When are decentralized infrastructure networks preferable to centralized ones?

Paul D. H. Hines
University of Vermont
paul.hines@uvm.edu

Seth Blumsack
Penn State University
sethb@psu.edu

Markus Schläpfer
ETH Future Cities Laboratory
schlaepfer@arch.ethz.ch

Abstract

Many infrastructure networks, such as power, water, and natural gas systems, have similar properties governing flows. However, these systems have distinctly different sizes and topological structures. This paper seeks to understand how these different features can emerge from relatively simple design principles. Specifically, we work to understand the conditions under which it is optimal to build small decentralized network infrastructures, such as a microgrid, rather than centralized ones, such as a large high-voltage power system. While our method is simple it is useful in explaining why sometimes, but not always, it is economical to build large, interconnected networks and in other cases it is preferable to use smaller, distributed systems. The results indicate that there is not a single set of infrastructure cost conditions that cause a transition from centralized networks being optimal, to decentralized architectures. Instead, as capital costs increase network sizes decrease gradually, according to a power-law. And, as the value of reliability increases, network sizes increase abruptly—there is a threshold at which large, highly interconnected networks are preferable to decentralized ones.

1 Introduction

Water, natural gas, and electric power networks share the following operating principle: all of the product flow that goes into a node must also come out. While there are differences in the physics of flow in these networks (most notably in their relative time scales), similarities in the underlying physics can result in similar design principles and thus similar topological properties. For example, all three network types sometimes use meshed topological designs, which provide redundancy, and sometimes use tree-like structures that allow only one path between sources and sinks.

On the other hand, there are substantial differences in the average sizes and typical structures for these networks. Consider, for example, electric power and water distribution networks. In the United States alone there are more than 150,000 public drinking water networks that serve at least 25 people [1]. In contrast, the continental U.S. has only three major power networks: the Eastern, Western, and Texas interconnections. The U.S. natural gas network lies between these two extremes with the U.S. system divided into around 200 distinct networks [2]. Why is it that in the case of drinking water, smaller, more decentralized networks seem to be optimal, whereas for electric power, larger systems that span continents seem to be preferable? Both systems have similar physical properties that govern flows. Both systems transport largely interchangeable goods: one electron is as good as another, just as one water molecule is as good as another (given appropriate standards for cleanliness). However, these two systems have fundamentally different size scales. A drinking water system serves, on average, 2000 people. A power network serves, on average, 100,000,000 people.

In addition to differences in existing networks, two distinct transitions are pushing infrastructure systems toward, or away from, centralized architectures.

The first transition is the growing push toward decentralized, "local" systems in more developed countries. The energy infrastructures in most industrialized countries have evolved into large, complex networks [3] that depend heavily on large centrally-located power plants. However the growing movement toward the use of local resources (e.g, microgrids) is a push back toward small, relatively independent systems [4]. The historical case for large interconnected systems (economies of scale in generation and transmission) and increased redundancy through interconnection is being challenged by falling costs for distributed power generation [5], increased interest in smart microgrids, and the insight that distributed systems may offer improved local reliability in some cases [6].

The second transition is the rapid growth of infrastructure, including electric power, water, transportation and natural gas, in less developed countries [7]. Prior work has argued that for political, geographic and economic reasons, the greenfield build-out of highly interconnected electric power infrastructure may not be desirable in developing-nation contexts, particularly in locations that are subject to elevated levels of stress [8]. However, international development projects often focus on the design of large networks that depend on central power plants [9].

While the contexts for infrastructure decisions in more-developed and less-developed nations differ, the basic question remains the same: given the need to meet

URI: http://hdl.handle.net/10125/41549
ISBN: 978-0-9981331-0-2

growing demand for infrastructure products (water, electricity, gas), what mix of small-scale versus large-scale system architectures will most effectively balance cost and reliability? Under what assumptions about cost, reliability and other factors would it be more advantageous to use a decentralized architecture? This work seeks to describe the conditions under which the transformation of large-scale systems into multiple smaller-scale systems, or the greenfield construction of multiple smaller-scale systems to serve a large geographic area, would yield improvements in cost and reliability.

