Preserving Key Topological and Structural Features in the Synthesis of Multilevel Electricity Networks for Modeling of Resilience and Risk

Scott Thacker[1]; Jim W. Hall[2]; and Raghav Pant[3]

Abstract: Given the often-limited availability of real electricity network data, this paper presents methodology for the synthesis of multilevel electricity networks for use in applied network failure and risk analysis. The proposed algorithm is capable of producing networks that preserve a number of important spatial and topological properties of real-world networks including the multilevel structure of subsystems, the geographic distribution of network nodes, the node degree distribution, and the networks spatial connectivity. The algorithm is capable of integrating both synthetic and real data from a range of sources to produce spatially and topologically continuous representations. The flexibility of the algorithm is demonstrated through the synthesis of a regional-scale electricity network. The practicality of the algorithm, in terms of providing new data to conduct applied risk and resilience studies of interdependent infrastructures, is demonstrated through the synthesis of a unique representation of the national integrated electricity network for England and Wales, bridging the transmission, subtransmission, and distribution scales and consisting of more than 160,000 nodes. **DOI: 10.1061/(ASCE)IS.1943-555X.0000404.**

Introduction

Infrastructure network systems play a critical role in modern societies by facilitating the distribution of resources and services across broad spatial extents at a range of scales. Electricity networks form a central part of the national infrastructure, with all other infrastructure sectors having a dependence on electricity, and electricity being dependent on many other infrastructures sectors (Rinaldi et al. 2001). This complex and interdependent arrangement creates a systemic vulnerability in which localized failures can cascade, resulting in large-scale and spatially widespread consequences. Such vulnerability has been highlighted multiple times over the last two decades including the 1998 Auckland blackout (Davis 1999), the 2003 Northeast blackout in the United States (U.S.–Canada Power System Outage Task Force 2004), and the disruptions associated with the 2007 U.K. summer floods (Pitt 2008). In the future, interdependent coupling between electricity power networks and other infrastructure network systems is set to increase, creating new potential vulnerabilities but also opportunities for risk reduction (PNNL 2015).

In recent years, tools and methods from the study of complex networks revealed significant insights into the risk and resilience of current and future infrastructure systems (e.g., Lambert and Sarda 2005; Lambert et al. 2005; Dueñas-Osorio et al. 2007; Johansson and Hassel 2010; Thacker et al. 2014; Ganin et al. 2016; Gao et al. 2016). Central to these studies are network models (consisting of nodes and edges) that are assembled to intuitively represent the composition and connectivity of real infrastructure systems. The spatial embedding of these networks allows not only for network assets to be intersected with known hazard data but also to be mapped to end users of the infrastructure services—the latter enabling the consequences of failures to be more comprehensively understood. The connectivity that is encoded inherently within infrastructure network models facilitates the use of techniques to estimate failure propagation and cascade.

Despite these and other advances, many potential applied studies are hindered by the lack of available data of both the spatial and topological components of electricity networks (Ouyang 2014). Reasons for the lack of available data include the following: (1) asset management systems have not collected all the salient data; (2) the multiple ownership of infrastructure systems; (3) security concerns resulting from the release of systems data; and (4) concerns over the release of potentially sensitive customer information.

Although a selection of data on real electricity networks is available to the research community, it often has limited spatial coverage or inconsistent topological information. This can be problematic when constructing systemwide models of interdependent infrastructures that typically operate and interact with failure-inducing hazards over large spatial extents at a range of scales (Thacker et al. 2017; Moini 2016). Such limitations can result in underestimations of failure consequence, and hence, risks. In most situations, the lack of data prohibits analysis or at least curtails it. Where these data are not available, one solution is to produce a realistic synthetic representation of the network system. Although synthetic networks will be different from reality on the ground, if the salient spatial and topological properties of the network can be preserved, they provide a worthwhile basis for infrastructure network analysis that would otherwise be impossible.

A number of models exist for the synthesis of networks, including the random graph (Erdös and Rényi 1959), small-world graph (Watts and Strogatz 1998), and scale-free graph (Albert and Barabaśi 1999). These models produce networks with characteristic statistical properties and network metrics that describe their

[1]Postdoctoral Researcher, Environmental Change Institute, Univ. of Oxford, South Parks Rd., Oxford OX1 3QY, U.K. (corresponding author). E-mail: scott.thacker@ouce.ox.ac.uk

[2]Professor, Environmental Change Institute, Univ. of Oxford, South Parks Rd., Oxford OX1 3QY, U.K.

[3]Postdoctoral Researcher, Environmental Change Institute, Univ. of Oxford, South Parks Rd., Oxford OX1 3QY, U.K.

Note. This manuscript was submitted on July 17, 2015; approved on June 12, 2017; published online on November 22, 2017. Discussion period open until April 22, 2018; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

structure. Understanding that the network structure is highly correlated to network function (Newman 2003), it is important that any synthetic representation preserves a number of these topological and spatially important characteristics. Multiple metrics exist and have been used to describe real-world network properties (Newman 2010; Barthelemy 2011). Of particular interest to the infrastructure community is the risk of network failure; one such property used to characterize failure is degree distribution (Albert et al. 2000). Building on this theme, LaViolette et al. (2006) and LaRocca and Guikema (2015) characterize the robustness of synthetic networks that have been generated to maintain a power-law degree distribution. Methods exist for the production of synthetic networks with predefined degree distributions, for example, the configuration model (Bender and Canfield 1978) or Newman's random graph with arbitrary degree distributions (Newman 2003). Wang et al. (2010) propose a similar method, but for the purpose of producing synthetic electricity networks, however, as with the work of LaViolette et al. (2006), LaRocca and Guikema (2015), Bender and Canfield (1978), and Newman (2003), the model does not maintain any of the important spatial properties of real-world electricity networks. Spatial variants of the range of generative methods exist (e.g., Aldous and Shun 2010; Ferretti and Cortelezzi 2011); these methods use distance as a means by which to constrain network growth in space during a preferential attachment process. The methods produce networks that preserve a specified degree distribution but are not flexible enough to accommodate known spatial information on the actual or likely location of nodes. Hines et al. (2015) use a similar bottom-up growth process to produce optimized synthetic infrastructure networks; this work differs, however, in that node location is specified and degree distribution (network connectivity) is unspecified.

Infrastructure network systems can be intuitively represented as multilevel networks (Kivelä et al. 2014; D'Agostino and Scala 2014). Such a representation can be useful, not only in mapping the multiple owners or operators of a particular system, but to characterize subsystems that form a part of the whole infrastructure that have distinct characteristics. Take, for example, electricity transmission networks, which are generally considered to have a mesh structure, whereas distribution networks are considered to be radial (Buchholz and Styczynski 2014). Developing a multilevel representation allows these distinct network structures to be preserved and integrated to form one continuous representation—bridging multiple operation scales while preserving the heterogeneous spatial and topological characteristics of the individual networks. An additional advantage on encoding this multilevel structure is that it facilitates a more explicit mapping of network interdependencies; for example, assets from different infrastructure sectors (e.g., railway stations, water pumps, and telecommunication masts) connect to different voltage levels in electricity transmission and distribution networks (Pant et al. 2016).

