Galvan, G., & Agarwal, J. (2018). Community detection in action: Identification of critical elements in infrastructure networks. *Journal of Infrastructure Systems*, *24*(1), [04017046]. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000400

COMMUNITY DETECTION IN ACTION: IDENTIFICATION OF CRITICAL ELEMENTS IN INFRASTRUCTURE NETWORKS

Giulio Galvan[1], Jitendra Agarwal[2*]

1: Department of Civil Engineering, University of Bristol, University Walk, Bristol, BS81TR, UK

2: Department of Civil Engineering, University of Bristol, University Walk, Bristol, BS81TR, UK

*: Corresponding Author: Jitendra Agarwal, J.Agarwal@bristol.ac.uk

ABSTRACT

Modern infrastructure systems form complex networks that are organised hierarchically in communities of tightly integrated elements. This paper presents three new community-based metrics to identify the critical elements of a network system. Two of these metrics assess intracommunity and intercommunity behaviour for any community structure and the third metric accounts for the multiple levels of community structure. First, these metrics are studied to establish their characteristics with different community structures and then the Great Britain Railway Network is used as a case study to demonstrate the usefulness of these new metrics. The results show that an assessment of the system using these metrics leads to the identification of not only those elements that are critical at the global level, but also that greatly affect the local performance of the communities. Such identification of the critical components at the community level and global level would enable a better understanding of the system behaviour by the stakeholders with competing demands.

Keywords: Infrastructure networks; community structure; critical assets; network metrics; decision-support tools.

INTRODUCTION

Infrastructure systems such as electricity, water and transportation form the backbone of modern societies (UNISDR 2015). These are organised in networks of assets which facilitate the flow of resources and the delivery of services (Zio 2016). To study the emerging behaviour of infrastructure systems, the graph-theoretic techniques, originally developed in the field of statistical mechanics (Barabasi & Albert 2002), are being used increasingly. In a graph model of an infrastructure network, the nodes represent the components where the service is generated or delivered and the edges represent the distribution channels (Table 1). The size of these networks varies considerably, for example, from potable water distribution networks which can be organised in small local networks (Dueñas-Osorio et al. 2007) with less than one hundred components, to the power grids with tens of thousands of nodes (Albert et al. 2004). A review of infrastructure systems analysis is given in (Ouyang 2014).

Whilst large-scale analyses of infrastructure systems are important to assess systemic risks (Lorenz et al. 2009), these can mask local criticalities. Evaluating the behaviour of the system at different levels of granularities allows for local information to emerge. This is particularly important for infrastructure systems that have multiple stakeholders each concerned with the performance at a different scale as shown in Figure 1. These scales may be hierarchical but this is not necessarily true for all of the infrastructures; some infrastructures go beyond regional or national boundaries to serve multiple regions across different nations (as is the case of the European Gas Network (Poljansek et al. 2012)). Further, they may be of interest to consumer (or industry) associations that operate at any of the intermediate scales shown in Figure 1.

As an example, the storm surge in February 2014 in the UK that washed away 80 meters of railway line in Dawlish (Network Rail 2014), could be considered a minor event from a whole-system perspective because the station represents only 0.02% of the annual network throughput (Office of Rail and Road 2015). From a local perspective, however, the damage was unacceptable, as it completely cut off the region from the rest of the national railway network. It is not unusual for each stakeholder to be concerned with the functionality of the system at a particular scale. Planning in a multi-stakeholder environment requires the right tools so that the performance of each of the parts can be assessed with clarity at the onset of the process (Blockley & Godfrey 2000) and stakeholders are able to have a meaningful discussion about the system criticalities.

The objective of this paper is to provide tools to foster such a discussion among stakeholders with a diverse range of priorities. To achieve this, new metrics to identify the critical elements of an infrastructure system at multiple scales are proposed. The paper is organised as follows: first, it reviews the approaches to identify communities in a infrastructure network; Then, it defines two new metrics to assess the performance of communities and explores their behaviour for different network parameters; in the following section, it introduces a third new metric to account for multiple levels of community structure within the system; a Case Study is then

presented to demonstrate the application of the new metrics to a model of the Great Britain railway system before conclusions are drawn in the final section.

LITERATURE REVIEW

COMMUNITY DETECTION

In an effort to better map the behaviour of the system to the needs of its stakeholders, the concept of network community is used. Network communities can be defined as "locally dense connected subgraphs in a network" (Barabási 2014). In an infrastructure system, community structure emerges because it is required to achieve a thorough distribution of service at the local level, while guaranteeing whole-system connectivity through efficient, but scarce, long-range communication channels. The elements of an infrastructure system community are much more likely to interact with each other than with the rest of the system when delivering their service to society. Electric power, for example, is produced and distributed at the local level whenever possible, with long-range transmission (i.e. at a national scale or above) being the costly alternative to be used when local generation capacity cannot satisfy peaks in demand (Mureddu et al. 2015).

Community detection is the process of identifying any partitions that may exist in a network, and a partition is a subdivision of a network in communities. Community detection on networks is a fertile research field and there exist several alternative procedures that can be used for this purpose (Fortunato 2010). The Stability Optimisation (SO) method described in (Delvenne et al. 2010) and implemented in (Le Martelot & Hankin 2011) was selected here (see Appendix A).

SO leverages the association between graphs and random walk processes. Any graph can be associated to a random walk process where the transition probabilities are proportional to the edge weights. The groups of nodes that a random walker is unlikely to leave, because of the number and the quality of the connections between them, become basins of attraction and are referred to as communities. Partitions are identified by maximising stability, which is the likelihood that a random walk of length t, referred to as scale parameter, terminates in the same community where it started (see (Lambiotte 2010) for the full derivation). When the length of the random walk is set to 1, SO corresponds to the widely used Modularity Optimisation methods (Newman & Girvan 2004) and by changing the length t of the random walk, it allows the discovery of partitions in which the size of communities is above or below the resolution limit of modularity (Fortunato & Barthelemy 2007). Each value of t can be associated to a different partition. To discern partitions between those actually characterising the system and those that are only a product of the algorithm sweeping through t values, it is necessary to assess their robustness against small perturbations such as small increments in the length of the random walk t in the algorithmic procedure (this is different from the robustness of operations within a network or community). Partitions identified for an adequate range of values of t are considered robust and used in the analysis. There are other methods sharing the features of SO, such as (Danon et al. 2006) and (Reichardt & Bornholdt 2006),

but they will not be discussed here. The choice of the community detection method does not affect the development of the metrics presented here, as long as it is able to deliver a sequence of partitions.

TRANSPORTATION SYSTEMS AND NETWORK SCIENCE

Network science methodologies, such as Community Detection, are powerful tools to examine the system-level performance of infrastructure networks (e.g. their robustness) or of their constituent elements (e.g. the criticality of individual nodes or edges). An infrastructure domain where network science has had significant success is that of transportation networks.

For example, it has been shown that in road networks the betweenness centrality distribution of their nodes shows heavy tails (Lämmer et al. 2006). The betweenness centrality of a node is the fraction of shortest paths going through it (Borgatti 2005). As the cost of going from origin to destination is proportional to the distance travelled (in absence of congestion), shortest paths represent the preferred choice for the users of the network. As such, betweenness centrality has been used as an indicator of the flow of service going through the nodes of an infrastructure network. This is also supported by the correlation of betweenness centrality and economic activity identified by (Strano et al. 2007), as economic activity in urban areas is known to attract vehicle traffic. The fact that the betweenness centrality distribution is heavy-tailed highlights the disproportionate amount of traffic that some parts of the road network attract from the peripheral regions. Vulnerability analyses have also been performed using network science as the overarching framework: for example investigating the robustness of an Italian road network with respect to the loss of its road links (Zio et al. 2011) or assessing the systemic impact of multiple hazards on a transportation network in New Zealand (Dalziell & Nicholson 2001).

Topological studies of railway networks investigated their structure both at the national (Dunn et al. 2016; Sen et al. 2003; Kurant & Thiran 2006) and continental (Kurant & Thiran 2006; Hong et al. 2015) scale, showing that their degree distributions are also compact, with little evidence of the heavy tails characterising scale-free networks (Barabási 2009). The vulnerability of these networks has been investigated (Kurant et al. 2007) using a dual network representation, building the model from the layout of the physical infrastructure network and the distribution of the train trips. By progressively removing larger fractions of edges from the network and investigating the consequences, it was found that both representations of the network provide similar results under random removal of nodes but if the attacks are targeted towards the most heavily loaded edges, the network built on the distribution of the train trips degraded much faster. This duality was not considered in subsequent research when the vulnerability of the Chinese railway network (Ouyang et al. 2014) was assessed by disconnecting randomly selected nodes and links, or when the Swedish railway network and its ancillary infrastructure systems were assessed against the removal of an increasingly large fraction of randomly selected nodes, for the removal of specific high criticality elements, and in the scenario of spatially-correlated set of contemporary failures (Johansson & Hassel 2010).

The structure of the air transportation system as a complex network was examined at the national (Cai et al. 2012; DeLaurentis et al. 2008), continental (Cardillo et al. 2013) and worldwide (Mossa et al. 2005) scales. While

at the national level, these networks show compact degree distributions, when the models are extended to the continental scale and beyond, the distribution becomes a power law, highlighting the role of hub airports in long distance connectivity. Further, as links do not represent physical entities but flight routes, two different representations can be used, as in the case of railway networks: models can be built either of the networks connecting the airports, or of the networks operated by single airlines (DeLaurentis et al. 2008).

The vulnerability of these networks to disruptions has become of interest for the research community in the aftermath of the 2011 eruption of the Eyjafjallajökull volcano, which caused the inoperability of the northern part of the European Air Transportation network for six days, affecting 10 around million travellers (Bye 2011). Wilkinson *et al.* (Wilkinson et al. 2011) tackled directly the vulnerability of the European Air Transportation Network and found that it has a scale-free degree distribution with an exponential cut-off, with most of its hub nodes located in the northern part of the system. This makes the network vulnerable to spatially concentrated hazards, such as aforementioned eruption, as a relatively spatially concentrated disruptive event has the potential to affect airports outside of it, effectively having a much larger footprint.

In general, there are two distinct traditions for vulnerability of transport networks - one, based on topological studies and the other based on demand and supply. A review of both these approaches is provided in (Mattsson and Jenelius 2015) where they also highlight the importance of collaborations between authorities, operators and researchers to improve the resilience of transport networks. The role of connectivity has also been reviewed in (Reggiani et al. 2015). In this paper, we provide new perspectives on the identification of critical elements using the concept of communities while addressing the needs of a variety of stakeholders.

COMMUNITY DETECTION AND INFRASTRUCTURE NETWORKS

Community detection has mostly been used in the infrastructure systems research literature to reduce the computational complexity of the problem under scrutiny. In (Mena et al. 2014) an affinity-based community detection procedure was used to arrive at the most efficient configuration for an electric power system. In (Gómez et al. 2014) a version of Markov Chain clustering was used to provide coarser, yet faster, assessment of damage propagation through a highway system. A different point of view was taken in (Rocco & Ramirez-Marquez 2011) and (Fang & Zio 2013) where modularity optimisation and hierarchical spectral partitioning were used respectively for the common goal of identifying system criticalities based on the position within the clusters of system elements. In these works (and those reported in Table 1), however, critical nodes are identified from the perspective of the whole system; the consequences of their impairment are never computed at the community level, the perspective taken in this paper. In this paper we also show how to integrate a global analysis with the information obtained at the community level. We assume that a community structure is present which is often the case with infrastructure systems.

COMMUNITY-BASED METRICS

In this section, we build new performance metrics for communities drawing upon the idea of node centrality in weighted graphs.