This paper builds on prior work focused on the optimal planning of networks that deliver services or otherwise provide connectivity over physical space. Our approach is to model a single planner making optimal resource decisions, which differs from game-theoretic approaches of network generation or the literature in random generation of synthetic networks [10], [11]. While the design of infrastructure networks has been of interest to geographers since the 1960s [12], [13], spatial network design has emerged only more recently as an area of scientific research [14]. Many applications of spatial network analysis focus on traffic or transportation networks [15], [14], [16], [17], [18] or physical infrastructures that deliver information, such as the internet or mobile telephony [14], [19]. Others find that the constraints of geographic space can dramatically change the implications found in abstract network models [20], [21]. A recent study of a complex systems model of cascading blackouts [22] suggested that there are optimal sizes for infrastructure networks in which the risk posed by outages is balanced by the benefit given the redundancies that are possible in larger networks.

Research on the spatial aspects of network design has largely focused on two areas. The first is how the cost of adding edges or otherwise connecting nodes in space influences network structure and design choices [14], [23]. The second strand uses known or theoretical spatial properties of networks to understand their performance in the case of attacks, failures or other contingency events [6], [18], [24].

We build on this extensive body of work, and add to its relevance for electrical networks, in two ways. The first is to integrate some of the most salient properties of electric power networks (namely Kirchhoff's Current Law) into the cost-driven spatial network design problems discussed in [14] and [23]. These properties are important for electric power networks specifically because while expansion costs may be straightforward to parametrize in terms of spatial distance, actual network flows are not so simply represented. Further, production costs in real electrical networks are heterogeneous by technology and in space, because of regional variations in resource endowments and technology choices. The second is to consider a design objective that incorporates the costs of network operation, network expansion, and network (un)reliability. While joint planning and operations models have been devised for incremental expansion decisions [25], [26], [27], and for optimal topology control applications (e.g., [28], [29]), our approach is different in its consideration of a flexible greenfield infrastructure build problem that does not, for example, restrict infrastructure expansion options to pre-defined paths or represent branches as binary integer variables. Detailed approaches such as the above are useful for solving particular engineering problems, but the results do not often reveal statistical insight that could lead to the discovery of more general principles underlying the design of optimal networks. It is our conjecture that simpler models, such as the attachment patterns of slime mold [30] or the approach proposed in this paper, may be helpful in uncovering such general design principles.

2 Notation

This paper uses the following notational conventions. Bold lower-case symbols, $\mathbf{x}$, represent column vectors and non-bold symbols with single subscripts, x_i, represent elements of those vectors. Matrices are represented by upper-case bold letters, $\mathbf{Y}$, with subscripted bold lower-case letters, $\mathbf{y}_i$, representing columns of those matrices; scalar elements of matrices are represented by lower-case script, y_{ij}. This paper simultaneously co-optimize the design of infrastructure networks and the operation of those networks. To clarify this distinction in our notation, variables representing the quantity of infrastructure to build are written with an overscore, $\overline{\mathbf{z}}$ and those representing the operating state of a particular element are not, $\mathbf{z}$ or z_i. Sets of indices are indicated with upper-case italic symbols, Q and elements of those sets with corresponding lower-case letters, q.

3 Optimal Infrastructure Network Design: Cost alone

To start our exploration of optimal network design, let us consider a system planner who has the task of designing (or modifying) an infrastructure system to provide a particular infrastructure product (water, natural gas, electricity, etc.) for a set of n locations (towns, buildings, etc.). Each location (vertex) v has some known demand for the product d_v and also has the ability to produce this product locally with a production cost that varies with geography. In order to satisfy the demand at node v one can either produce d_v locally or build an interconnection to some nearby node that can produce less expensively.

Given a model of this sort we can ask a number of important questions. Under what conditions is it optimal to produce locally, rather than building interconnections? If the goal of the planner is to minimize overall cost, what type of network would one want to build? Should one build many small networks or

one large one? Should the planner build a meshed network that allows redundancy, or a radial network that provides only one path between sources and sinks?

In this section we study a model that seeks to understand what types of infrastructure networks one should build if we solely care about minimize cost. Section 3 extends the study to add reliability to the objective. Both models use a "greenfield" approach, in which we assume that there is no existing infrastructure and then seek to find the optimal network that satisfies the total demand.