This study develops a novel approach for the synthesis of multilevel electricity networks that preserve key structural and topological features to enable risk and resilience studies. The study addresses gaps in the literature by producing an algorithm that is capable of encoding synthetic or real data on the multilevel structure of the system, including the location of node assets, the network degree distribution, and the networks spatial connectivity. The modular nature of the algorithm allows for differing levels of known systems data to be incorporated into the synthetic network. This can be used to derive parts of or complete networks that best represent actual networked systems. Although data (multilevel structure, census population, network degree distribution, nearest-neighbor rank distribution) is required to drive the algorithm, similarities observed between the spatial and topological properties of real electricity network systems (Barthelemy 2011; Cotilla-Sanchez et al. 2012) support that when absent, this data can be substituted with known data from similar systems. Data that encode properties not observed in real-world systems can also be used. Although such an application can have merits (e.g., designing future electricity networks), it should be considered that as less real data are used, the synthetic networks produced become further detached from those observed in practice. In addition to generating individual networks, the algorithm is capable of generating families of realistic networks that can form a useful test bed for community-based developments of models to characterize the risk and resilience of interdependent infrastructure systems.

A mathematical formalization of multilevel infrastructure networks is developed, which then leads on to a characterization of the most salient spatial and topological properties of multilevel electricity infrastructure network systems. A synthesis algorithm, and details for its implementation, is then introduced. Next, applications of the algorithm for regional and national scales is outlined. Finally, a discussion of the results, including conclusions and insights gained from the study, are offered.

Multilevel Infrastructure Network Formalization

A mathematical formalization for multilevel infrastructure networks is outlined that introduces the formal notation that is later used to characterize these systems and develop the synthesis algorithm. It also acts as an exemplar as to the complete set of data and attribution that is required to build any systemwide infrastructure representation. Notation is mapped throughout to the engineering properties of real electricity power systems.

The infrastructure system S is described as a set of interconnected assets whose collective function is to facilitate the production and transfer of a resource or service to customers, such as electricity infrastructures. Many infrastructure systems are comprised of multiple interacting subsystems, in which an individual subsystem S_k is defined as fulfilling a specific function within the system and can be characterized as having unique attributes. Within electricity networks, subsystems are typically classified as the transmission, subtransmission, and distribution network infrastructure. In England and Wales, the classification is based on the subsystems' attributed operational voltage, typically 400, 275, 132, 33, and 11 kV. The system S is comprised of $|S|$ number of individual subsystems which are denoted as $S = \{S_1, \ldots, S_{|S|}\}$. Infrastructure systems that contain multiple subsystems are classified as multilevel, with individual subsystems forming levels (nested) in the multilevel structure. This multiscale, property is seen in multiple infrastructures that have evolved to efficiently distribute resources across space (Barthelemy 2011).

Within this paper, infrastructure systems are represented as networks consisting of nodes and edges. Considering electricity infrastructure systems as networks, edges are used to represent overhead lines and underground cables that transmit flow between nodes. Nodes can be classified into different types. Source nodes are where the flow is generated (e.g., power stations), sink nodes are where the flow is consumed (e.g., demand hubs), and intermediate nodes are where flow is transformed (e.g., substations, through electrical transformers).

The nodes that are associated with a particular subsystem are sampled from the exhaustive set of all nodes $N = \{n_1, \ldots, n_{|N|}\}$ in the system, where $|N|$ denotes the number of nodes. The node set N_k of the subsystem S_k is defined as the subset of the Cartesian product set of N and S_k, as given in Eq. (1)

$$N_k \subseteq N \times S_k = \{(n_1, S_k), \ldots, (n_{|N|}, S_k)\} \tag{1}$$

Nodes that are present in two or more subsystems are described as having an attribute plurality. Electricity substations perform this function in electricity networks—stepping down or stepping up the voltage between subsystem levels—establishing the multiscale structure. This is described in Eq. (2), where node n_i is assigned the set $\tilde{S}$ of all subsystems where it is present

$$n_i \equiv (n_i, \tilde{S}) \tag{2}$$

The connectivity of the nodes within and across different subsystems is defined in terms of the edge set E, which is denoted as

$$E = \{e_{i\alpha,j\beta} = [(n_i, S_\alpha), (n_j, S_\beta)], \quad \forall\ i, j, \alpha, \beta\} \tag{3}$$

All edges in the present system are directed edges, i.e., $e_{i\alpha,j\beta} \neq e_{j\beta,i\alpha}$. From the preceding set definition, the following two distinct edge types are identified:

- Intrasectorial edges E_k, which are given as $E_k = \{e_{ik,jk} = [(n_i, S_k), (n_j, S_k)]; \forall i, j, k, i \neq j\}$, where it is assumed that within the subsystems, nodes do not have self-edges incident on themselves. In electricity infrastructures, these edges are overhead lines or underground cables.
- Bridging edges B, which are given as $B = \{e_{i\alpha,i\beta} = [(n_i, S_\alpha), (n_i, S_\beta)]; \forall i, \alpha, \beta, \alpha \neq \beta\}$, establishing the connectivity of nodes between subsystem levels. In electricity infrastructure systems, these edges are the bridge between the two halves of a transformer.

The edge types not belonging to the preceding two sets form the set of intersectorial or interdependency edges.

Infrastructure systems are common entities in the landscape with nodes and edges within S having a specific location given in Cartesian coordinates, derived by the location function $l(S)$

$$l(S){:}S \to \mathbb{R}^2 \tag{4}$$

Fig. 1 provides a simplified network example to demonstrate the multilevel infrastructure network formalization.

Spatial Topological Infrastructure Network Characterization

Representations of real infrastructure systems, such as those previously introduced, provide a means to understand the behavior of infrastructure systems for a range of applied analyses. The aim is to provide methodology in the form of an algorithm that is capable of producing realistic synthetic representations of the network system being investigated. This section describes the set of spatial and topological characteristics that the synthetic representation of the real infrastructure should preserve. In encoding these properties in the synthetic network, it is argued that the synthetic networks produced by the algorithm will provide a worthwhile basis for infrastructure network analysis that would otherwise be impossible.

Four requirements for synthetic networks are identified as follows: (1) to preserve the number of node assets distributed within each level of the multilevel structure; (2) to preserve the spatial distribution of these nodes; (3) to preserve global network connectivity for different levels; and (4) to preserve the local spatial connectivity between node assets for different levels.