GLOBAL CENTRALITY

The cost of delivering a service from generation to distribution is an important metric to assess the viability of an infrastructure. It is assumed here that such cost is univocal for an infrastructure network in its design state, and further that the design state of the system is characterised by the service being routed through the shortest paths between the generation and the distribution point. This represents a best-case scenario for agents operating infrastructure systems, as the cost incurred in running these networks is proportional to the distance over which the service is delivered. While it is not always possible to route the service along the shortest paths due to the finite capacity of the network components, this assumption provides a baseline to understand the structure of the system.

Let *G=(N,M)* be the graph representing the network infrastructure of interest, with *N* being the node set and *M* being the edge set. The cardinalities of such sets are respectively *n* and *m*. The average shortest path *D* between two nodes of the system is equal to:

$$D = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} d_{ij} \quad (1)$$

where d_{ij} is the shortest path between origin node *i* and destination node *j.* In the design state of the network the average cost of delivering service across the system is proportional to the average shortest path. However, when two nodes are disconnected the distance between them tends to infinity and the average cost diverges, too. In order to circumvent this issue, the average efficiency *E* of a network has been proposed in (Latora & Marchiori 2007):

$$E = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \frac{1}{d_{ij}} \quad (2)$$

With this formulation, when two nodes become disconnected and the distance between them tends to infinity, the efficiency between them $1/d_{ij}$ drops to zero, making it possible to account for disconnected components. Efficiency is thus a suitable indicator of infrastructure system performance; it contains information about the operability of the system while also being computable on disconnected graphs. Besides network efficiency, other performance indicators can be devised, such as the connectivity among the set of nodes where the service is produced and those where the service is delivered ((Murray et al. 2008), (Johansson & Hassel 2014)) or the total network throughput (Ouyang et al. 2009), but in this paper the discussion is limited to efficiency and its application to a network partitioned in communities.

When the system is stressed by removing a node (or an edge), the efficiency of the system reduces as shortest paths originally going through the removed node (or edge) become longer. This efficiency drop can be used to represent the degradation of the system performance. Computing the efficiency drop allows for a representation of performance that does not simply assess whether the nodes are connected, but also accounts for the quality of those connections. The efficiency drop associated to the removal of a node is defined Information Centrality (Latora & Marchiori 2007) and in this paper it is labelled as global centrality, *GC,* to underscore its global nature:

$$GC^k = \frac{E^0 - E^k}{E^0} \quad (3)$$

where E^0 is the efficiency of the network in its original configuration and E^k is the efficiency of the network after the removal of node *k*. It has been shown for several systems (e.g. (Dunn & Wilkinson 2013), (Ouyang 2013) and (Fang et al. 2015)) that there is a good correlation between the critical elements identified with network metrics such as *GC* and more physically accurate engineering-based models.

GC is an indicator that expresses the criticality of nodes for the operations of the whole infrastructure system. Therefore, it caters only to the need of stakeholders that are concerned with the operations of the entire network, such as the system operators. Next section provides them with new metrics to assess the local importance of infrastructure assets.

INTRACOMMUNITY AND INTERCOMMUNITY CENTRALITIES

DEFINITIONS

Once the network is partitioned into network communities, it is possible to de-average network efficiency *E* and introduce a Community Efficiency matrix ***CE***. The elements of ***CE*** are computed as follows:

$$CE_{ij} = \frac{1}{n_{C_i}(n_{C_j} - \delta_{ij})} \sum_{k=1}^{n_{C_i}} \sum_{l=1}^{n_{C_j}} \frac{1}{d_{kl}} \quad (4)$$

where d_{kl} is the shortest path between origin node *k* and destination node *l*, n_{ci} and n_{cj} are the number of nodes in communities *i* and *j*, and δ_{ij} is the Kronecker delta function required to adjust the total number of *(i,j)* pairs when the origin and destination communities are the same. Each CE_{ij} element represents the efficiency of communication between community *i* and community *j* in the current configuration of the system. The diagonal elements of ***CE*** map the efficiency within communities, whereas the extra-diagonal elements reflect how efficient the connection of each community to the rest of the system is. This allows for a first comparison among communities; it is possible to assess whether any of them is underperforming its peers and, if necessary, improve its internal or external efficiency. We are not suggesting that all communities should have an equal level of efficiency, as their development at the local level is demand-driven, but the use of ***CE*** in conjunction with the requirements on the different communities of the system would allow for a transparent discussion among system stakeholders about what improvement each community needs.

By removing each node in turn two new centrality indicators can be obtained from the change of the elements of the ***CE*** matrix. These two centrality indicators are called intracommunity centrality, *IC,* and intercommunity centrality, *EC*. For node *k* belonging to community *i*, they are defined as (Galvan & Agarwal 2015):

$$IC^k = \frac{CE_{ii}^0 - CE_{ii}^k}{CE_{ii}^0} \quad (5)$$

$$EC^k = \frac{1}{c-1} \sum_{\substack{j=1 \\ j \neq i}}^{c} \frac{CE_{ij}^0 - CE_{ij}^k}{CE_{ij}^0} \quad (6)$$

where CE_{ij}^{k} is the efficiency between communities *i* and *j* after the removal of node *k*, and *c* is the total number of communities in the system. These two indicators provide the same kind of information as *GC*, but at the local level. By leveraging the local average defined by Equation 4, the two indicators identify how much the node they refer to is important for the operations of those communities. *IC* expresses how critical a node is internally for the community where it belongs, whereas *EC* expresses how the removal of the same node influences the transactions between that community and the rest of the system. Figure 2 expresses the difference among the three indicators of centralities. Firstly, in order to assess the *GC* of Node *k*, the change in the efficiency is computed by considering all the residual pairs of origin and destination nodes within the network (Panel A). Secondly, when *IC* is under scrutiny, the change in efficiency is computed only among the nodes of the same community (violet nodes in Panel B). Thirdly, when *EC* is of interest, the efficiency drop is calculated between the community of Node *k* (violet nodes in Panel C) and the other communities (green and orange nodes).

It should be noted that *EC* averages the impact of the node removal over all of the communities, rather than focusing only on the community where the disrupted node originally belonged to. This allows the criticality analysis to maintain a global perspective, as the disruption of a node within a community has the potential to affect other communities at the same time, in terms of their efficiency with respect to the rest of the system. The focus on the individual community could be restored whenever necessary by restricting the computation of *EC* exclusively to those paths between the nodes of the community under investigation and the rest of the system.

Ranking the nodes of a community according to their *IC* and *EC* values highlights which of them are the most critical for its efficiency. Comparing the different values of *IC* and *EC* across different communities allows for an assessment of which communities are dependent on a few nodes and which are most robust to such perturbations. Further, since IC and EC are community level properties, these do not suggest where the node with highest value of IC or EC is located within a community.

NUMERICAL EXPERIMENTS

The *IC* and *EC* metrics depend on the local structure of the community, expressed by variables such as its size and edge density, as well as the global topological parameters such as the total number of communities and the number of connections between them. The behaviour of the newly proposed metrics was examined through synthetic networks with pre-defined community structure. The purpose was to explore which parameters affect the two metrics and thus understand which modifications to infrastructure networks would be the most favourable for maintaining the efficiency of the communities when their nodes are disrupted.

The numerical experiments were carried out in a Matlab programming environment (Mathworks 2017), using the MatlabBGL library as a support for component identification, the Graph Theory Toolbox for shortest path computations and the Community Detection Toolbox for its implementation of the Stability Optimisation algorithm.

The synthetic networks were generated with the Girvan-Newman (GN) model (Newman & Girvan 2004) for reasons given subsequently. The GN model requires the specification of the number of communities *c*, the

number of nodes in each community n_1, the average number of edges each node has within the same community z_i, and the average number of edges every node has outside of its community z_e. The GN model assigns each node a community at the start of the process.

In each realisation of the model, edges are generated with likelihood $p_i=z_i/(n_1-1)$ between nodes of the same community and with likelihood $p_e=z_e/(n_1*(c-1))$ between nodes of different communities. The GN model generates communities which have an intercommunity degree distribution following that of the Erdos-Renyi (ER) random graph model (Erdos & Renyi 1959), with a few extra connections between the communities. Random graphs generated with the ER model have the desirable property of putting the least amount of assumptions on the connectivity pattern of the system, thus representing a suitable choice for a null model used to investigate the behaviour of the new metrics.

Three sets of analyses were performed on graphs generated with the GN model. First, the size and the number of communities in the network were fixed to 25 nodes and 4 communities respectively, and the effects of the number of internal and external edges per node were investigated by varying the values of z_i and z_e. The parameters under scrutiny were varied in the [2:10] and [0.1:2] ranges respectively. Both variables were sampled at 20 points each in the respective intervals, thus 400 (z_i, z_e) combinations were tested. For every (z_i, z_e) pair 100 realisations of the GN model were generated, *IC* and *EC* were computed for every node in each of these networks and the median values of the two variables were recorded. Then, the expected values of the median *IC* and *EC* were calculated by performing the average of the results obtained over the 100 networks. The results are plotted in Panel A and Panel B of Figure 3.

In the second set of analyses, the number of communities c and the average number of external edges per node z_e were fixed to 4 and 0.2 respectively and the effects of the community size n_1 and the average number of internal edges per node z_i were investigated. The parameters under scrutiny were varied in the [2:10] and [6:25] and ranges respectively. Both variables were sampled at 20 points each in the respective intervals: again, 400 (n_1, z_i) combinations were tested. For each (n_1, z_i) pair 100 different graphs were generated, and expected values of the median *IC* and *EC* were computed using the same methodology as before. The results are plotted in Panel C and Panel D of Figure 3.

The third set of analyses was designed to investigate the effects of the number of communities c and the average number of external edges z_e. In this case, n_1 and p_i were fixed to 10 and 2, for each (c, z_e) pair in the [2:12] and [0.1:2] ranges 100 different graphs were again generated at each of the 400 sampling point, and the same methodology as before was followed. The results are plotted in Panel E and Panel F of Figure 3.

Findings and Discussion

Panel A and Panel B of Figure 3 show that *IC* and *EC* are both affected by the internal and the external average degree of the nodes. Increasing z_i results in lower values of the median value of *IC* and *EC*: as the number of efficient paths through the communities increases with z_i, the loss of any single node becomes decreasingly relevant for their internal operations, hence lower *IC*. This applies to *EC* as well, as higher levels of internal

connectivity allow the bypass of intra-community bottlenecks and maintain efficiency in the connection to the other communities. The same applies to z_e: the median value of the *EC* decreases for increasing z_e, as higher values imply that more nodes within each community are connected to other communities. The relation between *IC* and z_e, however, may seem counter-intuitive: *IC* decreases with increasing z_e. The external connectivity of a community mitigates its internal efficiency loss by allowing for communities the use of neighbouring external nodes to efficiently bridge their internal disruptions. This last finding is particularly significant: it means that, if network communities are considered in isolation from each other, the impact of disruptions to their internal operations is overstated by neglecting the possibility of using the neighbouring communities to compensate for the loss. This is most relevant for communities with low average internal degree: low z_i can be efficiently compensated by increasing z_e if $z_i < 4$. Provided that many infrastructure systems (Barthélemy 2011) are only sparsely connected even at the local level, this finding represents an opportunity to improve the local performance of infrastructure networks while at the same time improving the efficiency of communication between different regions.