3.1 Formulation

Let us assume that each of the n locations is a vertex ($v \in V$) that has coordinates x_v, y_v in some 2d space, known demand d_v, and a per unit cost for production χ_v. This production cost term combines the marginal cost of producing one unit of the good c_v, and the levelized cost of the production capacity κ_v needed to produce at a rate of one unit per unit time at vertex v: $\chi_v = c_v + \kappa_v$. For example, if a particular power plant technology costs \$1000/kW to build, it would have a levelized capacity cost of $\kappa_v = \$80$ per kW-year when amortized over a 20-year horizon at a 5% discount rate. If its production cost were \$10 per MWh, it would have a annual production cost of $c_v = \$88$/kW-year. In addition assume that there is a maximum feasible set of undirected edges $e \in E$ that one might choose to build. For example, one might allow into E all possible vertex pairs, thus allowing at most $n(n-1)/2$ edges. To represent this maximum feasible network in our optimization problem, we transform E into a sparse $m \times n$ edge matrix $\mathbf{E}$, with m representing the number of feasible edges. The matrix $\mathbf{E}$ has single entries of 1 and -1 on each row indicating the end points of each edge. The cost of building any one particular edge $e_{i \leftrightarrow j}$ depends on two factors: the length of the edge $\ell_{e_{i \leftrightarrow j}} = \left((x_i - x_j)^2 + (y_i - y_j)^2\right)^{1/2}$ and the cost of one unit length-capacity of edge construction, w.

Given these input data, the following formulation allows one to compute an "optimal" infrastructure network design.

$$\min_{\overline{\mathbf{f}},\mathbf{f},\overline{\mathbf{g}},\mathbf{g}} \quad \sum_{v \in V} (c_v g_v + \kappa_v \overline{g_v}) + w\sqrt{n} \sum_{e \in E} \ell_e \overline{f_e} \tag{1a}$$

$$\text{s.t.} \quad 0 \le g_v \le \overline{g_v}, \forall v \tag{1b}$$

$$\overline{f_e} \ge 0, \forall e \tag{1c}$$

$$-\overline{f_e} \le f_e \le \overline{f_e}, \forall e \tag{1d}$$

$$\mathbf{g} - \mathbf{d} = \mathbf{E}^\intercal \mathbf{f} \tag{1e}$$

where $\overline{f_e}$ and f_e are the undirected flow capacity and actual directed flow on edge e, g_v is the actual amount of production at vertex v, $w\sqrt{n}$ is an interconnection cost parameter (cost per-unit length·capacity), and $\mathbf{E}$ is an $m \times n$ edge matrix that comes from the set of feasible edges E.

The objective (1a) is to minimize the combined cost of production, $\mathbf{c}^\intercal \mathbf{g} + \boldsymbol{\kappa}^\intercal \overline{\mathbf{g}}$, and interconnection, $\boldsymbol{\ell}^\intercal \overline{\mathbf{f}}$, while satisfying constraints (1b)-(1e). Constraint (1b) defines locational production limits; (1c) ensures that we do not build negative quantities of interconnection capacity; (1d) constrains flows to be less than the chosen flow limits; and (1e) ensures that the net flow into and out of each vertex must be zero (Kirchhoff's Current Law for electrical networks). Because there is, in this case, no cost advantage to building surplus capacity, the amount of capacity built will be the same as the operating state: $\overline{\mathbf{g}} = \mathbf{g}$, and the absolute edge flow will be equal to the edge flow capacity: $\overline{\mathbf{f}} = ||\mathbf{f}||$. As a result the incremental and capital production cost terms can be combined into a single cost term $\sum_v \chi_v g_v$ and an optimal solution can be found without explicitly solving for $\overline{\mathbf{g}}$.

Because we are interested to find properties of this system that do not depend on the number of nodes, the cost function (1a) is designed so that both the production and the edge construction terms grow linearly with n. In order to implement this, we first observe that (at least for the case of uniformly distributed node locations) the distances between randomly selected node pairs decreases according to: $\ell_e \sim n^{-1/2}$.[1] As a result, ensuring linear growth of the edge cost term requires that we scale the second cost term by $\sqrt{n}$. This allows us to study different spatial scales with the same formulation: a large country with many nodes, or a small city with fewer nodes.