Definitions of the requirements and examples of their use through a characterization of the spatial and topological properties of a real-world electricity network system (the electricity network of England and Wales) are provided subsequently.

Multilevel Assignment of Nodes

The first requirement is to ensure that the correct number of nodes $|N_k|$ is assigned to each level of the multilevel structure. This characteristic follows the understanding that in the upper levels of the system (that are responsible for the bulk transmission of the service or resource), there are relatively fewer nodes with high capacities compared to assets in lower levels that are greater in number but have lower capacities. This is demonstrated with data (ENA 2009) that is provided for the number of nodes (substations) for given levels of the integrated electricity network system in England and Wales (Table 1).

As recognized earlier in this paper, substation assets function to step up or down the voltage between separate network levels. These assets therefore have a plurality and are formally recognized in two or more levels of the system, connected by a bridging edge. Considered locally, a sink node on the upper side of the transformer is therefore regarded as a source node on the lower side of the transformer. It is this node property that makes it possible to build a continuous representation that bridges operational scales.

Spatial Distribution of Nodes

Not only is it important to preserve the number of nodes, it is also important to preserve their spatial distribution. The spatial distribution of node assets is characterized next.

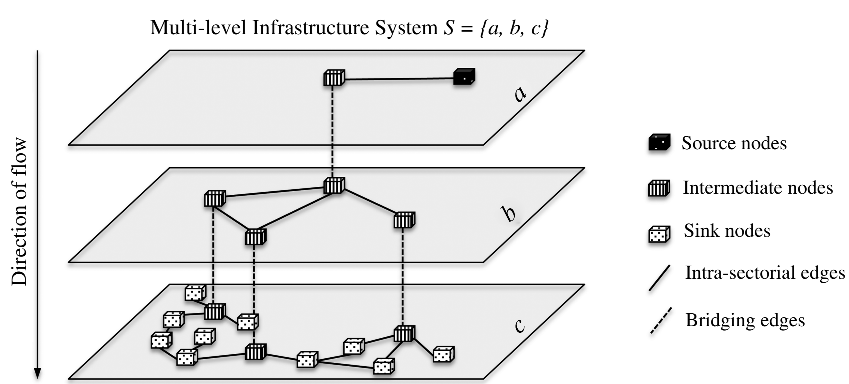

Fig. 1. Multilevel infrastructure system S decomposed into the subsystems: a, b, and c represent individual levels; flows of resources and services are facilitated from source to sink nodes within subsystems by intrasectorial edges and between levels (subsystems) by bridging edges

Table 1. Details for Substation (Node) Assets in England and Wales for Four Types of Substations That Correspond to Four Levels in the Integrated Electricity Network System

Substation (node) type	Typical voltage transformation levels	Approximate number	Typical number of customers supplied
Grid (transmission)	400–132 kV	377	200,000/500,000
Bulk (subtransmission)	132–33 kV	1,000	50,000/125,000
Primary (subtransmission)	33–11 kV	4,800	5,000/30,000
Distribution	11–415 V	160,000	1/500

Consider $Q \subset \mathbb{R}^2$ the set of all two-dimensional coordinates of locations within the boundaries of a spatial extent such as the countries of England and Wales. Q can be partitioned into r smaller disjoint regions [regional boundaries, such as local authority district regions (LAD)], $Q_i \subset Q{:}i = 1, \ldots, r$. Each boundary area can be associated with a population count, such as those derived during a national census. The set of Cartesian coordinates $P \subset \mathbb{R}^2$ are the locations of residences of all the national populace. The notation $|P|$ is used to represent the national populace count, which is the total number of people in the residences within the coordinates of P. The intersection of P and each subregion Q_i is equal to the set of people living in one region, so the population count $|P_i|$ in Q_i is

$$|P_i| = |P \cap Q_i| \quad i = 1, \ldots, n \tag{5}$$

The number of nodes from the network system S located within the regional area Q_i is calculated as

$$|\tilde{N}| = |l(N) \cap Q_i| \tag{6}$$

where $l(N)$ denotes the locations of all nodes in the set N. Fig. 2 plots the number of electricity assets (substations and pole mounted transformers) intersecting LAD regional population boundaries for England and Wales. The plot highlights that there is a strong correlation ($R^2 = 0.78444$) between the population of the region and the number of nodes that reside in that area.

Global Network Connectivity

The next characteristic is the preservation of global network connectivity for different subsystem levels. Connectivity is established through edges, but can be viewed as a node characteristic, in which the degree δ of a node is the number of connections that the node has to other nodes. The degree distribution $\mathbb{P}(\delta)$ of a network is defined to be the fraction of nodes in the network with degree δ. Therefore, in a network with $|N|$ number of nodes, and $|N(\delta)|$ of them have degree δ, the degree distribution can be given by $\mathbb{P}(\delta) = |N(\delta)|/|N|$.

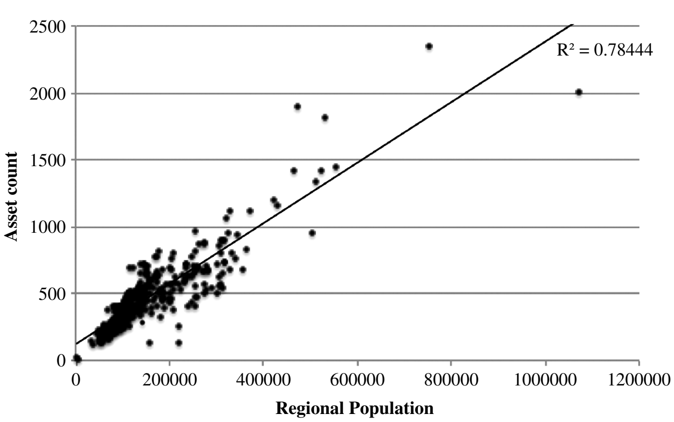

Fig. 2. Number of electricity node assets (substations and pole mounted transformers) intersecting LAD regional population boundaries in England and Wales

Recognizing that different subsystem levels can have different network topologies (Buchholz and Styczynski 2014), a further disaggregation can be made for the degree distribution for either source and sink nodes in any separate level S_k. Fig. 3(a) shows the degree distribution for the electricity transmission network of England and Wales disaggregated for both source and sink nodes. The plot shows that there are more sink nodes in each level than source nodes, reflecting the increase in the number of nodes at lower levels of the multilevel structure and difference in their respective probability distributions.

Local Network Connectivity

Recognizing the importance of local spatial connectivity between node assets for different subsystem levels, the metric of neighbor rank connectivity is proposed and derived next.