Panel C and Panel D of Figure 3 highlight another aspect of *IC* and *EC*: their median value depends on the size of the system. The higher the number of nodes in each community, the lower the median value of intra- and inter-community centrality of its nodes, and this size effect is much stronger than the effect of the internal connectivity. For example, when $z_i = 2$, the median value of *IC* obtained for communities with $n_1 = 12$ is 0.54 times the value obtained for communities with $n_1 = 6$, i.e. upon doubling the size of the community, the median *IC* roughly halves. When the number of internal edges per node doubles, however, the median value *IC* stays about the same; for $n_1 = 6$, doubling the average number of internal edges per nodes from 2 to 4 yields a reduction of *IC* of only 5.13%. Increasing the local connectivity of smaller communities is not a viable alternative to cope with their inherent lack of redundancy: the operations of these communities are much more susceptible to any disruption simply because of their small size, and as such they need to be treated carefully by the system owners.

Panel E and Panel F allow for some final insights: first, as it would be expected, the total number of communities *c* in the system does not influence sensibly the median value of *IC*, while the average number of external edges per node z_e does. The median value of *EC*, on the other hand, is influenced by both *c* and z_e.

CROSS-SCALE CENTRALITY

The two new community metrics, *IC* and *EC* defined above, account for the criticality of a system component at one particular scale. However, the capacity of infrastructure systems to deliver service can be evaluated at multiple scales, each represented by the different partitions identified during community detection. A node found to be critical at most scales would be of interest to all stakeholders. In order to allow for an assessment of component criticality across scales, a third metric, cross-scale centrality *XC*, is introduced here. It is obtained for every node by averaging over all partitions a linear combination of its normalised *IC* and *EC* values. The normalisation is done with respect to the maximum value in the each community.

For node k belonging to communities $i_1, i_2, \ldots, i_{np}$ in the n_p different partitions considered in the analysis, *XC* can be written as:

$$XC^k = \frac{1}{n_p}\sum_{j=1}^{n_p}\gamma_j\left(\alpha\frac{IC^k}{\max(IC)_{i_j}} + (1-\alpha)\frac{EC^k}{\max(EC)_{i_j}}\right) \quad (7)$$

where $max(IC)_{ij}$ is the maximum value of IC in community i_j where node k belongs in the j^{th} partition, and the same for *EC*. In order for *XC* to be bounded between 0 and 1, coefficient α in Equation (7) varies between 0 and 1 and the weights γ_j need to sum to one.

Coefficient α can be used to weigh more heavily the internal or the external operations of the communities. The two extreme cases of α = 1 and α = 0 represent the choice of using only one of the two community centrality indicators to compute *XC*. Selecting α = 0.5, on the other hand, implies attributing the same value to nodes affecting the internal and external efficiency of communities.

The weights γ_j can be applied to the results obtained at every scale to emphasize the information yielded by some partitions, reflecting the power dynamics between the stakeholders interested in functionality at that scale. For example, if a partition is of particular interest because budgeting for the infrastructure system is done at a comparable scale, then it can be weighted more heavily than others.

XC has the highest values for those nodes that are critical at multiple scales of system description, while having low or average values otherwise. Once *XC* has been obtained, it can be used in conjunction with *GC* to express the criticality of the network elements from both a global and a local perspective. The advantage of using *XC* consists in having a clearer view of the different stakeholder needs. By accounting for all the scales at which the system delivers its service, *XC* facilitates the discussion among stakeholders about which sections of the network require additional investments. If it is considered sufficient, only one partition can be used (for example to represent the interplay between the national and regional performance of a railway system), and in that case *XC* reduces to a linear combination of the intra- and inter-community centralities.

As noted previously, EC metric in its current form averages out the impact across all communities in a partition. However, our objective has been to capture both the local and global impacts of a disruptive event and metric *XC* enables the restoration of information that might not be fully considered at one level of system description.

The identification of the most appropriate number of partitions n_p used to describe a system is an open research issue for multi-scale community detection methods (Fortunato & Barthelemy 2007) including Stability Optimisation used here. In the case of infrastructure systems, however, it is not an obstacle, rather it is an opportunity to tailor the analysis to the needs of the stakeholders. Community detection is used here to highlight the criticality of network components for multiple levels of system organisation. If the decision-maker requires accounting for a plurality of stakeholder needs, community detection can be used to yield a higher number of partitions by relaxing the requirements for a partition to be identified. Conversely, if too many partitions are

355 deemed to make the analysis too complex, then requirements that are more stringent can be set. An example of
356 how Stability Optimisation can be tuned to yield a different number of partitions is provided when the Case Study
357 is presented. Practical and system-dependent considerations might also guide the choice of the number of
358 partitions to use in the analysis. For example, when assessing a water distribution network, the analysis can be
359 limited to partitions characterised by communities not smaller than neighbourhoods served by the same pumping
360 station.

361 A final remark is about the normalisation of centralities in Equation 7. For each node, its centralities are
362 normalised by the corresponding maximum value within the same community. This implies that all communities
363 in every system partition are given the same importance, as multiple nodes in different communities can achieve
364 a unitary contribution, based on their performance within the region. This assumption can be removed by
365 normalising by the same value across different communities.

366 A CASE STUDY – GREAT BRITAIN RAILWAY NETWORK

367 THE MODEL

368 A model of the Great Britain Railway System was built to demonstrate the application of the proposed community
369 metrics. The topology of the model (Galvan & Agarwal 2015) was derived from the main stations and routes map
370 available on the National Rail website (National Rail 2015) and includes a selected subset of the stations and
371 routes constituting the system. The model is composed of n = 148 nodes representing the stations (see Table A1
372 in Appendix A for a list of stations) and m = 270 edges connecting them through the railway lines. The weights
373 assigned to the edges reflect the minimum travel time between the nodes they connect. Such a measure was
374 chosen over the physical distance between two stations because it represents a more realistic proxy of the cost
375 of travel: it implicitly includes non-spatial constraints such as the maximum speed of the vehicles and the
376 capacity of the lines. Data regarding travel time between adjacent stations was obtained from the National Rail
377 website. Figure 4 shows the frequency distribution of nodal degrees in Panel A, while Panel B shows the fit of the
378 degree distribution to a Poisson Cumulative Distribution Function (CDF). Table 2 illustrates the main topological
379 features of the system.

380 Table 3 reports the global centrality values for the ten highest-ranking nodes of the system and Figure 5 plots the
381 distribution of *GC* on the whole system. Seven out of the ten most central nodes according to *GC* are located in
382 the London region or in the immediate vicinity (Nodes 97, 98, 99, 100 101 and 109 are located in London, Node
383 105 is adjacent to it), reflecting the role of the city in the national railway system. As a large number of shortest
384 paths between the nodes of the system go through the London region, any disruption to these nodes has the
385 potential to reduce significantly the efficiency of the network. The only three peripheral nodes that make it in the
386 list, namely Exeter, Preston and York (Nodes 135, 32 and 38), are bottlenecks for their communities. Exeter is
387 the node governing the accessibility to the whole of Devon and Cornwall, whereas Preston and York lie
388 respectively on the East and West coastal paths from to the North of England and Scotland: the efficiency of

paths towards these three large regions of the system depend on them. These are the only peripheral nodes that rank as high as the London nodes in a global analysis: a global metric such as global information centrality discounts the importance that other nodes have at the community level.

COMMUNITY DETECTION

The communities on the railway network were identified using the Stability Optimisation: the results of community detection are presented in Figure 6, where the number of communities identified during each run of stability optimisation is plotted in Panel A against the value of the scale parameter t. The algorithm was run 100 times for values of t in the [0:1] interval and 100 times in the [1:100] interval. The value t = 1 was chosen as the threshold between the two regimes because for t = 1 the optimisation of stability and modularity are equivalent: below 1 stability optimisation finds partitions finer than the resolution limit, while above it yields larger partitions.

For every value of the scale parameter t, stability optimisation yields a partition. Plateaus in the stability plot correspond to partitions identified during consecutive runs of the community detection procedure. The larger the plateau, the more robust the partition. In order to discriminate between partitions that are simply a product of the algorithm sweeping through different t values and those that represent functional subsystems within the network, robustness threshold, the number n_t of successive values of t for which a partition is identified, is used (Lambiotte 2010). In Panel B of Figure 6 the number of partitions selected for analysis is plotted as a function of the n_t robustness parameter. The two partitions persisting for the largest ranges of t values are the nine communities partition identified for t = 1 and the three communities partition identified for t > 40, each identified for n_t = 26 consecutive runs of the algorithm. Progressively relaxing the n_t threshold allows for the inclusion of other partitions in the analysis. In the following sections the six partitions persisting for n_t = 5 are used to demonstrate the application of the new metrics. The choice of the most appropriate number of partitions (and therefore of the threshold n_t value), is still an open research question and in practice will be led by high-level considerations about the objective of the analysis. The main characteristics of the six partitions identified are given in Table 4. The number of communities in each partitions varies from 26 to 3, and the average number of nodes goes from 5 nodes in P_1 to 49 nodes in P_6.

P_1 is the first stable partition identified, and with 26 communities of an average size of 5.7 nodes sits at an intermediate scale between the county and the regional level. Some communities of P_1 adhere perfectly to the county subdivision of Great Britain, as shown in Panel A of Figure 7: Community 23, for example, fits nicely within the boundaries of the Dorset County on the southern coast of Great Britain. The neighbouring Community 17, however, is larger and includes nodes belonging to the counties of Somerset, Devon and Cornwall. Partition P_6 is at the opposite end of the spectrum: it is the coarsest subdivision identified by stability optimisation, and consists of only three communities (Figure 7, Panel B). They are articulated around London, as the city is a gateway to access the North of England, and splits the South in a region that is extremely well connected to it (the South-East) and one where the network is sparser (the South-West). The remaining six partitions have sizes in between P_1 and P_6, and in the following P_4 is examined in greater detail. Partition P_4 is identified at t = 1 and by

inspecting its communities and the nodes belonging to them, it is possible to assess the similarity between the 11 regions of Great Britain and the 9 communities of P_4, both shown in Figure 8. The boundaries of Community C_4, for example, largely correspond to those of Scotland (region 1), while Community C_8 is constituted by the nodes that model the London (region 8) railway stations, and Community C_7 covers the whole of the South-West of England (region 10).

This correspondence can be quantified by the Normalised Mutual Information (NMI) (Danon et al. 2006). Given two partitions *X* and *Y*, *NMI(X,Y)* is the amount of information that is gained about one by knowing the other: in this case it is used to represent the overlap between the partition found by the algorithm (say, partition *X*) and the regional structure of the network (say, partition *Y*). *NMI* is computed as:

$$NMI(X,Y) = \sum_{x,y} p(x,y)\, log \frac{p(x,y)}{p(x)\, p(y)} \qquad (8)$$

where the summation is performed over all communities represented by *x* and *y*, *p(x,y)* is, for each community in partition *X*, the number of nodes belonging to the corresponding community in partition *Y*, and *p(x)* and *p(y)* the number of the nodes in those communities according to the two partitions. *NMI* is bounded between 0 and 1. Between Partition P_4 and the regional subdivision of Great Britain a value of *NMI* = 0.714 is obtained. The 1% significance value of *NMI* computed with a permutation test is equal to 0.189. Such a test was performed by generating 10^4 random partitions for a network with 148 nodes and 11 communities, and computing the 99th percentile of the distribution of *NMI* between each of the synthetic partitions and P_4. A value of 0.714 reflects the high level of similarity between the two partitions: the community detection algorithm is thus able to detect the regional structure of the railway system.

Where the partitions do not overlap, however, additional insight can be gained on the functionality of the system by inspecting it. One example is the North of Wales. While it is operated as part of the Welsh railway network, community detection shows that it belongs to the same basin of attraction as the North East of England. Community detection highlights the fact that for its operations this part of Wales depends on the North East of England, and therefore it might be more effectively managed as part of the adjacent community.

