Supplementary Information for "Entangled communities and spatial synchronization lead to criticality in urban traffic"

Giovanni Petri*
ISI Foundation, Via Alassio 11/c, 10126 Torino - Italy

Paul Expert†
Centre for Neuroimaging Sciences, Institute of Psychiatry,
De Crespigny Park, King’s College London, London SE5 8AF, UK and
Complexity and Network Group, Imperial College London, London SW7 2AZ, UK

Henrik J. Jensen‡
Complexity and Network Group, Imperial College London, London SW7 2AZ, UK and
Department of Mathematics, Imperial College London, London SW7 2AZ, UK

John W. Polak§
Centre for Transport Studies, Department of Civil and Environmental Engineering,
Imperial College London, London SW7 2PG, UK

* giovanni.petri@isi.it
† paul.expert@kcl.ac.uk
‡ h.jensen@imperial.ac.uk
§ j.polak@imperial.ac.uk
I. THE FREEFLOW PROJECT AND ITS DATASET

The Freeflow Project, a collaborative effort of 15 partners in academia, government and industry led by Imperial College London, is one of the winners in the Future Intelligent Transport Systems (FITS) initiative, promoted by the UK Department of Transport. The aim of the proposals funded by FITS generally lies in the direction of developing novel solutions to transport problems, in particular by adopting multidisciplinary approaches that are able to stimulate and accelerate the industry investment toward innovation.

In this context, Freeflow's goal is to change the way traffic data is used through the inclusion of new types of data (CCTV feeds, for example), the development of reliable network performance metrics and intelligent decision support. At the moment, Freeflow is collecting data in London, York and Kent. In this article, we focus on the data coming from London for two reasons: they are at present the most complete, and they permit to investigate traffic dynamics in an extremely heterogeneous and dense urban environment. Across the Greater London Authority area, Freeflow is recording data from 3256 sensors, mostly distributed in the Central London area but reaching also Tottenham to the north, Brixton to the south, Stamford to the east and Chiswick to the west (Figure S.1).

![Map data ©2011 Google](Image)

FIG. S.1. The map of the part of Central London covered by the Freeflow sensors data (Image is ©2011 Google). There are in excess of 3000 ILDs, stretching from Wimbledon and Brixton in the south to Tottenham Hale and Walthamstow in the north, from Stratford in the east all the way to Kensington and Chiswick in the west, for a total catchment area of about 400 km².

Each sensor is an Inductive Loop Detector (ILD) able to measure the flow and occupancy, respectively the number of vehicles passing over the detector and the fraction of time that the sensor passes covered by a vehicle. The latter can be related to the traffic density \( \mu = 8.5, \sigma = 12.8 \), and one for very large densities, centered at an occupancy of \( \sim 92 \). The first one contains about 95% of the sensors in the network, the remaining belonging to the high density peak. This means that there exist locations in the street...
FIG. S.2. Top panel, the map shows the time-averaged occupancy mean $\langle o_i \rangle$ of sensors. For clarity of visualization, the sensors’ size and color both represent the average value of the occupancy for the corresponding sensor. There is a large spatial disorder in the distribution of traffic with the most central sensors experiencing the largest densities. Bottom panel, the values are distributed around a well defined peak, which is fit well by a normal distribution characterized by a mean of 8.6 and width 12.9 and accounts for 95% of the total sensors. The remaining 5% lies in the peak for very high densities, corresponding (close to occupancy 92%) to the red, large dots in the picture.

network that are systematically undergoing large traffic densities. Interestingly, they are not tightly clustered, being on the contrary rather mixed up with other sensors. Moreover, there is no uniform distribution of average occupancies; sensors tend to be either very congested or very free.