For any node n_i, the distance to all other network nodes is computed to derive the ordered set

$$D_E(n_i) = \{d_E(n_i, 1), d_E(n_i, 2), \ldots, d_E(n_i, t)\} \tag{7}$$

where $d_E(n_i, j)$ = Euclidian distance between node n_i and its jth closest neighbor. If the node pair (n_i, n_j) corresponding to the distance $d_E(n_i, j)$ are actually connected, then this connectivity is represented as an edge e_{ij}. For each node in the network, an edge set is constructed as the ordered set of nearest neighbor distances, which results in the rank nearest neighbor edge set

$$C(n_i) = \{1_j e_{ij}\}{:}1_j = \begin{cases} 1{:}\exists j \\ 0{:}\nexists j \end{cases} \tag{8}$$

The nearest neighbor connectivity within the network is established by constructing the preceding set $C(n_i)$ for all nodes in the network. Derived from this, the neighbor-rank distribution $\mathbb{P}(\theta)$ of a network is defined to be the fraction of edges in the network with neighbor rank θ. Therefore, in a network with $|E|$ number of edges, and $|E(\theta)|$ of them have neighbor rank θ, the neighbor-rank distribution is given as $\mathbb{P}(\theta) = |E(\theta)|/|E|$.

Fig. 3(b) presents the neighbor-rank distribution of the electricity transmission network for England and Wales. The graph highlights that although most assets intuitively connect to their nearest neighbors in space, some connect to assets that are further away, providing some level of long-range connectivity in the network.

Synthesis Algorithm

Algorithm 1 was developed to produce synthetic representations of multilevel electricity network systems. The algorithm iterates through successive levels of the multilevel structure. When missing

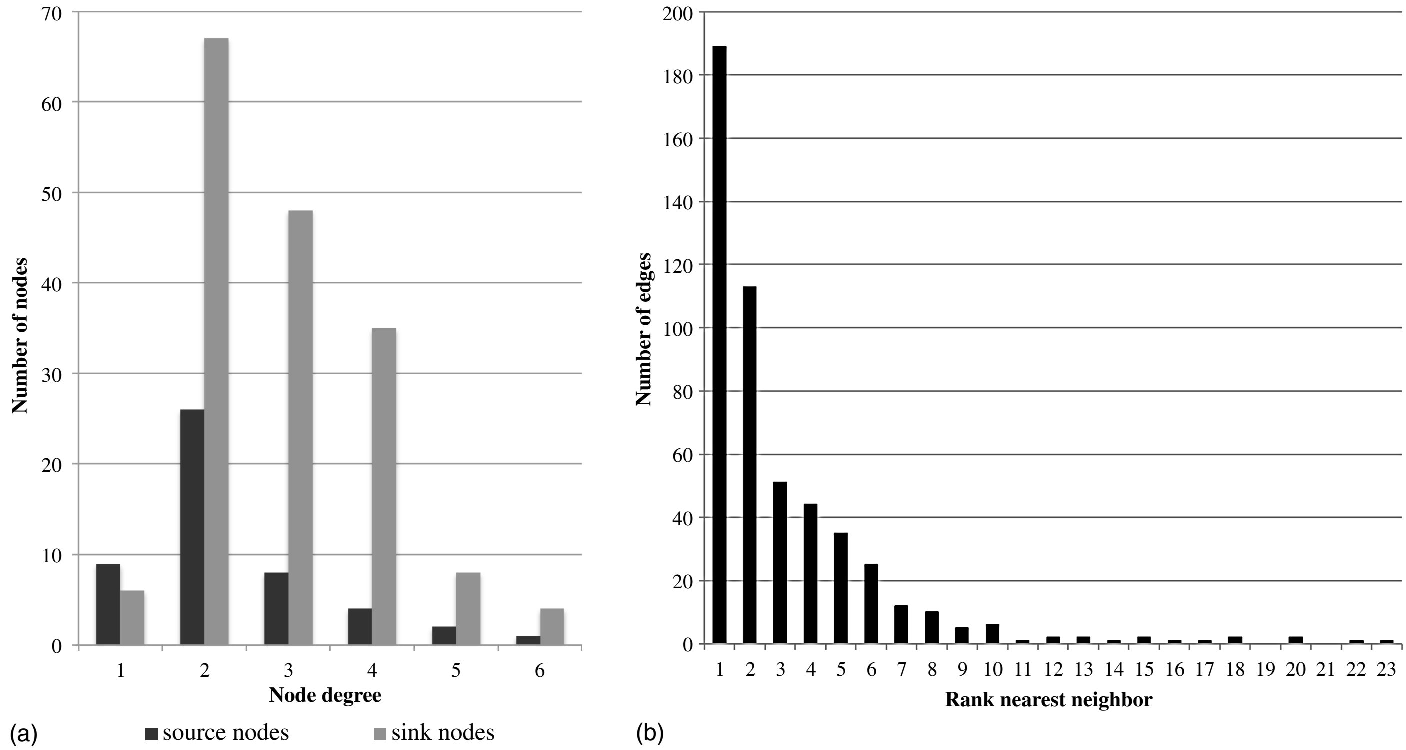

Fig. 3. Global connectivity measures for the electricity transmission networks of England and Wales: (a) network degree distribution with the degree of sink and source nodes being presented separately; (b) neighbor-rank distribution of edges within the network

spatial or topological information is identified, subroutines are implemented to produce synthetic representations of the missing elements that maintain the minimal spatial and topological requirements established in the previous section of this paper. This results in a final network representation that forms a continuous representation over the specified spatial extent, at a range of declared operational scales. The primary steps in the algorithm are outlined next, with Parts 1, 2.2, and 2.4 in bold because of their role in explicitly incorporating the required synthetic characteristics identified earlier in this paper:

Algorithm 1. Algorithm for the Production of Synthetic Representations of Multilevel Electricity Network Systems.

1. Given: Collect all available node N and edge E data and assign to subsystems depending on their attribution to give $S = \{S_1, \ldots, S_{|N|}\}$
2. For the subsystem S_k in the system:
  2.1. Does the full node set N_k exist?
    2.1.1. Yes: Go to step 2.3
    2.1.2. No: Go to step 2.2
  2.2. Create missing nodes and distribute in space:
    2.2.1. If sink nodes from subsystem S_{k-1} exist, then assign as source nodes in S_k (bridging node)
    2.2.2. For each regional area Q_i:
      2.2.2.1. Calculate population $|P_i| = |P \cap Q_i|$ of that area
      2.2.2.2. Calculate the number of nodes of correct type required in that area $|\tilde{N}| = |l(N) \cap Q_i|$ (taking into account existing source node allocations)
      2.2.2.3. Assign $|\tilde{N}|$ nodes to locations in Q_i
    2.2.3. Go to step 2.3
  2.3. Does edge set E_k exist?
    2.3.1. Yes: Go to step 2.5
    2.3.2. No: Go to step 2.4
  2.4. Assign edges to node:
    2.4.1. For all nodes with missing edges:
      2.4.1.1. Assign degree to node of specific type by sampling known degree distribution $\mathbb{P}(\delta)$
    2.4.2. Assign edges to nodes by sampling rank-neighbor distribution $\mathbb{P}(\theta)$ to match expected degree distribution:
    2.4.3. Go to step 2.5
  2.5. Subsystem S_k representation complete: If $S_k = S_{|N|}$ Go to step 3, else return to step 2 to evaluate S_{k+1}
3. End—System S network representation complete

Implementation of Algorithm 1 is heavily dependent on the raw data that may be available. In most cases, it will be beneficial to include as much data from the actual system as possible, therefore, only synthesizing parts of the system that have data missing from them. By reducing the constraint of real data, other generic distributions of node locations and connectivity can be used to build abstract representations that can be considered as alternative configurations. For example, one could specify the location of nodes based on future projected maps of population growth or land-use change data, or may incorporate new distribution network topologies in alignment with changes expected to occur through the increased uptake of distributed generation and electric vehicles.

