map into four regions with similar size in order to obtain smaller size networks. We took contiguous subnetworks in order to reflect the network. The results of the three regions are given in Figure 11 which are all the same, the results are obtained with sampling algorithm.

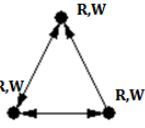

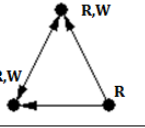

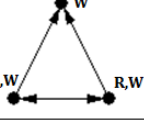

COMMUTING OA LEVEL		Frequency (Original)	Mean- Frequency (Random)	Standard- Deviation (Random)	Z-Score
		0.04134% 0.02740% 0.00935%	1.088e-007% 1.310e-007% 2.267e-007%	8.0677e-009 6.4409e-009 8.9574e-009	51242 42552 10438
		0.48294% 0.39586% 0.1398%	0.00012182% 0.00019388% 0.00029591%	3.5769e-007 6.1396e-007 1.3587e-006	13498 6444.6 1026.8
		0.064322% 0.047015% 0.015943%	1.629e-005% 1.743e-005% 1.596e-005%	8.01e-008 8.8586e-008 8.586e-008	8028.2 5305.3 2035.4

Figure 11. Commuting network motifs at output area level.  $R$  represents the areas with residential buildings and  $W$  represents the areas with workplaces, the donation added after the results have been obtained from FANMOD.

## Epidemic Spread

We found the dominant eigenvalue and epidemic threshold of the commuting network at local authority level by using the Equation (9). The results depend on different sets of  $(\alpha, \nu)$  (see Table 4).

**Table 4. The dominant eigenvalue and epidemic threshold**

$\alpha$	$\nu$	$\max(\lambda)$	$\tau$
$10^{-10}$	$10^{-10}$	7.68	0.130
$10^{-5}$	$10^{-10}$	13.57	0.073
$10^{-10}$	$10^{-9}$	85.88	0.011
$10^{-5}$	$10^{-9}$	91.77	0.010

## Conclusion

In this paper we presented a comparison of migration and commuting network measures between different hierarchy of regions by using the UK 2001 census data. For the local authority level, movement networks show sink/source behavior: In the migration network, the in- and out-degree are highly correlated because migration happens between home locations relatively symmetrically. On the other hand, in the commuting network, we observe distributed sink/source behavior over the system, which shows that in the network, there are some locations likely to be commercial and industrial which behave as a sink, and there are suburban areas which behave as a source.

In the output area level, we observe symmetry between nodes rather than sink/source behavior. In order to deepen our analysis on the output area level, we decided to examine motif structures with three nodes. We observed different network motif structures for commuting and migration networks. The motif with the highest  $Z_{\text{score}}$  is a clique in the migration network, which implies a stronger symmetry between patches than reciprocity alone,  $r = 0.74$ . However, in the commuting network we observe less symmetry between patches because the regions may represent work places, residential areas or areas containing both. Also we found a lower reciprocity factor,  $r = 0.61$ , which implies less symmetric movement in the network.

Understanding the movement behavior of commuters are essential in order to model the spread of disease. There are several reasons why scientists are interested in epidemic modeling. The first reason is to understand commuters behavior and how the parameters in the network effect the spread mechanism. The second goal is to predict the disease spread in future, and finally the most important one is to determine ways to control and stop epidemic. In previous studies in the literature, it has been shown a relationship between the epidemic threshold of a system and the dominant eigenvalue of the corresponding matrix. Looking at the spectrum might be a good starting point which is very basic, however it is not sufficient to model the spread of diseases. Givan *et al.* provide a summary of the most general way to find an exact relations and predict the epidemic threshold of a system which can be more complex in general than the previous approach, and can interpret all of the spectral properties of commuting data [18]. Epidemic is a process very sensitive to network structure, and it is hard to make concrete analysis when having networks with larger size. For a future work, the motivation from this study might be to link everything such as reciprocity, motif structure, basic network measures and spectrum of the network in order to model the disease spread, to make the model more closer to the real process.

## Acknowledgments

The project is funded by the European Commission under the Erasmus Mundus master programme.