INTRACOMMUNITY AND INTERCOMMUNITY CENTRALITIES

The community efficiency matrix ***CE*** of the railway network for the 9-community partition P_4 is shown in Figure 9. The diagonal elements are the internal efficiencies of the communities: by comparing them between each other the dramatic difference between the London community C_8 and the rest of the system becomes apparent. The London community has an efficiency of 0.079 while the other communities have an average efficiency of 0.016. Further, by comparing the off-diagonal elements of a column or row to the corresponding diagonal element it is possible to assess the relative efficiency between the internal and external operations of a community and how this ratio compares with the other communities. While for all other communities the ratio between the internal and

456 the average external efficiency is between 2 and 3, the London community scores a value of 8.2, highlighting
457 once again its superior level of internal connectivity.

458 Table 5 provides, for each community, the nodes with the highest values of *IC* and *EC*. The results for the whole
459 system are presented in Figure 10. The differences and the analogies in the outcome of the global and the
460 community-based analyses can be assessed by comparing the elements of Table 5 to the most critical nodes
461 according to *GC* (Table 3 and Figure 5, or *GC* rank column in Table 5). Some nodes critical at the global level,
462 for example, also emerge as local hubs. These are King's Cross in London, Preston in the North-West of
463 England and Exeter in the South-West. Further, the community centrality analysis also embeds system-wide
464 information: the nodes of Reading and Stratford show the highest *EC* in their respective communities because of
465 their vicinity to the London community C_8, which is a hub for the long-range efficient transport.

466 The community centrality analysis however, yields new information: nodes such as Ashford (128) in the South or
467 Newcastle (22) in the North are the most central for the internal operations of their communities, although they
468 only rank 41st and 57th in terms of global information centrality. Disruptions to these two nodes have serious
469 consequences at the local level: disregarding those means neglecting the needs of the stakeholders that operate
470 at the community level. It is also interesting to note that in South-East, South-West and North-West communities
471 (C_1, C_7 and C_9) the nodes ranking the highest in terms of *IC* (Nodes 128, 135 and 135 respectively) are also the
472 most critical node when assessed for *EC*. Not only any disruption to those nodes would markedly reduce the
473 internal performance of the community where they belong, but it would also impair the efficiency of service
474 delivery between those communities and the rest of the system.

475 The *EC* results highlight the role of the London community C_8 as a hub of whole-system connectivity: the highest-
476 ranking node in all of the adjacent communities is a direct neighbour to the London community. In absolute
477 terms, the highest value of *EC* is obtained for Node 100, at the interface between communities C_8 and the East
478 of England community C_5. This community has a single point of connection with the rest of the system, and this
479 situation is unique in the network, as all other communities are connected in several positions due to the meshed
480 structure of the railway system. The result is a value of *EC* = 8.41% for the London community and 6.88% for the
481 East of England community, both substantially above those obtained for the other communities. The opposite is
482 true for the station of Darlington in the Scotland community C_4: the community is very remote and presents two
483 equally valid alternative group of paths (on the East and West coast) to reach the southern regions of the
484 network. This means any disruption to either has very little consequences on the efficiency of communication
485 with the rest of the system, with *EC* = 1.7%.

486 CROSS-SCALE CENTRALITY AND CRITICAL ELEMENTS

487 The community centrality analysis performed for partition P_4 and illustrated above was replicated for the other
488 five partitions identified with community detection, and the results were combined in the cross-scale centrality
489 indicator *XC* in an effort to trace the criticality of the different infrastructure assets at multiple scales. The relative

contributions of *IC* and *EC* towards *XC* depends on the α coefficient and the partition weights. Every partition was equally weighted, while three different values of α, equal to 0, 0.5 and 1, were considered.

Table 6 provides the results of a cross-scale centrality analysis by reporting the 10 nodes reaching the highest *XC* values for three different values of α. Some nodes attain high levels of *XC* independently of the α value selected for the assessment. This is the case of nodes 135, 98, 32 and 105: these nodes represent true criticalities within the system, governing its efficiency at multiple scales of description, within and across communities. As a whole, however, the list of the top 10 nodes by *XC* presented in Table 6 inevitably changes for different α. Higher values rank more heavily the nodes governing local efficiency of communities, such as nodes 16, 22 or 49. The opposite is true for nodes such as 55, 52 or 28, which are not central for the efficiency of connections within their community, but reach high levels of *EC* across different partitions, and this results in high *XC* scores for lower values of α.

The use of *XC* in a multi-stakeholder environment is exemplified in the following. Let us assume that there are three stakeholders concerned with the assessment of the railway network. The first consists of municipal authorities (A) interested with the functionality of the network at the County scale (P_1), because that maps to the needs of their constituencies. The second is Network Rail (B), which manages the network by routes that coarsely correspond to the regions of Great Britain (P_4) and needs to allocate resources to each of them. The third and last one is the Office of Rail and Road (C), which needs to assess the performance of the system as a whole. When evaluating the criticality of every asset within the network, the stakeholders in Group A use *XC* that is the weighted average of *IC* and *EC* computed on the communities of P_1 to assess the impact on the functionality of communities from the impairment of each node. Stakeholder B uses the same metrics but computed on the communities of P_4, while Stakeholder C, concerned with the whole system, uses the outcome of a global analysis: *GC*. This produces a different ranking for the criticality of the individual network components, because stakeholders defined criticality at different scales. The Spearman correlation coefficient of the three rankings are, respectively ρ_{AB} = 0.82, ρ_{AC} = 0.81 and ρ_{BC} = 0.85, indicating that while there is disagreement in the prioritisation of the different nodes, there also are some converging interests. Aggregating the local analyses using cross-scale centrality *XC* and comparing these values with their *GC* score (Fig. 11) allows for a transparent evaluation of the needs of the three stakeholders, identifying the assets that are of interest to all of them and the ones where there is disagreement.

The horizontal axis of Fig. 11 expresses how much every node affects the global behaviour of the system, while the vertical one synthesises the importance of nodes across the (possibly many) local descriptions of the network performance yielded by *IC* and *EC* at different scales. This way, no information is neglected and this can be used to facilitate discussion about resource allocation among the three stakeholder groups.

It is possible to identify nodes such as Exeter or King's Cross (Nodes 135 and 99) which fall into a high-priority class (red nodes – Q1) for everyone, as they achieve high values of both *GC* and *XC*. Not only these are critical at the global level, but their removal jeopardises the efficiency of their communities at multiple scales. The

opposite is true for nodes falling in the low *GC* – low *XC* group (purple nodes – Q3): nodes such as 31 (Blackpool) are peripheral for the whole system and smaller communities are not dramatically affected by any disruption to them.

Nodes in the low GC – high XC quadrant (green nodes – Q2), represent a more interesting case. Nodes such as 117, representing the station of Ramsgate, do not attain high scores in a global analysis, due to the redundancies of the system that can cope well with their removal. If they are evaluated from a local perspective, however, they stand out as those nodes that can compromise the efficiency of the communities they belong to. The system owner (in this case, the Office of Rail and Road) should treat these elements with caution, as stakeholders concerned with local operations may consider them to be of the highest importance. Those nodes falling in the high *GC* – low *XC* region (blue nodes – Q4), represent the final case: elements such as Node 58, which models the station of Sheffield, are a concern for the system operator, but their importance can be downplayed by local stakeholders such as municipal authorities.

CONCLUSION

Global assessments of criticality fail to identify the system elements governing the service delivery performance at the local level, as this is governed by short-range interactions between neighbouring elements. Community detection forms the basis of new metrics that play a central role in the new approach. The first metric, intracommunity centrality, *IC*, accounts for the efficiency loss within a community that follows the removal of a node. The second, intercommunity centrality, *EC*, maps the efficiency loss between that community and the rest of the network.

Larger communities generate lower values of *IC* and *EC* for individual nodes indicating that the nodes are more disposable. *IC* is influenced by both the internal average degree of its communities, as well as by their external average degree. For sparse networks, adding edges between communities can be more efficient than adding edges within communities, if the objective is to reduce *IC* across the board. *EC* is influenced by the number of communities within the network and the external average degree of their nodes. At the same time, the internal average degree plays a role as it allows for the bypassing of internal bottlenecks.

A third metric, cross-scale centrality, *XC,* combines the *IC* and *EC* scores obtained by the network nodes to enable an assessment of their criticality accounting for the role they play at all the meaningful scales of system description. *XC* can be tuned to account more heavily for *IC* or *EC*, depending on the objective of the specific analysis. Used in conjunction to global indicators such as *GC*, cross-scale centrality allows for a synthesis of the information for decision-making.

A global analysis of the railway network of Great Britain reveals that most of the high-centrality nodes are located in London or immediately around it. The application of the new community centrality metrics, however, shows that, at the local level, some peripheral nodes can be just as critical. This information was suppressed by the

558 global average used to compute *GC*, but is restored with the use of *IC*. *EC*, on the other hand, accounts for long-
559 range interaction within communities, thus restoring a global perspective in the community-based assessment. *IC*
560 and *EC* obtained at different scales are then combined in the *XC* indicator. By using *XC* in conjunction with the
561 outcome of the global analysis some low *GC* nodes, although neglected by global assessments, emerge as
562 having the potential to be mission-critical for a wide variety of stakeholders concerned with local performance.

563 While the approach presented in this paper facilitates the identification of critical elements and provides decision-
564 makers with tools to explore the local behaviour of infrastructure systems at multiple scales, in future work it will
565 be expanded to include multiple contingencies, other indicators of service delivery and interactions with the
566 natural hazards threatening the performance of the network and that of its communities.

567 ACKNOWLEDGEMENTS

568 All of the numerical experiments were carried out in a Matlab programming environment (Mathworks 2017), using
569 the MatlabBGL library as a support for component identification, the Graph Theory Toolbox for shortest path
570 computations and the Community Detection Toolbox for its implementation of the Stability Optimisation
571 algorithm.

572 The authors would like to thank the EPSRC (DTA Grant EP/L504919/1) and the Systems Centre at the University
573 of Bristol, the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1), for financial support
574 to the first author. The funding organisations were not involved in the design of the study nor in the writing of the
575 report.

576 The authors also thank the anonymous reviewers for their constructive comments.

577 APPENDIX A

578 Please refer to the supplementary material for (i) brief notes on Stability Optimisation (Delvenne et al. 2010) and
579 (ii) Table A1 identifying the nodes with station names.

580 REFERENCES

581 Albert, R., Albert, I. & Nakarado, G.L., 2004. Structural vulnerability of the North American power grid. *Physical*
582 *review. E, Statistical, nonlinear, and soft matter physics*, 69(2 Pt 2), p.25103.

583 Barabási, A.-L., 2014. *Network Science*, Cambridge University Press.

584 Barabási, A.-L., 2009. Scale-free networks: a decade and beyond. *Science (New York, N.Y.)*, 325(5939),
585 pp.412–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19628854 [Accessed July 12, 2014].

586 Barabasi, A. & Albert, R., 2002. Statistical mechanics of complex networks. *Review of Modern Physics*,
587 74(January), pp.47–97.

588 Barthélemy, M., 2011. Spatial networks. *Physics Reports*, 499(1–3), pp.1–101.

589 Blockley, D. & Godfrey, P., 2000. *Doing it differently: systems for rethinking construction*, Thomas Telford.

590 Borgatti, S.P., 2005. Centrality and network flow. *Social Networks*, 27(1), pp.55–71.

591 Bye, B.L., 2011. Volcanic eruptions: science and management. *Science*, 2(27).

592 Cai, K.-Q., Zhang, J., Du, W. & Cao, X., 2012. Analysis of the Chinese air route network as a complex network.
593 *Chinese Physics B*, 21(2), p.28903.