Fig. 4 provides an example of how the synthesis algorithm would be implemented to produce a subsystem c for the infrastructure network system $S = \{a, b, c\}$ introduced earlier in the paper. Having collected all known data and assigned them to levels in the multilevel system, an iteration through the system is initiated, beginning at subsystem a to find that all data is complete; the next subsystem is b, which is also complete. The next subsystem is c, and this is incomplete and therefore the focus of the synthesis algorithm implementation. The process of iteration is represented in

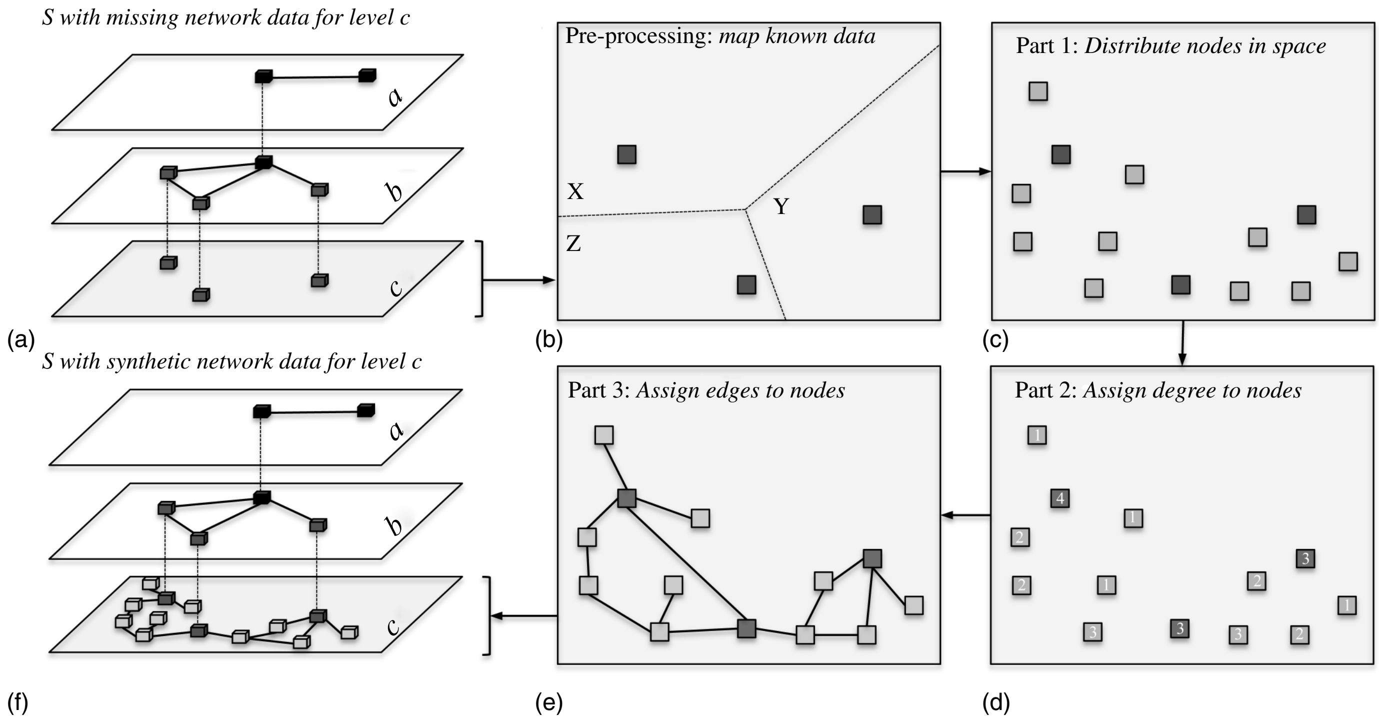

Fig. 4. (a) Representation of infrastructure system S with network node and edge information missing for level c; (b) preprocessing stage where known information on the position of nodes and regional population data is mapped; (c) Part 2.2 of the algorithm, where the missing distribution network nodes are distributed in space; (d) Part 2.4.1.1 of the algorithm, where a degree value is assigned to each node; (e) Part 2.4.3 of the algorithm, where edges are added between nodes to match predefined degree values; (f) the resultant continuous representation of S with real and synthetic data integrated

Fig. 4(a). Fig. 4(b) highlights the data that is available for that level; this is the source nodes (sink nodes from subsystem b). Next, the sink nodes that are identified as missing from level c are distributed in space as given in Part 2.2 of the algorithm [Fig. 4(c)]. The next stage of the algorithm ascertains that no edges are present. A degree is therefore assigned to each node based on sampling known degree distributions for nodes that belong to level c. Finally, edges are assigned in an iterative process to meet the expected degree distribution [Figs. 4(d and e)] and completes Part 2.4 of the algorithm. The algorithm terminates as all nodes and edges are present for the infrastructure network system s [Fig. 4(f)] and given as Step 5 of the algorithm.

Application of the Algorithm

Two practical applications of Algorithm 1 are presented at different scales: (1) Application of the algorithm for a regional electricity network, which highlights the ability of the algorithm to produce synthetic network representations that preserve the most salient properties of the desired network for different data availability levels; and (2) building the integrated electricity network of England and Wales and demonstrating the usefulness of the algorithm for building a continuous systems-level representation that enables a number of studies on interdependent infrastructure network risk and resilience.

Application of the Algorithm for a Regional Electricity Network

The following application of the algorithm centers on its ability to reproduce the encoded characteristics set out previously in this paper. This is performed using a regional electricity network from the northwest of England (ENW 2010). The network operates at 132 kV, distributing electricity between local sources (275 kV/132 kV substations) and local sinks (132 kV/33 kV substations). The network contains 83 nodes (15 sources and 68 sinks) that are located in different land-use types, from the urban city of Manchester to the rural Cumbrian hills. Fig. 5(a) gives a fully topographic representation of the network, and Fig. 5(b) provides a spatial network representation, showing edge geometries as straight lines between nodes.