This static picture is useful to develop an intuition regarding the traffic distribution on the street network. However it is unable to provide insights into the dynamical evolution of the network traffic states. More interesting is the distribution of sensor instantaneous occupancies (Figure S.3). We find that the distribution is characterized by an initial quasi-plateau for low occupancies and a power law tail for larger occupancies. This implies that, in analogy to other systems displaying such distributions [18–20] and although low-density states are very likely, the probability of finding a sensor in a state of large density is significant due to the presence of the thick tail. The aggregated observations for the whole day mix together states of the network that are radically different, for example rush hours...
FIG. S.3. The distribution of instantaneous occupancies for the Freeflow sensors separated for periods of the day. The single sensor occupancy distributions show an initial plateau followed by a power-law decay. The global distribution of occupancies does not show significant differences between different time of the day.

and the night. We see however that the distributions obtained separating different sections of the day are very similar, with a small bump around 20-40% in the morning and a small increase for the probability of high events in the day and evenings. Therefore, from the point of view of the distribution of network densities, the heterogeneity is not accounted by the inter-day variability of the network traffic. Differences emerge if one studies the distribution of single sensor occupancy for classes of different average network occupancy $O(t) = 1/N \sum_{i=1}^{N} o_i(t)$. In the dataset $O(t)$ varies between 0 and 30. Define then the sets $T_{low}$, $T_{med}$, $T_{high}$ to be:

$T_{low} = \{ t \in (0, T) | O(t) < 10 \}$

$T_{med} = \{ t \in (0, T) | 10 < O(t) < 20 \}$

$T_{high} = \{ t \in (0, T) | O(t) > 20 \}$

In Figure S.4 we plot the distribution of occupancies obtained restricting to the three different sets of times. In this new cut of the data, the occupancy distributions are different. For low total occupancy, the distribution shows a nice power law behaviour, $P(o_i) \sim o_i^{-\omega}$ with $\omega = 2.4 \pm 0.1$. When the density increase, as expected the distribution shifts to the right, enlarging the plateau, but the best fit to the tail has the same exponent as before (within errors), implying that, although somewhat reduced, the heterogeneity of traffic distribution carries through the whole total network occupancy range.

This implies also that the network traffic density field is very disordered. This observation is not unexpected: travelers tend to concentrate on few major roads, reducing the density of traffic on the smaller roads. As first argued in [21], this disorder is likely to create instabilities, especially during periods of intense traffic, reducing the carrying capacity of the network and effectively decreasing its performances.

II. SENSOR PROFILES AND CORRELATION MATRICES

In order to make sense of the information contained in the second layer of the Partition Decoupling Method (PDM), we want to make use of the explicit distance information about the sensors through the spatial modularity approach. However, while the PDM works directly by clustering sensors through similarities between the timeseries, the spatial
modularity requires the underlying network to be provided. Moreover, the second layer of the PDM refers to the correlations between sensors’ timeseries after the effects of the first layer are removed. As explained in the main text, this entails constructing a characteristic timeseries for each community found at a given partition layer and using it to "scrub" [2] (detrend) the original data. In the light of these consideration, we produce a typical daily profile for each sensors and remove it from the data.

For sensor $i$, denote by $f_i(t)$ and $o_i(t)$ its flow and occupancy timeseries and build its typical daily flow profile $\Gamma_i^f(\tau)$ and occupancy profile $\Gamma_i^o(\tau)$:

$$\Gamma_i^f(\tau) = \frac{1}{n_{days}} \sum_{m=0}^{n_{days}} f_i(\tau + mT_{day})$$

$$\Gamma_i^o(\tau) = \frac{1}{n_{days}} \sum_{m=0}^{n_{days}} o_i(\tau + mT_{day})$$

where $\tau$ is the time of the day ($\tau \in (0, T_{day})$), $n_{days}$ is the number of days the timeseries lasts and finally $T_{day}$ is the length in time steps of a day. The analyzed timeseries span a month and have a temporal resolution of 15 minutes, so $n_{days} = 30$ and $T_{days} = 24 \times 60 / 15 = 96$. We then build the detrended timeseries $\hat{f}_i(t)$ and $\hat{o}_i(t)$ by subtracting the profile function:

$$\hat{f}_i(t) = f_i(t) - \Pi_i^f (mod(t,T_{days}));$$

$$\hat{o}_i(t) = o_i(t) - \Pi_i^o (mod(t,T_{days}));$$

where $mod$ is the modulo function. Figure S.5 reports an example of the resulting profiles (sensor n00/005g1). The correlation matrices are then built from the detrended timeseries in the standard way, by interpreting pair-wise Pearson correlation coefficients as entries of the adjacency matrix of a fully connected, weighted graph:

$$\Omega_{ij}^f = \frac{\sum_i (\hat{f}_i(t) - \mu_{\hat{f}_i})(\hat{f}_j(t) - \mu_{\hat{f}_j})}{\sqrt{\sum_i (\hat{f}_i(t) - \mu_{\hat{f}_i})^2} \sqrt{\sum_j (\hat{f}_j(t) - \mu_{\hat{f}_j})^2}}$$

$$\Omega_{ij}^o = \frac{\sum_i (\hat{o}_i(t) - \mu_{\hat{o}_i})(\hat{o}_j(t) - \mu_{\hat{o}_j})}{\sqrt{\sum_i (\hat{o}_i(t) - \mu_{\hat{o}_i})^2} \sqrt{\sum_j (\hat{o}_j(t) - \mu_{\hat{o}_j})^2}}$$
FIG. S.5. An example of the flow and occupancy profile (sensor no. n00/005g1). The errors are the standard deviation of the flow/occupancy values at given time.

III. METHODS AND RESULTS

A. Partition Decoupling Method

The Partition Decoupling Method (PDM) is a topological clustering technique proposed by Leibon et al. [2] and successfully applied to the analysis of a section of the equities market [2], of gene expressions [3] and the roll call data of the US Congress [4]. Through recursive clustering and "scrubbing" steps, the method allows to progressively uncover finer features in the data and finally yields a multi-layered community structure, where each layer encodes qualitatively different information in contrast to standard hierarchical clustering methods.

The algorithm needs two ingredient: a vector of data (e.g. coordinates in a high dimensional space or a time series) for each node, which, borrowing the notation of [2], we denote as $D(i)$ for node $i$; and a node similarity measure derived from the data, for example Euclidean distance, Pearson correlation or mutual information. With these two elements, the algorithms works as follows:

1. Using a clustering method on the similarity measure, a partition $P = \{C_0, \ldots, C_m\}$ is produced, where $\alpha$ specifies the layer of the decomposition ($\alpha = 0$ for the first layer) and $m$ the number of communities in the partition. For the first step, $\alpha = 0$.

2. For each community $l = 1, \ldots, m$ in partition $P^\alpha$, a characteristic vector $V^\alpha_l$ is constructed by averaging over the vectors associated to the nodes belonging to $l$, that is $V^\alpha_l = \text{mean}\{D^\alpha(i) | i \in C^\alpha_l\}$.

3. The set $\{V^\alpha_1, \ldots, m\}$ is then used to produce the "scrubbed" data vectors of the following layer $D^{\alpha+1}$, by expressing the vectors $D^\alpha$ in the form $D^\alpha = F^\alpha + R^\alpha$, where the first term represents the projection on the $\{V^\alpha_1, \ldots, m\}$ set and $R^\alpha$ the residual. The vectors $\{D^{\alpha+1}\}$ are then obtained by normalizing the residuals, $D^{\alpha+1}(i) = \text{norm}(D^\alpha(i) - F^\alpha(i))$.

4. The steps above are iterated until the residuals cannot be discerned anymore from the corresponding Gaussian Ensemble null model [4].

The output is a list of layers, each defined by a partition $P^\alpha = \{C_0^\alpha, \ldots, C_m^\alpha\}$, a set of characteristic vectors $\{V^\alpha\}$ and the projection, $\pi_{k=1, m}^\alpha(i)$, of the $(\alpha - 1)$ data vectors on the communities in partition $\alpha$. Each layer describes the finer structure of the data after having accounted (scrubbed) the effects of the previous ones.
B. Spatial modularity