594 Cardillo, A., Zanin, M., Gómez-Gardeñes, J. et al., 2013. Modeling the multi-layer nature of the European Air
595 Transport Network: Resilience and passengers re-scheduling under random failures. *European Physical*
596 *Journal: Special Topics*, 215(1), pp.23–33.

597 Carvalho, R. , Buzna, L., Bono, F., et al., 2014. Resilience of Natural Gas Networks during Conflicts , Crises and
598 Disruptions. , 9(3), pp.1–9.

599 Dalziell, E. & Nicholson, A., 2001. Risk and impact of natural hazards on a road network. *Journal of*
600 *transportation engineering*, 127(April), pp.159–166.

601 Danon, L., Díaz-Guilera, A. & Arenas, A., 2006. The effect of size heterogeneity on community identification in
602 complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(11), pp.P11010–
603 P11010.

604 DeLaurentis, D., Han, E.-P. & Kotegawa, T., 2008. Network-Theoretic Approach for Analyzing Connectivity in Air
605 Transportation Networks. *Journal of Aircraft*, 45(5), pp.1669–1679.

606 Delvenne, J.-C., Yaliraki, S.N. & Barahona, M., 2010. Stability of graph communities across time scales.
607 *Proceedings of the National Academy of Sciences of the United States of America*, 107(29), pp.12755–60.

608 Dueñas-Osorio, L., Craig, J.I., Goodno, B.J. & Bostrom, A., 2007. Interdependent response of networked
609 systems. *Journal of Infrastructure Systems*, 13(3), pp.185–194.

610 Dunn, S., Wilkinson, S. & Ford, A., 2016. Spatial structure and evolution of infrastructure networks. *Sustainable*
611 *Cities and Society*.

612 Dunn, S. & Wilkinson, S.M., 2013. Identifying Critical Components in Infrastructure Networks Using Network
613 Topology. , pp.157–165.

614 England, J., Blockley, D. & Agarwal, J., 2008. The vulnerability of structures to unforeseen events. *Computers*
615 *and Structures*, 28, pp.1042–1051.

616 Erdos, P. & Renyi, A., 1959. On random graphs. *Publicationes Mathematicae Debrecen*, 6, pp.290–297.

617 Fang, Y.-P. & Zio, E., 2013. Unsupervised spectral clustering for hierarchical modelling and criticality analysis of
618 complex networks. *Reliability Engineering & System Safety*, 116, pp.64–74.

619 Fang, Y., Pedroni, N. & Zio, E., 2015. Optimization of cascade-resilient electrical infrastructures and its validation
620 by power flow modeling. *Risk analysis*, 35(4), pp.594–607.

621 Fortunato, S., 2010. Community detection in graphs. *Physics Reports*, 486(3–5), pp.75–174.

622 Fortunato, S. & Barthelemy, M., 2007. Resolution limit in community detection. *Proceedings of the National*
623 *Academy of Sciences of the United States of America*, 104(1).

624 Galvan, G. & Agarwal, J., 2015. Community detection and infrastructural criticality. In: Zio et al (Eds), Safety and
625 reliability of complex engineered systems - Proceedings of the 25th European Safety and Reliability
626 Conference (ESREL) 2015, CRC Press. pp. 1209-1217.

627 Gómez, C., Sánchez-Silva, M. & Dueñas-Osorio, L., 2014. An applied complex systems framework for risk-based
628 decision-making in infrastructure engineering. *Structural Safety*, 50, pp.66–77.

629 Grubesic T. H., M.A.T., 2006. Vital Nodes, Interconnected Infrastructures, and the Geographies of Network
630 Survivability. *Annals of the Association of American Geographers*, 96, p.21.

631 Hong, L., Ouyang, M., Peeta, S., He, X. & Yan, Y., 2015. Vulnerability assessment and mitigation for the Chinese
632 railway system under floods. *Reliability Engineering & System Safety*, 137, pp.58–68.

633 Johansson, J. & Hassel, H., 2010. An approach for modeling interdependent infrastructures in the context of
634 vulnerability analysis. *Reliability Engineering and systems safety*, 95, pp.1335–1344.

635 Johansson, J. & Hassel, H., 2014. Impact of Functional Models in a Decision Context of Critical Infrastructure
636 Vulnerability Reduction. *Vulnerability, Uncertainty, and Risk*, pp.577–586.

637 Kurant, M. & Thiran, P., 2006. Extraction and analysis of traffic and topologies of transportation networks.

Physical review. E, Statistical, nonlinear, and soft matter physics, 74(3 Pt 2), p.36114.

Kurant, M.I., Thiran, P. & Hagmann, P., 2007. Error and attack tolerance of layered complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 76(2), pp.1–5.

Lambiotte, R., 2010. Multi-scale modularity in complex networks. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*. pp. 546–553.

Lämmer, S., Gehlsen, B. & Helbing, D., 2006. Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications*, 363(1), pp.89–95.

Latora, V. & Marchiori, M., 2007. A measure of centrality based on network efficiency. *New Journal of Physics*, 9(6), pp.188–188.

Lorenz, J., Battiston, S. & Schweitzer, F., 2009. Systemic risk in a unifying framework for cascading processes on networks. *The European Physical Journal B*, 71(4), pp.441–460.

Le Martelot, E. & Hankin, C., 2011. Multi-scale community detection using stability optimisation within greedy algorithms. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*. pp. 216–225.

Mattsson, L. & Jenelius, E., 2015. Vulnerability and resilience of transport systems - a discussion of recent research. *Transportation Research Part A*, 81, pp.16-34.

Mathworks, 2017. MATLAB. Available online: https://uk.mathworks.com/products/matlab.html.

Mena, R., Hennebel, M., Li, Y-F., Zio, E., 2014. Self-adaptable hierarchical clustering analysis and differential evolution for optimal integration of renewable distributed generation. *Applied Energy*, 133, pp.388–402.

Mossa, S., Turtschi, A, & Amaral, L.A., 2005. 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), pp.7794–7799.

Mureddu, M., Caldarelli, G., Chessa, A., Scala, A., & Damiano, A., 2015. Green Power Grids: How Energy from Renewable Sources Affects Networks and Markets. *PloS one*, 10(9), p.e0135312.

Murray, A.T., Grubesic, T.H. & Matisziw, T.C., 2008. A Methodological Overview of Network Vulnerability Analysis. *Growth and Change*, 39(20), p.573.

National Rail, 2015. *Maps of the UK Railway Network*, Available at: http://www.nationalrail.co.uk/stations_destinations/maps.aspx.

Network Rail, 2014. *Dawlish*, Available at: http://www.networkrail.co.uk/timetables-and-travel/storm-damage/dawlish/.

Newman, M. & Girvan, M., 2004. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), p.26113.

Office of Rail and Road, 2015. *Station usage 2014 - 2015 data*, Available at: http://orr.gov.uk/__data/assets/excel_doc/0019/20179/Estimates-of-Station-Usage-in-2014-15.xlsx.

Ouyang, M., Hong, L., Mao, Z-J., Yu, M-H., & Qi, F. 2009. A methodological approach to analyze vulnerability of interdependent infrastructure. *Simulation Modelling Practice and Theory*, 17, p.12.

Ouyang, M., Pan, Z., Hong, L., & Zhao, L., 2014. Comparisons of complex network based models and real train flow model to analyze Chinese railway vulnerability. *Reliability Engineering & System Safety*, 123, pp.38–46.

Ouyang, M., 2013. Comparisons of purely topological model, betweeness based model and direct current power flow model to analyze power grid vulnerability. *Chaos*, 23, p.10.

Ouyang, M., 2014. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering and systems safety*, 121, p.18.

Poljansek, K., Gutierrez, E. & Bono, F., 2012. Seismic risk assessment of interdependent critical infrastructure systems: The case of European gas and electricity networks. *Earthquake Engineering and Structural Dynamics*, 41, p.19.

Reggiani, A., Nijkamp, P. & Lanzi, D., 2015. Transport resilience and vulnerability: The role of connectivity. *Transportation Research Part A*, 81, pp.4-15.

Reichardt, J. & Bornholdt, S., 2006. Statistical mechanics of community detection. *Physical Review E*, 74(1).

Rocco, C.M. & Ramirez-Marquez, J.E., 2011. Vulnerability metrics and analysis for communities in complext networks. *Reliability Engineering and systems safety*, 96, p.7.

Sen, P., Dasgupta, S., Chatterjee, A. et al., 2003. Small-world properties of the Indian railway network. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 67(3 Pt 2), p.36106.

Strano, E., Cardillo, A., Scellato, S., et al., 2007. Street centrality vs. commerce and service locations in cities: a Kernel Density Correlation case study in Bologna, Italy. , p.14.

UNISDR, 2015. *Sendai Framework for Disaster Risk Reduction*, Available at: http://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf.

Vugrin, E.D., Warren, D.E. & Ehlen, M.A., 2011. A resilience assessment framework for infrastructure and economic systems: Quantitative and qualitative resilience analysis of petrochemical supply chains to a hurricane. *Process Safety Progress*, 30(3), pp.280–290.

Wilkinson, S.M., Dunn, S. & Ma, S., 2011. The vulnerability of the European air traffic network to spatial hazards. *Natural Hazards*, 60(3), pp.1027–1036.

Yazdani, A. & Jeffrey, P., 2011. Complex network analysis of water distribution systems. *Chaos (Woodbury, N.Y.)*, 21(1), p.16111.

Zio, E., Sansavini, G., Maja, R. & Marchionni, G., 2011. An analytical approach to the safety of road networks. *International Journal of Reliability, Quality and Safety Engineering*, 15(1).

Zio, E., 2016. Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety*, 152, pp.137–150.

Zio, E. & Sansavini, G., 2013. Vulnerability of Smart Grids With Variable Generation and Consumption : A System of Systems Perspective. , 43(3), pp.477–487.

TABLES

Table 1. Infrastructure systems modelled as networks in the scientific literature

System	Nodes	Edges	Reference
Electric Power Transmission	Generators and Substations	Power Lines	(Vugrin et al. 2011)
Water Distribution	Pumping Stations, Storage Tanks,	Pipelines	(Yazdani & Jeffrey 2011)
Natural Gas Distribution	Pumping Stations, Storage Tanks	Pipelines	(Carvalho et al. 2014)
Roads Systems	Origins and Destination, Intersections	Roads	(Grubesic T. H. 2006)
Railway Networks	Railway Stations	Railway Lines	(Ouyang et al. 2014)
Telecommunications	Routers	Signal Channels	(Zio & Sansavini 2013)
Airport Networks	Airports	Flight Routes	(Wilkinson et al. 2011)
Structures	Joints	Beams	(England et al. 2008)

Table 2. Topological features of the Great Britain Railway Network model

Network Properties		
Nodes	n	148
Edges	m	270
Average node degree	$\langle k \rangle$	3.64
Degree distribution	$P(k)$	Poisson-like
Average edge weight	$\langle w \rangle$	38.97
Efficiency	E	0.0138

Table 3. Nodes with the 10 highest GC values, and the Region of Great Britain to which they belong

Rank	Node	Station	*GC*	Region
1	98	St. Pancras	6.43%	London
2	99	King's Cross	6.33%	London
3	100	Liverpool Street	5.85%	London
4	135	Exeter	5.06%	South West
5	97	London Euston	4.88%	London
6	32	Preston	4.41%	North West
7	105	Didcot Parkway	4.28%	South West
8	109	Paddington	4.27%	London
9	38	York	4.13%	Yorkshire and the Humber
10	101	Stratford	4.00%	London

Table 4. Number of communities in each of the stable partitions (for $n_t = 5$) and their average size

Partition	Communities	Mean Community Size (number of nodes)
P_1	26	5.69
P_2	12	12.33
P_3	10	14.80
P_4	9	16.44
P_5	6	24.67
P_6	3	49.33