Constraining the algorithm with differing levels of the actual systems data tests the sensitivity of each stage of the algorithm. Stages of the algorithm correspond to the characteristics outlined previously:

1. Multilevel structure that requires the algorithm to preserve the number of node assets distributed within each level of the system;
2. Node location that requires the algorithm to preserve the spatial distribution of these nodes;
3. Node degree that requires the algorithm to preserve global network connectivity for different levels; and
4. Edge connectivity that requires the algorithm to preserve the local spatial connectivity between node assets for different levels.

Table 2 provides an overview of the testing framework, highlighting the data constraints that result in four different realizations of the network. The table shows that the real network contains all the data and therefore is completely constrained. Types A, B, and C all preserve the first constraint multilevel structure, because within each representation, only a single synthetic level of the system (132 kV level) is being produced and hence it is not required to be implemented for multiple levels in the system. Types A, B,

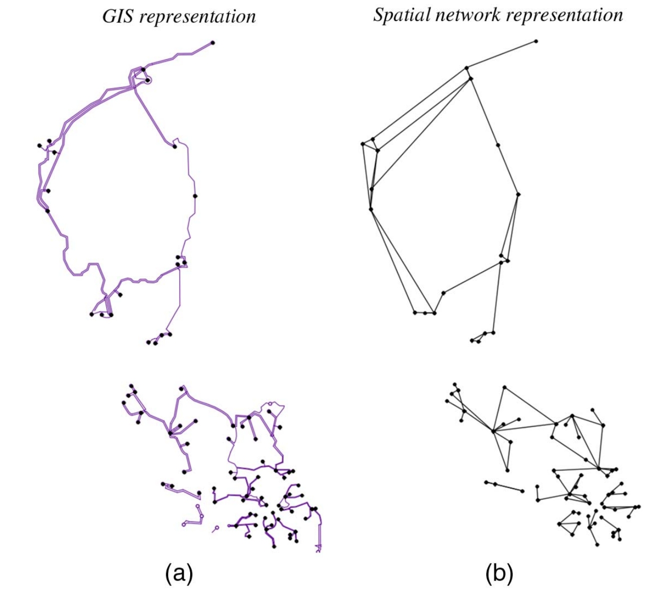

Fig. 5. Regional electricity network used for the application of the network synthesis algorithm: (a) GIS representation highlights the geographic spatial locations of nodes and edges that make up the real network; (b) spatial network representation shows the network using real node locations, connected by simplified shortest-distance straight line edge geometries

Table 2. Testing Constraints for the Network Synthesis Algorithm

Network	Constraint 1: multilevel structure	Constraint 2: node location	Constraint 3: node degree	Constraint 4: edges
Real network	1	1	1	1
Synthetic type A	1	1	1	0
Synthetic type B	1	1	0	0
Synthetic type C	1	0	0	0

Note: 0 = network synthesis algorithm generated data; 1 = actual system data.

and C differ however based on Constraints 2–4, providing a means to test the sensitivity of the algorithm to the amount of real data that may be available.

To test the algorithm's ability to reproduce the encoded characteristics, synthetic representations of the network that are constrained with differing levels of available data have been produced. A number of randomly selected samples are presented in Fig. 6 for visual inspection. Networks follow a good visual adherence to the actual network; however, as expected this coherence lessens where networks are created with synthetic instead of real data. This is particularly apparent in synthetic Type C networks that do not preserve the peripheral structure of the northern part of the network (Cumbria region), which has evolved owing to geographic constraints. Analyses that use synthetic networks are therefore limited by the amount and quality of input data. Insights gained from such analyses should reflect this point and should be presented at an appropriate scale, with explicit mention of associated uncertainties.

Building the Integrated Electricity Network for England and Wales

A demonstration of the algorithm to produce a representation of the integrated electricity network for England and Wales is given in this section. Table 1 provides an overview of the data used in the representation. Although data are available for the location of power generators (DECC 2012) and the transmission grid (National Grid 2012), no data are available for either the subtransmission or the distribution networks. Synthetic networks for each voltage level are produced sequentially to represent the multilevel structure of this system. This starts with the transmission 132-kV level, where transmission level sinks and medium-level power generators are used as local source nodes. The algorithm outlined in this paper was then implemented to complete the 132-kV level before then producing synthetic representations of the 33-kV and finally the 11-kV levels. In running the algorithm, Constraints 1–4 are preserved, in which Constraint 1 is maintained using national data on the composition of network layers as described in Table 1. Constraint 2 is maintained by sampling from known substation

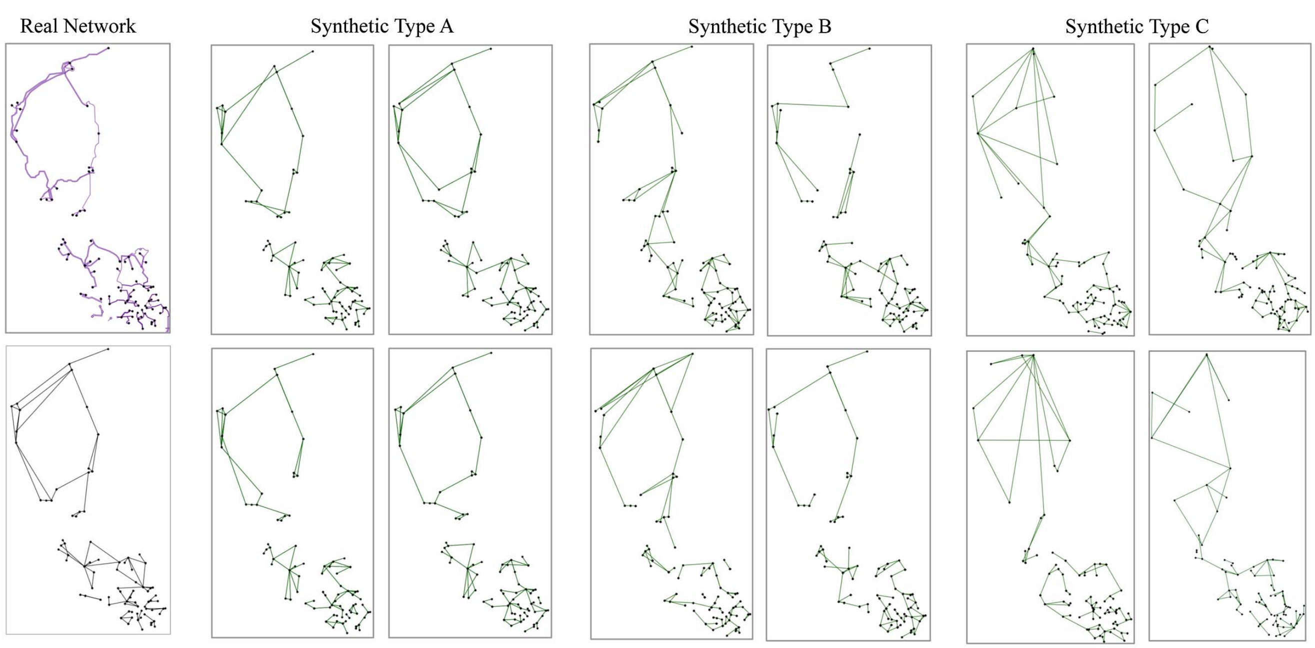

Fig. 6. Synthetic networks generated under varying constraints to produce multiple variants of the network, characterized as Type A, Type B, and Type C