The optimisation of Newman-Girvan (NG) modularity [5] is the prominent tool for community or cluster detection in complex networks:

\[
Q = \frac{1}{2m} \sum_{C \in \mathcal{P}} \sum_{i,j \in C} \left| A_{ij} - P_{ij} \right|, \quad P_{ij} = \frac{s_i s_j}{2m},
\]

where \( i, j \in C \) is a summation over pairs of nodes \( i \) and \( j \) belonging to the same community \( C \) of \( \mathcal{P} \) and therefore counts links between nodes within the same community. \( s_i \) is the strength of a node and \( 2m \) the total strength of the network. Optimising modularity thus groups together nodes that have more in common than what is expected by the null-model \( P_{ij} \). Its power, but also its weakness, comes from considering solely the pair-wise interactions, in other words the elements of the adjacency matrix of the network and compares the actual topology against an expected random topology constrained with the degree distribution of the original network. There are cases in which this approach overlooks important features of the network as, for example, when nodes and links have properties that are not encoded in the topology, such as distances or fitness. The emblematic case is provided by spatial networks, where nodes and edges are embedded in space, putting severe constraints on many network properties, for example the reduction of hubs due to the cost of long links [6–12]. It is possible to account for the effects of space by carefully modifying the null model of the standard NG. An example that is very relevant to spatial networks are the gravity models (GM) [13, 14]. The idea in such models is that the flow or interaction between two nodes \( i \) and \( j \) can be effectively modelled by a distance-dependent term and by a second term that is proportional to some node-related quantity \( N_i, N_j \): in the case of passenger flow between cities, one reasonable quantity is the cities’ populations; for money flows between banks, the total bank social capital. So, in general the gravity model is written as:

\[
P_{ij}^{\text{Spa}} = N_i N_j f(d_{ij})
\]

where \( f \) can be any (generally decreasing) function of the distance \( d_{ij} \) between \( i \) and \( j \) and the \( \omega_{ij} \) play a role similar to the mass. The function \( f \) is defined as:

\[
f(d_{ij}) = \frac{\sum_{i,j|d_{ij}=d} \Omega_{ij}}{\sum_{i,j|d_{ij}=d} N_i N_j}
\]

and usually referred to as deterrence function. Here \( \Omega_{ij} \) is the correlation between two sensor’s time series for a given time delay. In this approach, the deterrence function is determined directly from the data and so does not require a fit to some functional form [14], thus making the method independent from external hypothesis. The method presented in [1] was used to optimize the spatial modularity, thus overcoming the limitations of community detection based only on network topological properties.

In the present work, the data do not provide flows of vehicles between sensors, but occupancy time-series from which we calculated correlations matrices. Using correlation automatically renormalise the relative importance of the nodes, so the mass term in the null-model can be set to 1 for all nodes. The null-model then reduces to the deterrence function which gives the average correlation between sensors as a function of distance:

\[
P_{ij}^{\text{Spa}} = f(d_{ij}) = \sum_{i,j|d_{ij}=d} \Omega_{ij},
\]

C. Results

We calculated the correlation matrices for both flows and occupancies and applied the community detection methods described in the main text. In our case, we assume the same mass \( (N_i = 1 \forall i) \) for all sensors, because the relative strength of nodes is already taken into account by the Pearson coefficient (see Eqs. (8-9)). Following [22], we test the distance function we use the expected correlation values \( C^{o,f} \) at a certain distance \( d \):

\[
f^{o}(d) = \sum_{i,j|d_{ij}=d} \Omega^{o}_{ij} = \frac{1}{n_{d_{ij}=d}} \sum_{i,j|d_{ij}=d} \Omega^{o}_{ij} = P_{ij}^{\text{Spa},o},
\]

\[
f^{f}(d) = \sum_{i,j|d_{ij}=d} \Omega^{f}_{ij} = \frac{1}{n_{d_{ij}=d}} \sum_{i,j|d_{ij}=d} \Omega^{f}_{ij} = P_{ij}^{\text{Spa},f}.
\]
FIG. S.6. Results for the Newman-Girvan community detection on the correlation network $\Omega^f$. The Newman-Girvan modularity is optimized by a partition of the network containing 4 communities. The two largest ones (red and black) account for about 85% of the sensors and correspond respectively to the northern central area and the southern peripheral areas of London. The third (yellow) appears to be spatially scattered across the central area of London, but has no evident functional structure. The fourth is almost invisible in the network. All communities show a large degree of spatial overlap and it is not easy to identify units allowing to decompose the traffic in different dynamical units. The right figure shows the size of the different communities in decreasing size order.