Implied in this formulation are a number of important assumptions. First, we assume that interconnections can be built at any size scale and that construction costs scale linearly with the capacity of the edge. It is certainly possible to think of particular examples, such as transmission line construction, where costs exhibit returns to scale, such that building a 1 MW transmission line is more than 1/100 of the cost of building a 100 MW transmission line. However, if we consider that the edge might be either a large transmission line or a small distribution line, this assumption is not as obviously incorrect. Modeling scale economies of this sort would require knowledge about the details of a particular infrastructure system at a particular place and time. In this paper we are more interested to identify general trends that appear in optimal infrastructure designs. Second, formulation (1) models a single snapshot of demand, whereas all real infrastructure systems have demand that varies in time. This allows us to combine the two production cost terms c_v and κ_v in a way that melds together the capital and operating costs associated with supplying

[1] This can be derived by first observing that the number of points per unit area is n. Thus each point occupies a space that would be (if it were square) $A = \frac{1}{n} = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}}$. Thus as $n \to \infty$ the distance between a point and its nearest neighbor is approximately $n^{-1/2}$. A more formal derivation of this remains for future work.

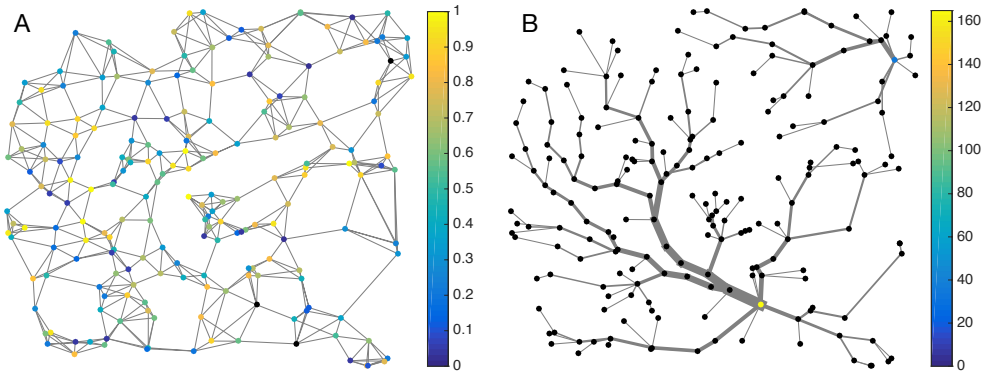

Fig. 1. Illustrative results for random placement of "towns" on a 2d plane. Panel A shows an initial, maximum feasible graph for a $n = 200$ network, with node colors indicating the production/capacity cost, χ_v, at each "town." Panel B shows the optimal network configuration, after solving (1a)-(1e) for $w = 0.001$, with node colors indicating the amount of production at each node, g_v, and edge thicknesses indicating the flow capacity. This optimal network has two connected components and thus an average component size of $< n_s >= n/2 = 100$.

the demand d_v. If this assumption were relaxed and time-varying demand added to the problem, the results from (2) would likely be similar in many ways, except that the optimal network would need to have some surplus capacity to allow for peak demand, and thus it would no longer be the case that $g_v = \overline{g_v}$ and $|f_e| = \overline{f_e}$ Finally, we make the assumption that nodes do not include any storage, leading us to include constraint (1e), which is equivalent to Kirchhoff's Current Law. While some types of infrastructure networks, such as water and natural gas, tend to have some storage at some nodes, even with storage the long-run average net flow out of a node must sum to zero. In this way our model effectively captures the average long-run patterns of flows through a network.