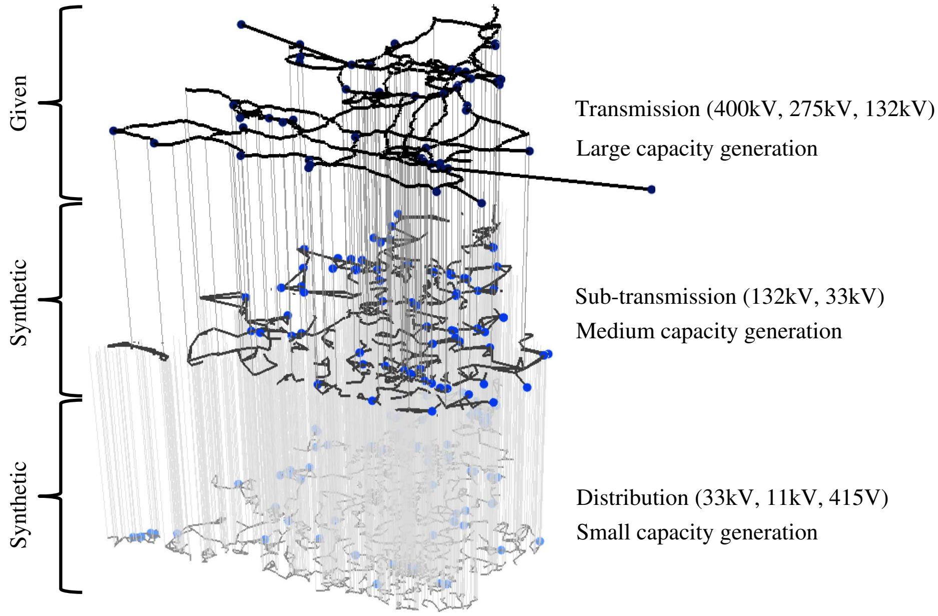

Fig. 7. Electricity network of England and Wales, showing the given real transmission network and synthetic subtransmission and distribution networks

locations (locations are known, but no voltage attribution is given) from OS MasterMap topography layer node data (Ordnance Survey 2015). Constraints 3 and 4 are derived directly from the electricity network data for the northwest of the country. A representation of the synthetic integrated network with details of the operational voltage at different levels is given in Fig. 7.

In England and Wales, National Grid, operates the transmission system, whereas seven different distribution network operators operate the subtransmission and distribution networks. The synthetic representation of the integrated electricity network for England and Wales provides a continuous network that bridges these operation scales—providing, for the first time, a network model that integrates the different distribution networks with a representation of transmission and power generation. Such a system-level representation provides a means to characterize hazard exposure and failure propagation within and between levels of the system and even between infrastructure sectors that are dependent on electricity at a certain voltage for their operation, such as water pumps, railway stations, and gas compressors. Doing so builds toward a system-of-systems-based understanding of the national infrastructure, providing key insights into potential failure and disruptions, highlighting areas of criticality for risk, resilience, and adaptation planning.

Discussion and Conclusions

Within this study, an algorithm for the synthesis of multilevel electricity networks for use in applied network failure and risk analysis is presented. The proposed algorithm is capable of producing networks that preserve a number of important spatial and topological properties of real-world networks, including multilevel structure of subsystems, the geographic distribution of network nodes, the node degree distribution, and the networks, spatial connectivity. Because of its modular structure, the algorithm is highly flexible and can accommodate different levels of data availability. This property of the algorithm lends itself to the synthesis of families of networks that can form a test bed for community analyses.

The algorithm has been demonstrated on a regional electricity network to show the ability of the algorithm to generate synthetic networks that maintain the properties with which they have been encoded. This results in the synthesis of networks that are close approximations to real networks. The synthetic nature of the networks produced by the algorithm means that the networks generated may be different to the reality on the ground. Rather intuitively, differences become more pronounced as fewer real data are used to constrain the algorithm. Interpretation of analyses that make use of such representations should therefore take this into account. Despite this, it is proposed that the algorithm is capable of maintaining the most salient properties of multilevel electricity networks, providing a worthwhile basis for interdependent infrastructure network analysis at an appropriate scale and level of interpretation that would otherwise have been impossible. An application of the algorithm for the integrated electricity network of England and Wales further highlights this point, demonstrating the potential to produce continuous multiscale representations that bridge multiple network operators (who do not make their real data publicly available), facilitating the evaluation of hazard exposure, failure propagation, and risk at the systems level—the scale at which they can occur in reality.

Although a demonstration is offered for an electricity network, the methodology has the potential to be applied for a range of infrastructure network systems that take a networked and multilevel form, for example, fluid network systems such as gas or water infrastructures that typically have a high capacity transmission systems and local, low capacity, distribution systems. The applicability of the algorithm beyond electricity networks would be an intuitive next step for this research. The flexibility offered by the algorithm to encode different network properties at different levels provides an opportunity to experiment on future network topologies. For example, one future application of the algorithm could be to produce synthetic variants of distribution networks with different topological properties, representing the topological changes that are already taking place, such as the connection of increased levels of distributed generation and electric vehicles. The future failure risks associated with these changes can then be calculated, not only for the electricity sector, but also for other sectors that have a critical dependence on electricity.

Methodology from the study of complex networks has been used in a number of applied studies of infrastructure network systems, providing important insights into network failure and risk analysis. Many studies are, however, impossible owing to the lack of availability of spatial and topological network data. This paper provides the means to address these concerns, providing methods to produce realistic synthetic network representations and enabling a multitude of studies in applied interdependent infrastructure network analysis.

Acknowledgments

The research reported in this paper was part of the U.K. Infrastructure Transitions Research Consortium (ITRC) funded by the Engineering and Physical Sciences Research Council under Programme Grant EP/I01344X/1. In addition, Scott Thacker would like to thank ARUP for additional funding of the research.

References

Albert, R., and Barabaśi, A. L. (1999). "Emergence of scaling in random networks." *Science*, 286(5439), 509–512.