FIG. S.7. Results for the map-equation community detection on the correlation network $\Omega^f$. The optimal compression is obtained by by a partition of the network containing 5 communities. The largest one (black) accounts for about 95% of the sensors and covers the whole city. The second largest (red) is itself scattered across the central area of London and has no evident functional structure. The right figure shows the size of the different communities in decreasing size order.

Figure S.6 shows the results of the Newman-Girvan modularity maximization on $C^f$. The top panel contains a map of the sensors colored according to the community they belong to. The main feature are the two largest communities, which together account for $\sim 85\%$ of all the sensors. The red clearly identifies the central areas of London, while the black dots identify the peripheral sensors. The results are not unexpected, since the central area is where most of the general traffic is concentrated during the day, while the peripheral arteries sustain mostly the commute traffic in the morning and the evening. On the whole, this suggests that the whole traffic network is split in two large dynamical areas (we are correlating the flows here). At a closer look, communities show a large spatial overlap. In particular, such overlap appears not only near the spatial boundaries between the two, but also deep in the bulk. An overlap near the boundaries could be explained by small errors in identifying the maximum modularity partition...
FIG. S.8. Results for the spatial modularity community detection on the correlation network $\Omega^f$. $Q_{Spa}$ is optimized by a partition containing 78 communities. The largest community covers the whole catchment area and is followed by a series of communities with power law distributed sizes that do not provide much functional information.

(e.g. due to degeneracy near the maximum or resolution issues [23, 24]). Here instead the overlap between the two largest communities is significant (see for example the right part of the top panel in Fig. S.6) and the same applies to the other two communities detected (yellow and orange). This result cannot then be considered as an artifact of the method used and implies that the communities effectively overlap in space. However, it is hard to understand why a sensor deep in the (spatial) bulk of a community might belong to another spatially distant community while another sensor, close to the first, is not.

One way to think about this is to hypothesize that a hierarchical structure of communities exists. The Newman-Girvan modularity would not be able to resolve hierarchical structure since it looks only for the partition with the highest modularity. If some hierarchy is really present, then one should coarse-grain the network and obtain a highest modularity partition at each level of coarse-graining. This process is laborious and requires defining a process of box-counting, which on networks is an open problem in itself [25–27]. The multilevel map equation approach however is naturally tailored to identify the underlying hierarchical structure through the use of nested codebooks. In Figure S.7 we show the results of the map equation on the network. Surprisingly, we find that, beyond the trivial cases -all nodes in one community or each node in a different community- there is only another level of the community hierarchy, which contains one very large community, spanning the whole system and containing almost all the sensors. The remaining four communities detected are extremely small and spatially scattered. This means that the trajectory of a random walker on $C^f$ is very hard to compress and does not show modules with high persistence times, i.e. dense network structures. In addition, there is no hierarchy in the network partitions.

A very similar result is obtained from the optimization of spatial modularity. Also in this case, we found a large community that spans the whole system (about 65% of the sensors). In contrast to the map equation results, there are 77 smaller partitions. Incidentally, if communities are ordered by descending size and their position in this ranking denoted by $r$ (Figure S.8, bottom), one finds that the community size $s$ scales approximately like $s \propto r^{-\nu}$ with $\nu = 1.09 \pm 0.01$ (adj-$R^2 = 0.97$). The top panel of Figure S.8 shows how the communities are distributed in space. The red dots covering the whole network correspond to the largest community, while the other colors correspond to the smaller communities. This visualization is confusing however and does not allow to identify the details of the communities. While some of the smaller communities appear to identify recognizable elements of the street network, others instead are completely spatially spread out. For an example, the panels of Figure S.9 show some of the largest communities (large dots) against the background of the largest one (small dots). The second largest does not show any recognizable feature and is spread all over the catchment area (top left). The third largest (top right) instead appears to identify two different branches of an important corridor in London, the north-south path that leads from Vauxhall bridge along the Marble Arch area toward Edgware Road. However, in general the identified communities are very noisy, making it extremely hard to understand the reliability of the partition and the role of the sensors composing them.
D. Significance of the Spatial modularity partition