## Appendix A: the degree distribution plots

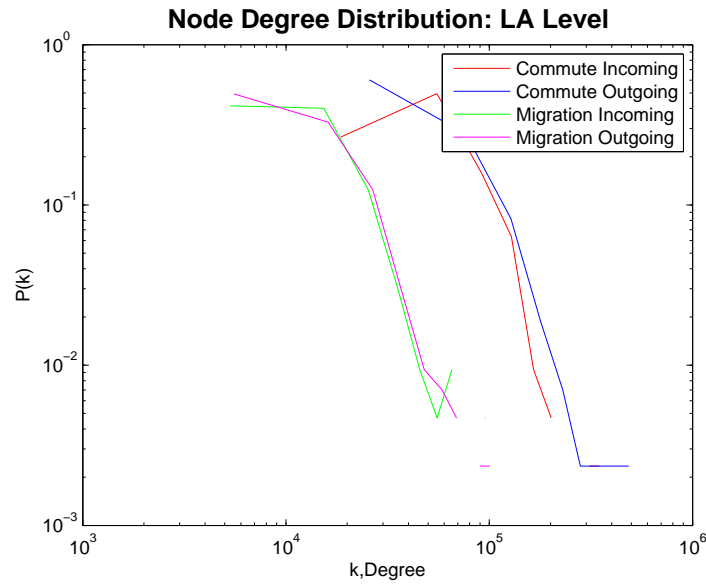


Figure 12. Degree distribution plot at local authority level.

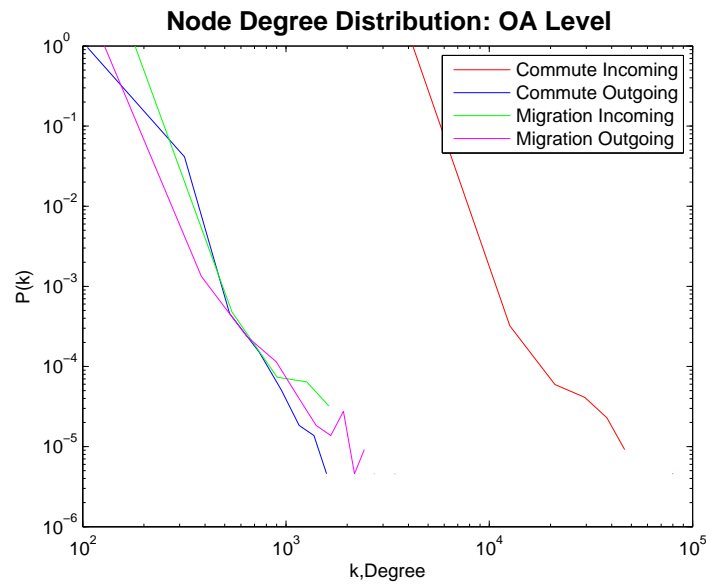


Figure 13. Degree distribution plot at output area level.

## Appendix B: the network motifs at local authority level

Network motifs at local authority level: In migration data, we have the higher  $Z$  – score for a clique, this implies a stronger argument than reciprocity. It shows there are lots of movements. However, in commuting data we don't observe a clique structure.







MIGRATION LA LEVEL	Frequency (Original)	Mean-Frequency (Random)	Standard-Deviation (Random)	Z-Score
	18.204%	13.825%	0.00729	6.000
	10.396%	7.571%	0.00481	5.862
	9.3515%	6.769%	0.00451	5.722
	0.571%	0.172%	0.0005	6.785
	1.397%	0.881%	0.0009	5.426
	1.617%	1.067%	0.0010	5.294

Figure 14. Migration network motifs with three and four nodes at local authority level.







COMMUTING LA LEVEL	Frequency (Original)	Mean-Frequency (Random)	Standard-Deviation (Random)	Z-Score
	7.0468%	6.8623%	0.00054	11.135
	6.7033%	3.9124%	0.00330	8.437
	6.4689%	3.8857%	0.00310	8.322
	1.2794%	0.3392%	0.00111	8.413
	1.5553%	0.3498%	0.00165	7.278
	0.2697%	0.1629%	0.00015	6.916

Figure 15. Commuting network motifs with three and four nodes at local authority level.