721 **Table 5**. Nodes with the highest IC and EC values and their ranking in the global assessment

	Highest *IC*				Highest *EC*			
Community	**Node**	**Station**	***IC***	***GC* rank**	**Node**	**Station**	***EC***	***GC* rank**
C_1 – South East	128	Ashford	30.73%	41	128	Ashford	4.66%	41
C_2 – Wales	91	Newport	26.94%	19	106	Reading	4.07%	11
C_3 – Yorkshire	49	Doncaster	28.04%	14	61	Grantham	4.05%	18
C_4 – Scotland	22	Newcastle	23.39%	57	28	Darlington	1.77%	40
C_5 – East of England	102	Colchester	32.76%	38	101	Stratford	6.88%	10
C_6 – West Midlands	76	Rugby	30.64%	24	96	Watford	5.35%	23
C_7 – South West	135	Exeter	31.42%	4	135	Exeter	4.34%	4
C_8 – London	99	King's Cross	34.01%	2	100	Liverpool St.	8.41%	3
C_9 – North West	32	Preston	37.26%	6	32	Preston	4.12%	6

722 **Table 6.** Results of *XC* analyses obtained with different α values

Cross-Scale Analyses						***Global Analysis***	
α = 1.00		***α = 0.50***		***α = 0.00***		***(normalised GC)***	
Node	*XC*	Node	*XC*	Node	*XC*	Node	*GC*
135	1.00	135	0.98	106	1.00	98	1.00
98	1.00	98	0.96	105	0.97	99	0.98
32	0.95	32	0.95	135	0.97	100	0.91
99	0.94	105	0.94	32	0.94	135	0.79
91	0.92	99	0.88	98	0.93	97	0.76
105	0.91	128	0.87	100	0.92	32	0.69
49	0.89	76	0.84	52	0.91	105	0.67
16	0.88	100	0.84	128	0.90	109	0.66
22	0.87	55	0.84	28	0.89	38	0.64
120	0.86	52	0.82	101	0.88	101	0.62

723

724

LIST OF FIGURES

747 **Fig. 6.** The results of stability optimisation on the model of the railway network. *(A) The number of*
748 *communities c identified for every value of t. (B) The number of partitions n_p as a function of the*
749 *value of n_t.*

750 **Fig. 7.** Partitions P_1 and P_6 (for n_t = 5). *(A) The detail of the South-West of England in partition P_1:*
751 *while Community formed by brown nodes corresponds to the Dorset country (identified by the*
752 *number 4), the neighbouring Community formed by light blue nodes includes nodes from the*
753 *Cornwall (1), Devon (2) and Somerset (3) countries. (B) Partition P_6: three large communities*
754 *articulated around London.*

755 **Fig. 8.** Partition P_4 and the regional subdivision of Great Britain. *The colours represent the nine*
756 *communities identified by stability optimisation for t =1 and n_t = 5. The solid lines correspond to the*
757 *boundaries of the regions of Great Britain: 1.Scotland, 2.North East, 3.North West, 4.Yorkshire and*
758 *the Humber, 5.East Midlands, 6.West Midlands, 7.East of England, 8.London, 9.South East, 10.South*
759 *West, 11.Wales.*

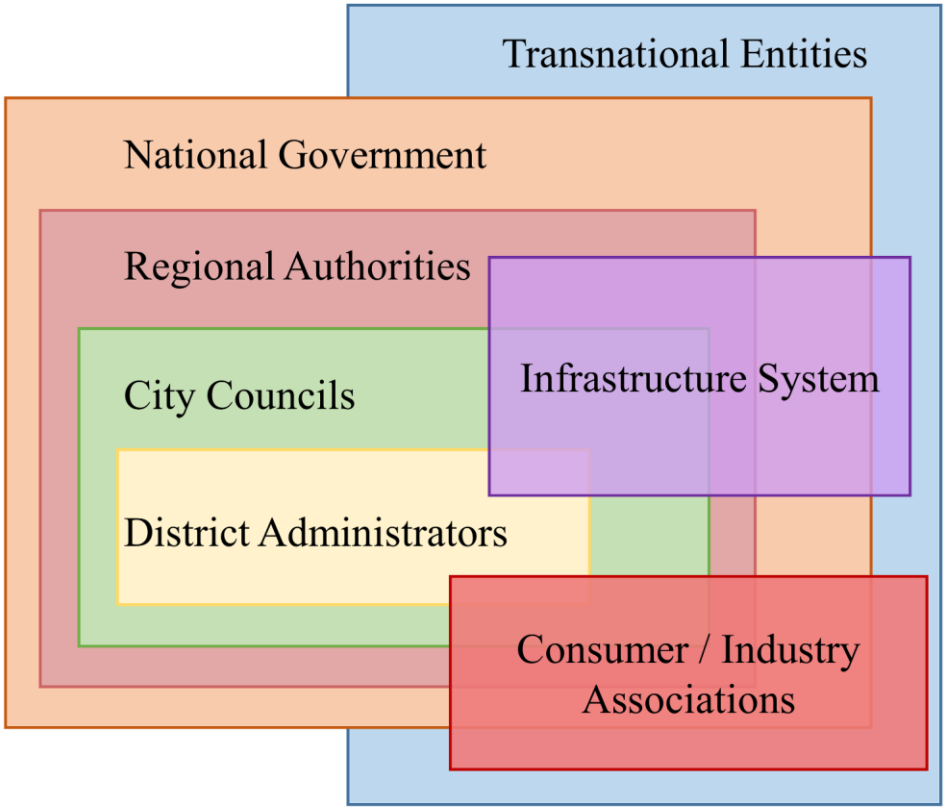
760 **Fig. 9.** The Community Efficiency matrix *CE* of partition P_4 of the railway network. *Each CE_{ij} element*
761 *was calculated using Equation 4: the diagonal elements (solid lines) represent the internal efficiency*
762 *of communities, whereas the off-diagonal terms (dashed lines) reflect the efficiency of transport*
763 *between different communities.*

764 **Fig. 10.** The distributions of *IC* and *EC* on the nodes of the network (nodes with the highest value in
765 each community labelled). *A) Distribution of intracommunity centrality IC on the system: nodes*
766 *located in the periphery of the system are able to reach high centrality values because of their local*
767 *importance. B) Distribution of intercommunity centrality EC: only nodes near the London community*
768 *C_8 achieve high centrality values.*

769 **Fig. 11.** The results of a *XC* analysis on the railway network for α = 0.5 and three stakeholders.
770 *Mapping of the nodes of the network in the (GC, XC) plane. The dash-dotted lines represent the*

771 *median of the distribution of the two variables, and were used to separate the nodes in four*
772 *quadrants. Q1 includes nodes critical in local and global assessments, Q2 nodes which are critical*
773 *only in local assessments, Q3 contains the peripheral nodes, while Q4 hosts those nodes that are*
774 *critical for global efficiency but not at the local level.*