As argued in the main text, the structure found in the clustering by the Spatial modularity stems from spatio-temporal synchronisation. To confirm that the structure unveiled by the Spatial modularity is genuinely caused by spatio-temporal synchronisation as argued below, we performed $N = 100$ randomisation of the sensor’s position, while keeping their time-series the same. And optimised modularity for the corresponding null-models. A $z$-score can be computed as follows:

$$z = \frac{Q_{\text{original}} - \mu_{Q_{\text{rand}}}}{\sigma_{Q_{\text{rand}}}},$$

where $\mu_{Q_{\text{rand}}}$ is the average value of the modularity for the randomised null-models and $\sigma_{Q_{\text{rand}}}$ is the standard deviation. The $z$-score obtained is 35, confirming that the spatial layout has a genuine effect. This easily seen in fig. S.10, which shows that the effects of randomising the position typically gives a flat profile to the deterrence function, destroying any effect of space. The temporal importance is seen in the correlation matrix being different from zero. Randomisation of the time-series trivially gives zero correlation between each pair of sensors.

IV. CORRELATION FUNCTION AND LARGEST CONNECTED COMPONENT

In the spatial modularity, a modification of the gravity model, the fundamental element is the deterrence function $f$, defined in Eq. (2) of the SI. In general, $f$ measures the expected strength of a link between nodes at a given
distance. In our case, the links are (equal time) correlations and therefore the deterrence function is the two-point

correlation function at zero delay ($\tau = 0$). The two regimes we found, separated by a characteristic distance $r_0 \sim 200$ m,

correspond to the the ultra-local, single road scale versus the network level. We show this result using the following

spatial model: place each sensor on the plane at its coordinates and consider it as a node; then link all the pairs of

nodes $(i, j)$ such that their spatial distance $d_{ij}$ is smaller than a certain distance $d_0$:

$$p(d_{ij}, d_0) = \begin{cases} 
1 & \text{if } d < d_0, \\
0 & \text{if } d > d_0.
\end{cases}$$ (17)

It is clear that varying $d_0$ between 0 and $\max(d_{ij})$, one passes from the empty network to the completely connected

graph. In Figure S.11, we show the visualization of the resulting network for four values of $d_0$. In the middle, there

must exist a value of $d_0$ for which most of the network becomes connected. In graph theory, this phenomenon is

usually referred to as the birth of the giant connected component (GCC) [28, 29]. Strictly, the GCC is defined as

the component whose size is a finite fraction of the total network in the limit of network size tending to infinite

($N \rightarrow \infty$). For example, in the Erdős-Renyi graph model, one can prove that the appearance of the GCC at a

certain linking threshold probability $p_c = 1/N$ occurs as a second-order phase transition, akin to percolation [30].

However, we are interested in a finite, albeit large, system and, although these concepts are rigorously defined only

in the thermodynamic limit, we can still exploit the concept of GCC in a finite size network to test the hypothesis

regarding the two regimes. If at $r_0$ the sensor network starts to develop a large component, it means that for $r < r_0$ we

are probing only local dynamics, pertaining to small disconnected components (e.g. a street and its related junctions),

while for $r > r_0$ we are studying the integrated dynamics taking place at the network level. In Figure S.12 we show

that this is the case: the GCC undergoes a rapid growth at a value of the linking distance $d_0 \sim 100$ m very close to

the value at which the correlation function becomes algebraic. Therefore, the algebraically decaying correlations are

a feature of the network traffic, that is lost when we consider single road segment or similarly sized areas of the city.

FIG. S.10. Typical deterrence function after randomisation of the sensors’ positions.
FIG. S.11. Some examples of the spatially embedded networks constructed varying the the linking probability $p(d_{ij}, d_0)$ (eq. (17)), for the cases $d_0 = 50, 100, 300, 400 m$. 
FIG. S.12. The top figure shows the size of the giant connected component of the spatial network defined by eq. (17) as a function of the linking distance $d_0$. The bottom figure presents $C(r, \tau = 0)$ as a function of the distance. Note that the size giant connected component starts to increase steeply at the same value of distance for which the correlation function start displaying a power law dependence.
[1] Richardson T, Mucha PJ, Porter MA Bettencourt LMA, Lobo J, Helbing D, Kuhnert C, West GB (2009) Spectral tripartitioning of networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 80:036111.