Albert, R., Jeong, H., and Barabaśi, A. L. (2000). "Error and attack tolerance of complex networks." *Nature*, 406(6794), 378–382.

Aldous, D., and Shun, J. (2010). "Connected spatial networks over random points and a route length statistic." *Stat. Sci.*, 25(3), 275–288.

Barthelemy, M. (2011). "Spatial networks." *J. Physrep.*, 499(1–3), 1–101.

Bender, E. A., and Canfield, E. R. (1978). "The asymptotic number of labelled graphs with given degree sequences." *J. Comb. Theory Ser. A*, 24(3), 296–307.

Buchholz, B. M., and Styczynski, Z. (2014). *Smart grids—Fundamentals and technologies in electricity networks*, Springer, New York.

Cotilla-Sanchez, E., Hines, P. D. H., Barrows, C., Blumsack, S. (2012). "Comparing the topological and electrical structure of the North American electrical power infrastructure." *IEEE Syst. J.*, 6(4), 616–626.

D'Agostino, G., and Scala, A. (2014). *Networks of networks: The last frontier of complexity science*. Springer, Cham, Switzerland.

Davis, G. (1999). "The Auckland electricity supply disruption 1998: Emergency management aspects." *Aust. J. Emergency Manage.*, 13(4), 44–46.

DECC (Department of Energy and Climate Change). (2012). "Digest of UK energy statistics (DUKES)." London.

Dueñas-Osorio, L., Craig, J. I., Goodno, B. J., and Bostrom, A. (2007). "Interdependent response of networked systems." *J. Infrastruct. Syst.*, 10.1061/(ASCE)1076-0342(2007)13:3(185), 185–194.

ENA (Energy Networks Association). (2009). "Resilience to flooding of grid and primary substations." *ETR 138*, London.

ENW (Electricity North West). (2010). "Long term development statement." Warrington, U.K.

Erdös, P., and Rényi, A. (1959). "On random graphs. I." *Publ. Math. Debrecen*, 6, 290–297.

Ferretti, L., and Cortelezzi, M. (2011). "Preferential attachment in growing spatial networks." *Phys. Rev. E*, 84(1), 016103.

Ganin, A. A., Massaro, E., Gutfraind, A., Steen, A., Keisler, J. M., Kott, A., Mangoubi, R., and Linkov, I. (2016). "Operational resilience: Concepts, design and analysis." *Sci. Rep.*, 6, 19540.

Gao, J., Barzel, B., and Barabasi, A. L. (2016). "Universal resilience patterns in complex networks." *Nature*, 530(7590), 307–312.

Hines, P. D. H., Blumsack, S., and Schläpfer, M. (2015). "Centralized versus decentralized infrastructure networks." ArXiv e-prints arXiv: 1510.08792v1.

Johansson, J., and Hassel, H. (2010). "An approach for modelling interdependent infrastructures in the context of vulnerability analysis." *Reliab. Eng. Syst. Saf.*, 95(12), 1335–1344.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). "Multilayer networks." *J. Complex Networks*, 2(3), 203–271.

Lambert, J. H., and Sarda, P. (2005). "Risk identification by superposition of infrastructure and terrorist networks." *J. Infrastruct. Syst.*, 10.1061/(ASCE)1076-0342(2005)11:4(211), 211–220.

Lambert, J. H., Schulte, B. L., and Sarda, P. (2005). "Tracking the complexity of interactions between risk incidents and engineering systems." *Syst. Eng.*, 8(3), 262–277.

LaRocca, S., and Guikema, S. (2015). "Characterizing and predicting the robustness of power-law networks." *Reliab. Eng. Syst. Saf.*, 133, 157–166.

LaViolette, R. A., Beyeler, W. E., Glass, R. J., Stamber, K. L., and Link, H. (2006). "Sensitivity of the resilience of congested networks to rolloff and offset in truncated power-law degree distributions." *Physica A*, 368(1), 287–293.

Moini, N. (2016). "Modeling of risks threating infrastructures: System approach." *J. Infrastruct. Syst.*, 10.1061/(ASCE)IS.1943-555X.0000263, 04015010.

National Grid. (2012). "Electricity ten year statement, UK electricity transmission." Warwick, U.K.

Newman, M. E. J. (2003). "The structure and function of complex networks." *SIAM Rev.*, 45(2), 167–256.

Newman, M. E. J. (2010). *Networks: An introduction*, Oxford University Press, New York.

Ordnance Survey. (2015). "OS mastermap topography." Southampton, U.K.

Ouyang, M. (2014). "Review on modelling and simulation of interdependent critical infrastructure systems." *Reliab. Eng. Syst. Saf.*, 121, 43–60.

Pant, R., Thacker, S., Kelly, S., Hall, J. W., Barr, S., and Alderson, D. (2016). "The risk perspective: Risk of failure in interdependent infrastructure networks." *The future of national infrastructure: A system of systems approach*, J. W. Hall, M. Tran, A. J. Hickford, and R. J. Nicholls, eds., Cambridge University Press, Cambridge, U.K.

Pitt, M. (2008). "The Pitt review: Learning lessons from the 2007 floods." ⟨http://archive.cabinetoffice.gov.uk/pittreview/thepittreview/final_report.html⟩ (Sep. 18, 2017).

PNNL (Pacific Northwest National Laboratory). (2015). "The emerging interdependence of the electric power grid & information and communication technology." *PNNL-24643 Final Rep.*, Richland, WA.

Rinaldi, S. M., Peerenboom, J. P., and Kelly, T. K. (2001). "Identifying, understanding, and analysing critical infrastructure interdependencies." *IEEE Control Syst.*, 21(6), 11–25.

Thacker, S., Pant, R., and Hall, J. W. (2014). "Characterizing the vulnerability of future configurations of Great Britain's electricity network infrastructure to climate-related hazards." *2nd Int. Conf. on Vulnerability, Uncertainty and Risk*, ASCE, Reston, VA, 648–657.

Thacker, S., Pant, R., and Hall, J. W. (2017). "System-of-systems formulation and disruption analysis for multi-scale critical national infrastructures." *Reliab. Eng. Syst. Saf.*, 167, 30–41.

U.S.–Canada Power System Outage Task Force. (2004). "Final report on the August 14th, 2003 blackout in the United States and Canada: Causes and recommendations." U.S. Dept. of Energy and Canadian Dept. of Natural Resources, Washington, DC.

Wang, Z., Scaglione, A., and Thomas, R. J. (2010). "Generating statistically correct random topologies for testing smart grid communication and control networks." *IEEE Trans. Smart Grid*, 1(1), 28–39.

Watts, D. J., and Strogatz, S. H. (1998). "Collective dynamics of 'small-world' networks." *Nature*, 393(6684), 440–442.

J. Infrastruct. Syst., 2018, 24(1): 04017043