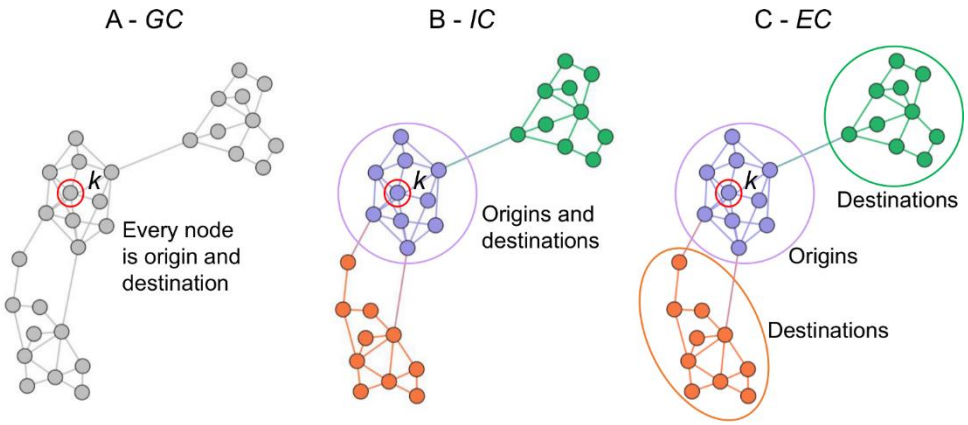
775



776

777 Fig 1

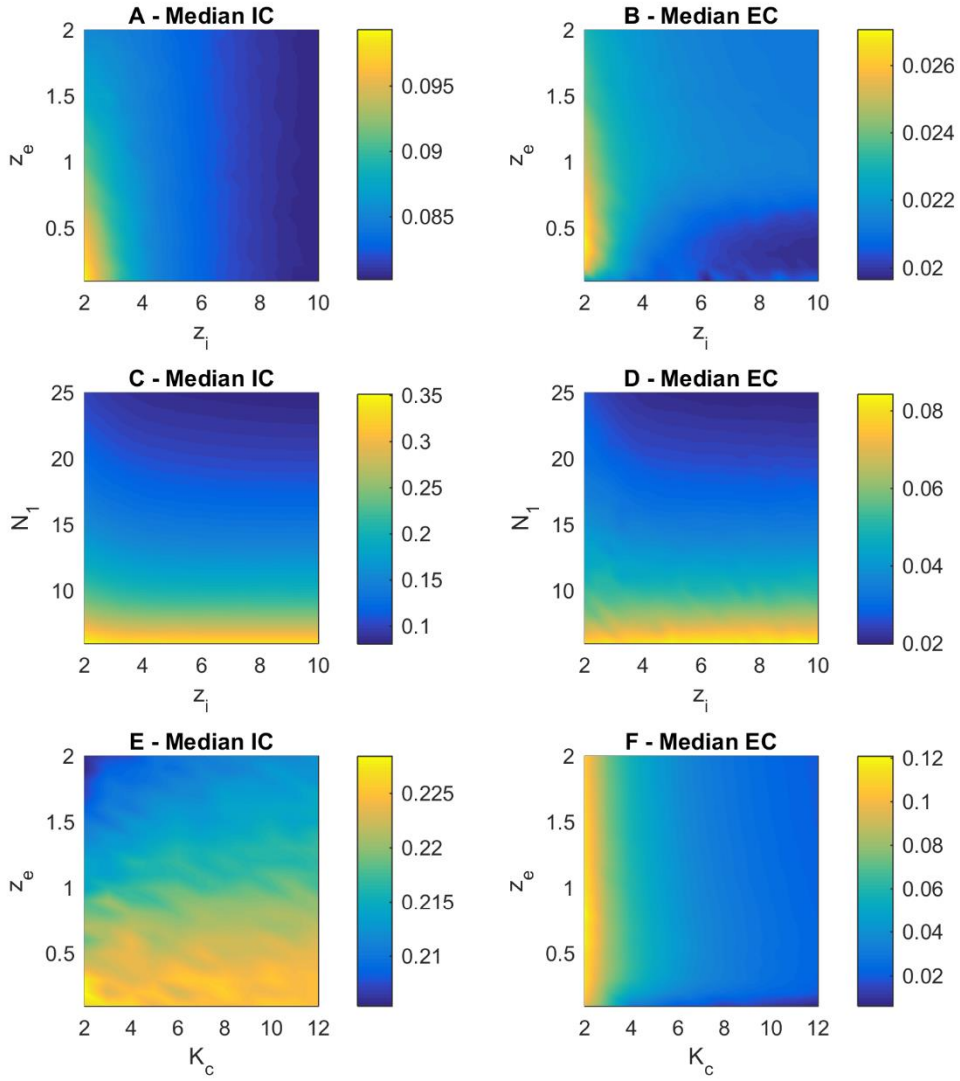
778



779

780 Fig 2

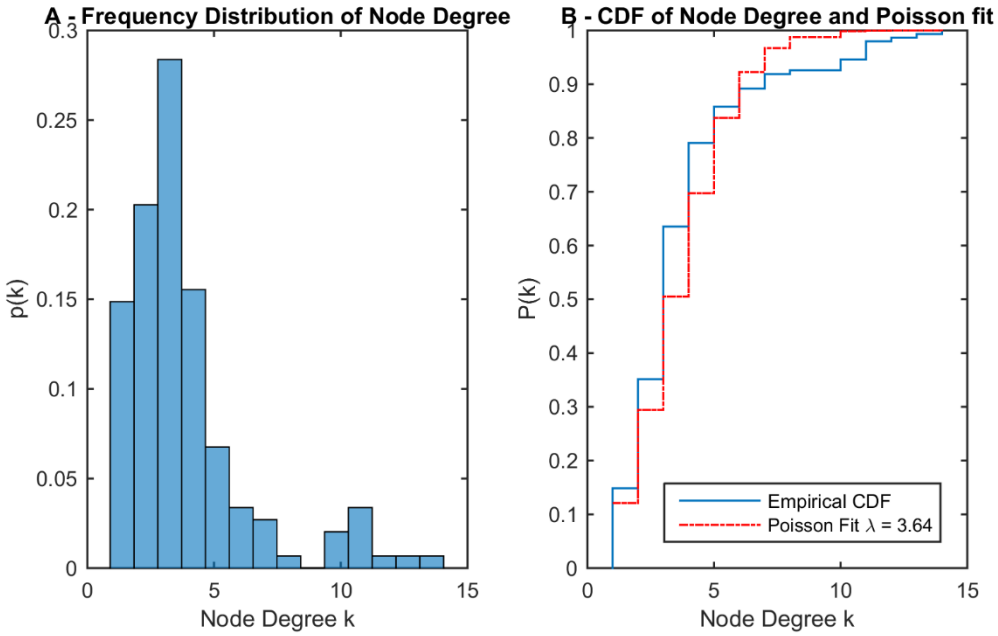
781



782

783 Fig 3

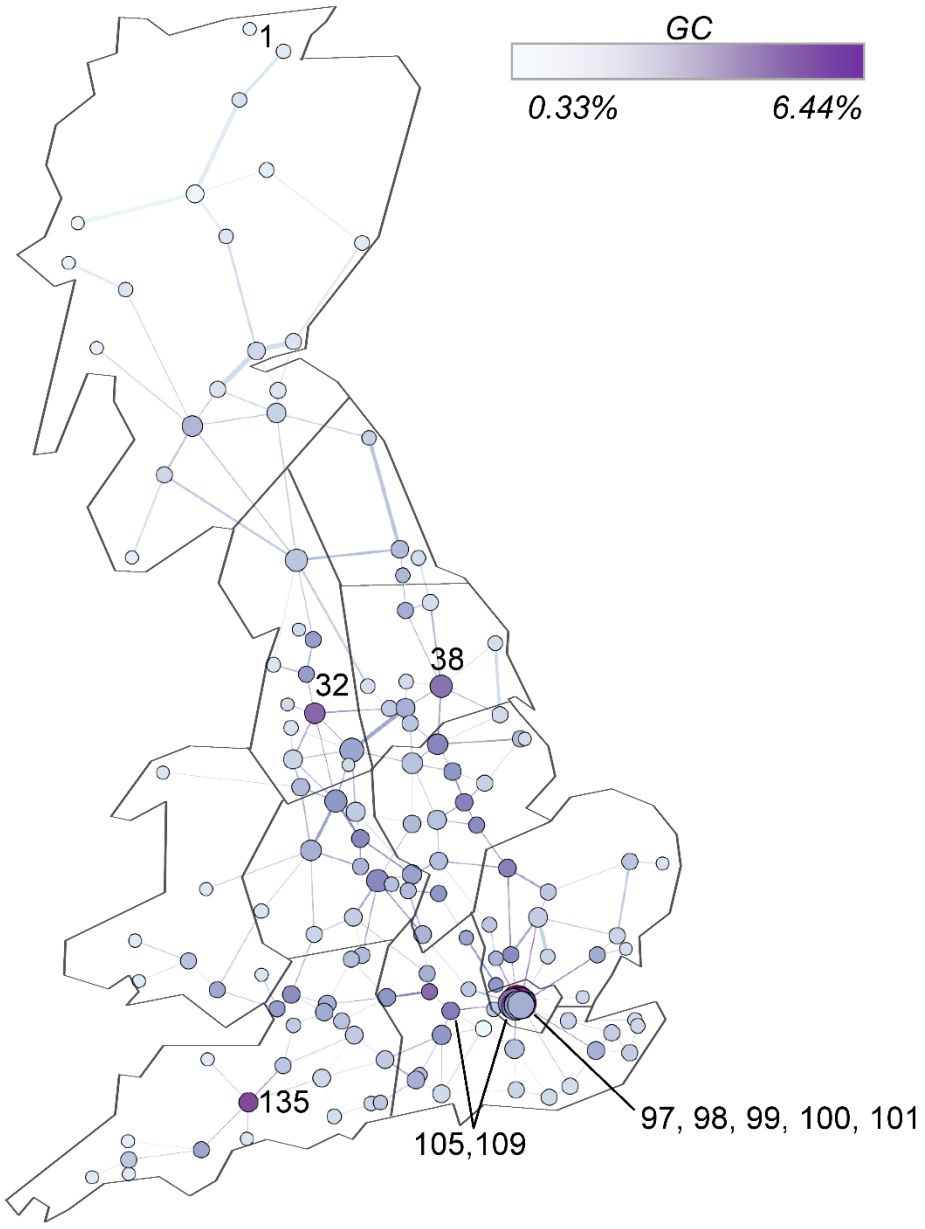
784



785

786 Fig 4

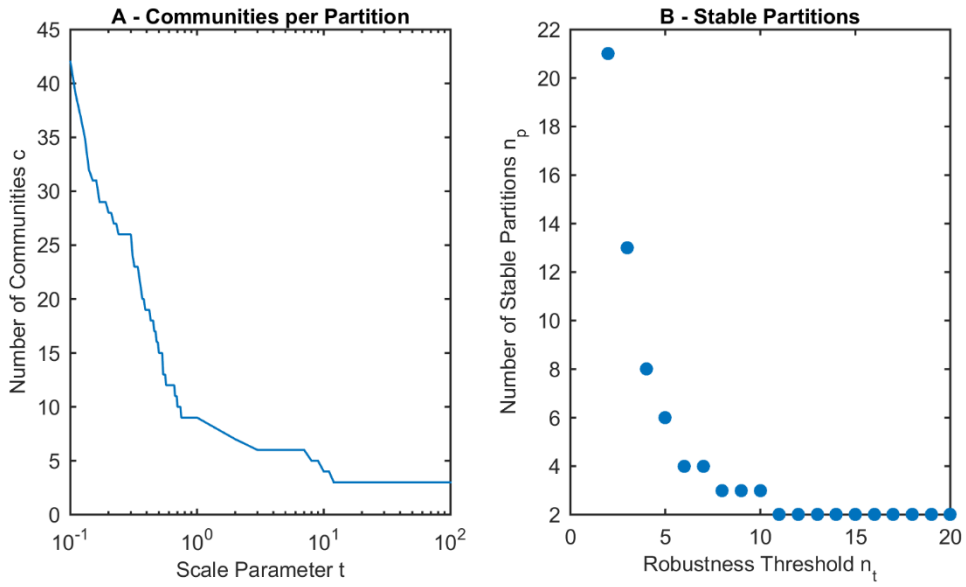
787



788

789 Fig 5

790



791

792 Fig 6

793

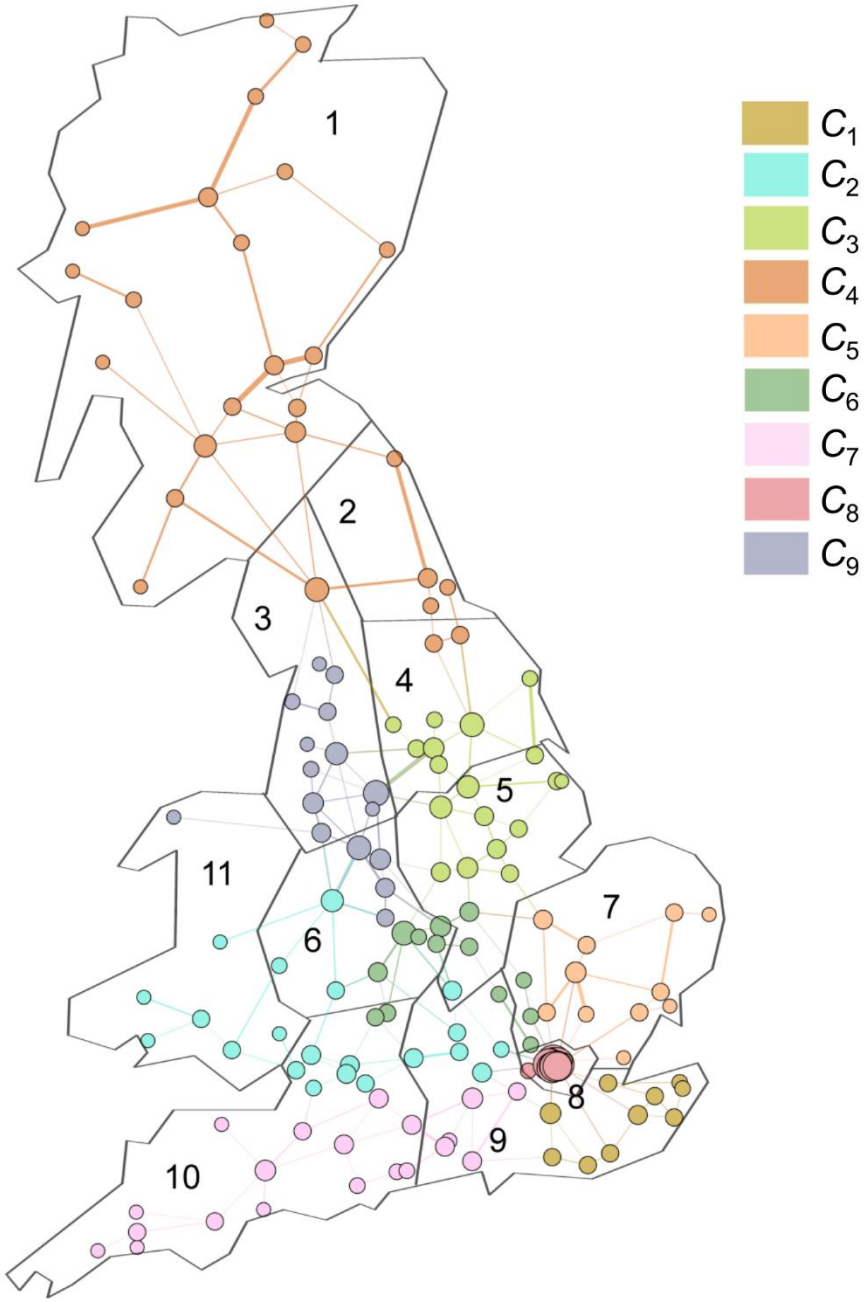
A – Detail of P_1

B – P_6

794

795 Fig 7

796



797

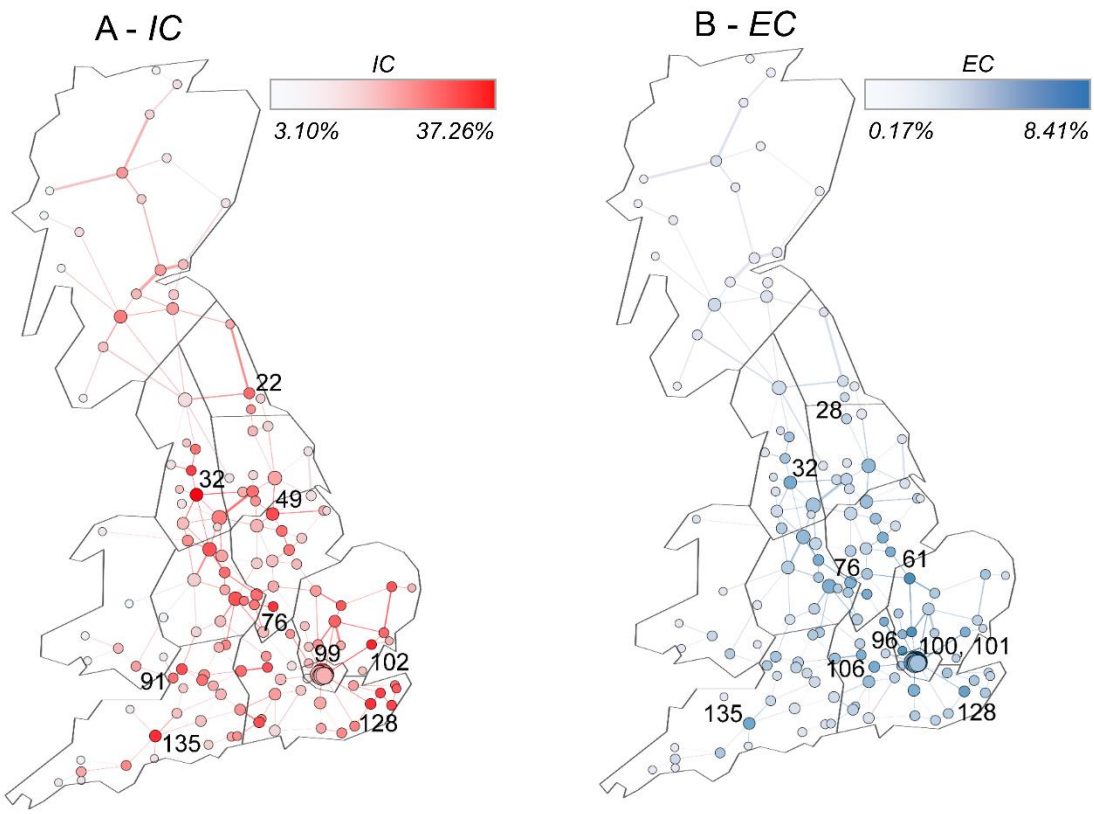
798 Fig 8

799

$$CE = \begin{bmatrix} 0.020 & & & & & & & & \\ 0.006 & 0.013 & & & & & & & \\ 0.005 & 0.005 & 0.019 & & & & & & \\ 0.002 & 0.003 & 0.005 & 0.007 & & & & & \\ 0.008 & 0.005 & 0.007 & 0.003 & 0.015 & & & & \\ 0.008 & 0.009 & 0.008 & 0.003 & 0.008 & 0.023 & & & \\ 0.005 & 0.007 & 0.004 & 0.002 & 0.005 & 0.006 & 0.011 & & \\ 0.014 & 0.009 & 0.007 & 0.003 & 0.015 & 0.015 & 0.008 & 0.079 & \\ 0.004 & 0.006 & 0.007 & 0.004 & 0.005 & 0.010 & 0.004 & 0.006 & 0.017 \end{bmatrix}$$

Sym.

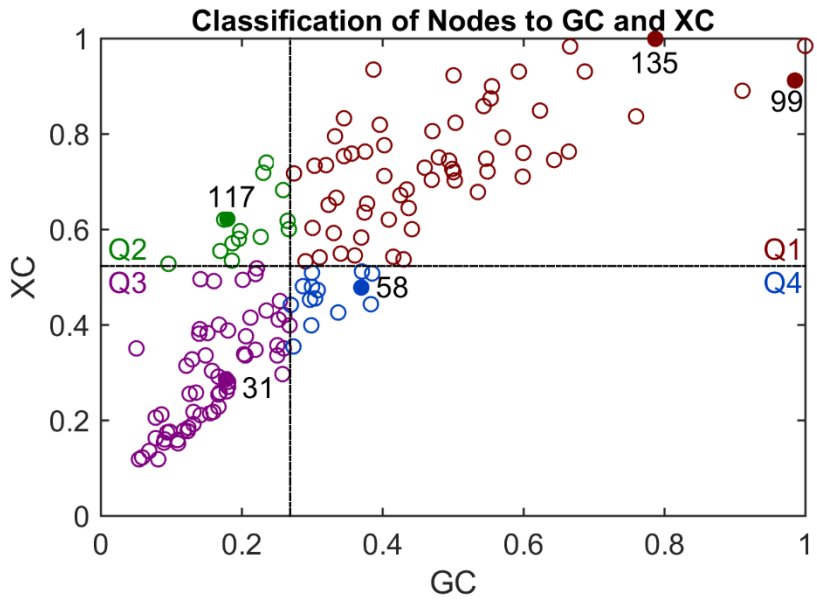
Fig 9



803

804 Fig 10

805



806

807 Fig 11

808

APPENDIX A: SUPPLEMENTARY MATERIAL

STABILITY OPTIMISATION

Stability Optimisation (Delvenne et al. 2010) is a Community Detection technique which leverages the relationship between graphs and Markov chains. This section provides a brief overview of the technique, but the reader is encouraged to refer to the original paper for a detailed explanation.

Every network can be associated a random walk, and thus a Markov chain, in which the states of the stochastic process are represented by the nodes, while edges describe the possible transitions between these states. The transition probabilities are proportional to the weight of the edges between the nodes, normalised by their sum of the weight of all the outgoing edges.

The transition matrix $\boldsymbol{M}$ of the Markov chain defined by this random walk is:

$$\boldsymbol{M} = \boldsymbol{D}^{-1}\boldsymbol{A}$$

where $\boldsymbol{D}$ is the diagonal matrix of node degrees, and $\boldsymbol{A}$ is the adjacency matrix of the network.

The Markov chain is hence described as:

$$\boldsymbol{p}_{t+1} = \boldsymbol{p}_t\boldsymbol{M}$$

where $\boldsymbol{p}_t$ is the normalised probability vector, expressing the likelihood of the random walker being in each node at step t of the process. Under these simple assumptions, the Markov chain is ergodic and reversible, with stationary distribution:

$$\boldsymbol{\pi} = \boldsymbol{d}/2\mathrm{m}$$

where $\boldsymbol{d}$ is the vector of node degrees, and m is the sum of the weight of all the edges.

This Markov Chain can be analysed in terms of transitions between communities, rather than between nodes. When a network with n nodes is partitioned in c communities, this partitioning can be encoded in the n x c indicator matrix $\boldsymbol{H}$, in which the elements of each columns are zero if the node belongs to the community associated to the column, and one otherwise.

The probability of transition from a community to another during the random walk can be then quantified using the clustered auto-covariance matrix of the network, expressed as:

$$\boldsymbol{R}_t = \boldsymbol{H}^T\left(\boldsymbol{P}\boldsymbol{M}^t - \boldsymbol{\pi}^T\boldsymbol{\pi}\right)\boldsymbol{H}$$

where $\boldsymbol{P}$ is the diagonal matrix of transition probabilities.

According to the standard definition of community used in network science (Newman 2010), its elements should be better connected to each other than to the rest of the system. Therefore, a partition of a network in communities should maximise the likelihood that the random walks which start in a community end up in the