[2] G. Leibon, S. Pauls, D. Rockmore, and R. Savell. Topological structures in the equities market network. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52):20589–20594, 2008.

[3] R. Braun, G. Leibon, S. Pauls, and D. Rockmore. Partition decoupling for multi-gene analysis of gene expression profiling data. *BMC bioinformatics*, 12(1):497, 2011.

[4] G. Leibon, S. Pauls, D.N. Rockmore, and R. Savell. Partition Decomposition for Roll Call Data. *Arxiv preprint arXiv:1108.2805*, 2011.

[5] MEJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8577–8582, 2006.

[6] L.A.N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97(21):11149–11152, 2000.

[7] R. Guimera. Modeling the world-wide airport network. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38:381–385, 2004.

[8] Barrat A, BarthÃ©lemy M., and Vespignani A. The effects of spatial constraints on the evolution of weighted complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05003, 2005.

[9] Y. He, Z. J. Chen, and A. C. Evans. Small-world anatomical networks in the human brain revealed by cortical thickness from mri. *Cerebral Cortex*, 17(10):2407–2419, 2007.

[10] R. Boschma and K. Frenken. The spatial evolution of innovation networks: A proximity perspective. Technical Report 0905, 2009.

[11] M. Barthelemy and A. Barrat. The role of geography and traffic in the structure of complex networks. *Advances in Complex Systems*, 10(1):5–28, October 2007.

[12] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. volume 102, pages 11623–11628, 2005.

[13] A. G. Wilson. A statistical theory of spatial distribution models. *Transportation Research*, 1:253–269, 1967.

[14] S. Erlander and N. F Stewart. *The gravity model in transportation analysis: theory and extensions*. Vsp, 1990.

[15] D. L. Gerlough and M. J. Huber. *Traffic Flow Theory*. Transportation Research Board, 1976.

[16] H. Greenberg. An analysis of traffic flow. 1959.

[17] J.H Banks. Freeway speed-flow-concentration relationships: More evidence and interpretations. *Transportation Research Record*, 1225:53–60, 1989.

[18] F. Caruso, A. Pluchino, V. Latora, S. Vinciguerra, and A. Rapisarda. Analysis of self-organized criticality in the Olami-Feder-Christensen model and in real earthquakes. *Physical Review E*, 75(5), May 2007.

[19] Adilson Motter and Ying-Cheng Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66(6), December 2002.

[20] X. Gabaix, P Gopikrishnan, V. Plerou, and H E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(9297):267–270, May 2003.

[21] A. Mazloumian, N. Geroliminis, and D. Helbing. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4627–4647, September 2010.

[22] P. Expert, T. S. Evans, V.D. Blondel, and R. Lambiotte. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, 108(19):7663, 2011.

[23] S. Fortunato and M. BarthÃ©lemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007.

[24] M. Sales-Pardo, R. Guimera, A. A. Moreira, and L. A. N. Amaral. Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 104(39):15224–15229, 2007.

[25] F. Radicchi, J. Ramasco, A. Barrat, and S. Fortunato. Complex Networks Renormalization: Flows and Fixed Points. *Physical Review Letters*, 101(14), 2008.

[26] S Dorogovtsev. Renormalization group for evolving networks. *Physical Review E*, 67(4), April 2003.

[27] S. Havlin. *Complex networks: structure, robustness and function*. Cambridge University Press, 2010.

[28] R. Albert and A.L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

[29] M. E. J. Newman. *Networks: an Introduction*. Oxford University Press, 2010.

[30] D. ben Avraham and S. Havlin. *Diffusion and Reactions in Fractals and Disordered Systems*. (Cambridge Nonlinear Science Series). Cambridge University Press, November 2000.