## References

- [1] D.J. Watts. *Small worlds: the dynamics of networks between order and randomness*. Princeton Univ Pr, 2003.
- [2] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *nature*, 393(6684):440–442, 1998.
- [3] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [4] R. Cohen and S. Havlin. *Complex networks: structure, robustness and function*. Cambridge Univ Pr, 2010.
- [5] M. Newman. *Networks: an introduction*. Oxford University Press, Inc., 2010.
- [6] D. Leon, F. Ashley P, H. Thomas, J. Chris P, K. Matt J, R. Gareth O, R. Joshua V, V. Matthew C, et al. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.
- [7] M.W. Horner. Spatial dimensions of urban commuting: A review of major issues and their implications for future geographic research\*. *The Professional Geographer*, 56(2):160–173, 2004.
- [8] R. Patuelli, A. Reggiani, S.P. Gorman, P. Nijkamp, and F.J. Bade. Network analysis of commuting flows: A comparative static approach to german data. *Networks and Spatial Economics*, 7(4):315–331, 2007.
- [9] A. De Montis, A. Chessa, M. Campagna, S. Caschili, and G. Deplano. Modeling commuting systems through a complex network analysis: A study of the Italian islands of Sardinia and Sicily. *Journal of Transport and Land Use*, 2(3), 2010.
- [10] G. Maier and M. Vyborny. *Internal migration between US-states*. Wirtschaftsuniv., Abteilung für Stadt-und Regionalentwicklung, 2005.
- [11] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science's STKE*, 298(5594):824, 2002.
- [12] IM Hall, JR Egan, I. Barrass, R. Gani, and S. Leach. Comparison of smallpox outbreak control strategies using a spatial metapopulation model. *Epidemiology and infection*, 135(07):1133–1144, 2007.
- [13] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124, 2004.
- [14] R. Guimera, S. Mossa, A. Turttschi, and L.A.N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7794, 2005.
- [15] DM Green, IZ Kiss, and RR Kao. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2729–2735, 2006.
- [16] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

- [17] O. Diekmann, JAP Heesterbeek, and JAJ Metz. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382, 1990.
- [18] O. Givan, N. Schwartz, A. Cygelberg, and L. Stone. Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches. *Journal of Theoretical Biology*, 2011.
- [19] European Commission et al. Nomenclature of territorial units for statistics-nuts statistical regions of europe’. URL: [http://ec.europa.eu/comm/eurostat/ramon/nuts/home\\_regions-en.html](http://ec.europa.eu/comm/eurostat/ramon/nuts/home_regions-en.html), 2005.
- [20] T. House, I. Hall, L. Danon, and M.J. Keeling. Contingency planning for a deliberate release of smallpox in Great Britain-the role of geographical scale and contact structure. *BMC infectious diseases*, 10(1):25, 2010.
- [21] Office for National Statistics. 2001 census: Sample of anonymised records (sars) (licensed) (England, Wales, Scotland and Northern Ireland). ESRC/JISC Census Programme.
- [22] E.D. Kolaczyk. *Statistical analysis of network data: methods and models*. Springer Verlag, 2009.
- [23] Office for National Statistics. 2001 census: Special migration statistics (United Kingdom). ESRC/JISC Census Programme.
- [24] Office for National Statistics. 2001 census: Special workplace statistics (England, Wales and Northern Ireland). ESRC/JISC Census Programme.
- [25] Office for National Statistics. 2001 census: Special travel statistics (Scotland). ESRC/JISC Census Programme.
- [26] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.
- [27] U. Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.
- [28] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20(11):1746–1758, 2004.
- [29] F. Schreiber and H. Schwöbbermeyer. Mavisto: a tool for the exploration of network motifs. *Bioinformatics*, 21(17):3572–3574, 2005.
- [30] S. Wernicke and F. Rasche. Fanmod: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006.
- [31] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemicsi. *Bulletin of mathematical biology*, 53(1):33–55, 1991.
- [32] JAP Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis, and interpretation*, volume 5. Wiley, 2000.
- [33] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34. IEEE, 2003.