same community. This is expressed by the trace of the clustered auto-covariance matrix, which is labelled stability:

$$r(t, \boldsymbol{H}) = min_{0 \leq s \leq t} trace(\boldsymbol{R_s})$$

and in order to ensure that the partition is indeed robust, Stability Optimisation prescribes that the minimum value of stability over all times up to *t* is taken as the value of its stability. The partitions achieving the highest levels of stability are those which encode the underlying community structure of the network.

By varying the length *t* of the random walk this method is able to explore the quality of partitions of different size: as the random walk lengthens, the clustered auto-covariance matrix changes and therefore different partitions will emerge as the most appropriate subdivisions of the network in communities.

The numerical implementation of the method used in this paper is the one provided in (Le Martelot & Hankin 2011), which utilises a greedy optimisation heuristic to identify the partitions which maximise stability: at the beginning of the process, each node is originally assigned to its own individual community, and then the stability of different configurations is then explored in order to converge to an approximation of the maximum.

Table A1. List of nodes and corresponding station names

Node ID	Station Name	Node ID	Station Name	Node ID	Station Name
1	Thurso	51	Cleethorpes	101	Stratford
2	Wick	52	Stattford	102	Colchester
3	Helmsdale	53	Sheffield	103	BristolParkway
4	KyleOfLochalsh	54	Retford	104	Swindon
5	Inverness	55	Newark	105	Didcot
6	Elgin	56	Lincoln	106	Reading
7	Mallaig	57	Aberystwyth	107	HeathrowAirport
8	Aviemore	58	Shrewsbury	108	Marylebone
9	Aberdeen	59	Derby	109	Paddington
10	FortWilliam	60	Nottingham	110	Victoria
11	Perth	61	Grantham	111	FenchurchStreet
12	Dundee	62	Wolverhampton	112	CharingCross
13	Oban	63	Birmingham	113	Waterloo
14	Stirling	64	Nuneaton	114	LondonBridge
15	Kirkcaldy	65	Leicester	115	Chatham
16	Glasgow	66	Peterborough	116	Margate
17	Edinburgh	67	Ely	117	Ramsgate
18	BerwickUponTweed	68	Norwich	118	SouthendAirport
19	Stranraer	69	GreatYarmouth	119	WestonSuperMare
20	PrestwickAirport	70	BirminghamIntl	120	BristolTempleMeads
21	Carlisle	71	Bedford	121	Bath
Node ID	**Station Name**	**Node ID**	**Station Name**	**Node ID**	**Station Name**

22	Newcastle	72	Cambridge	122	Westbury
23	Windermere	73	Ipswich	123	Basingstoke
24	Durham	74	Llandridod	124	Woking
25	Sunderland	75	Coventry	125	GatwickAirport
26	Barrow	76	Rugby	126	Canterbury
27	Oxenholme	77	LutonAirport	127	Dover
28	Darlington	78	Stevenage	128	Ashford
29	Middlesbrough	79	StanstedAirport	129	Barnstaple
30	Lancaster	80	Hereford	130	Taunton
31	Blackpool	81	Worcester	131	Yeovil
32	Preston	82	Banbury	132	Salisbury
33	Skipton	83	MiltonKeynes	133	SouthamptonAirport
34	Scarborough	84	Harwich	134	Newquay
35	Southport	85	Fishguard	135	Exeter
36	Bradford	86	PembrokeDock	136	Penzance
37	Harrogate	87	Carmarthen	137	Falmouth
38	York	88	Swansea	138	Truro
39	Liverpool	89	Cardiff	139	Plymouth
40	Manchester	90	Valleys	140	Paignton
41	Leeds	91	Newport	141	Weymouth
42	Hull	92	Chelthenham	142	Poole
43	ManchesterAirport	93	Gloucester	143	Bournemouth
44	Wakefield	94	Oxford	144	SouthamptonCentral
45	Holyhead	95	HighWycombe	145	Porthsmouth
46	Chester	96	Watford	146	Brighton
47	Crewe	97	LondonEuston	147	Eastbourne
48	StokeOnTrent	98	SaintPancras	148	Hastings
49	Doncaster	99	KingsCross		
50	Grimsby	100	LiverpoolStreet		

854

855