3.2 Cost only results

To explore the properties of networks that result from (1) consider n nodes randomly located within a 1×1 2d square, such that each node location x_v, y_v is a uniform random variable in $[0, 1]$. Each of these nodes has a production cost $\chi_v : c_v = \kappa_v = \chi_v/2$ that is also a uniform random variable in $[0, 1]$, and a demand $d_v = 1$. The set of feasible edges that we might decide to build (the feasible graph E) comes from initially setting E to be a modified form of the random geometric graph [31]. Whereas in the conventional random geometric graph nodes are typically connected to other nodes that lie within some prespecified radius, we instead connect each node i to i's k nearest (Euclidian) neighbors, while avoiding the addition of duplicate edges. Because it is possible that j is one of i's k nearest neighbors, but i is not one of j's k nearest neighbors, the resulting E has an average degree that is slightly larger than k. If one were to set $k = n - 1$ the result would be the full graph of $n(n-1)/2$ possible edges. However, allowing this many potential edges into E makes the optimization problem computationally impractical for all but the smallest problems. Instead the results in this section come from choosing $k = 5$. We found, via experimentation, that the key statistical properties of the resulting networks did not change substantially by increasing k above 5.

Figure 1 illustrates the application of this approach to a system with $n = 200$ nodes and $w = 10^{-3}$. From this figure a few observation can be made. First, we see that the algorithm tends to produce tree-like graphs in which the number of edges in each connected component, m_s, is one less than the number of nodes in that component n_s. The reason for this is fairly straightforward: creating a loop means that there are redundant paths between node pairs. Given a network with a loop, one can always reduce the edge-construction cost term $w\sqrt{n}\sum_{e\in E}\ell_e\overline{f_e}$ by removing one edge in the loop, without loss of functionality. As a result the graphs that result from (1) are always treelike, with precisely $m_s = n_s - 1$ edges in each component. Secondly, we see that there are two connected components in the optimal network and thus an average of $< n_s >= n/2 = 100$ nodes per component. In this illustration the two least-expensive production nodes had costs of $c_{v_1} = 0.0003$ and $c_{v_2} = 0.0046$, with the less expensive node supplying the larger sub-component. While it would have been feasible to connect the two components with a fairly short additional edge, supplying the whole network from the less expensive node v_1 would have required building additional capacity along the spidery path from v_1 to v_2. Doing so would have cost more than the additional cost of supplying the second component from the more expensive unit, a cost of $n_{s_2}(c_{v_2} - c_{v_1}) = 0.149$. For comparison purposes, the cost of building a length $\ell_e = 1$ edge that could supply the whole of the 35 node second component

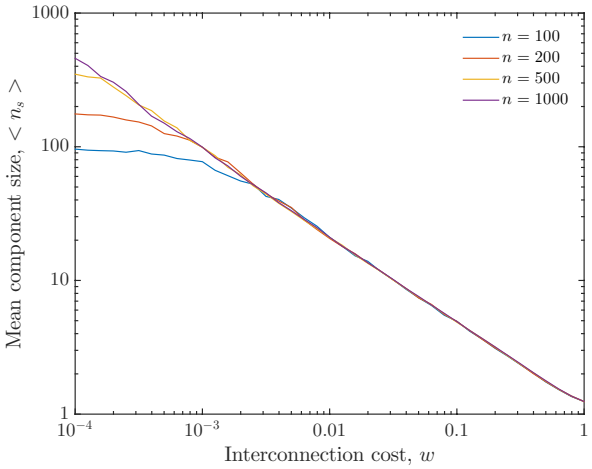

Fig. 2. Mean connected component sizes as a function of the interconnection cost parameter w. For each of the four different network sizes considered, we find that component sizes scale approximately as $w^{-2/3}$.

would be $35w\sqrt{n} = 0.495$. This result illustrates a property that occurs in all of the optimal networks resulting from (1): each of the subgraphs (components) in the network is supplied by precisely one production plant. Since there are no losses in formulation (1) the cost of transportation is independent of distance and there no reason to build more than one supply node in each sub-network.

Given that this approach can determine "optimal" network sizes, it is natural to ask how those network sizes change as the cost of building network infrastructure changes. For example, as w increases one might expect to see a relatively sudden phase transition from optimal networks that span the entire network to optimal networks with many small, decentralized sub-systems. In order to investigate this further and understand the cost conditions under which centralized, or decentralized, networks are optimal, we performed the following experiment. For several values of n, we computed optimal infrastructure networks using (1) over a range of w from 10^{-4} to 1. For each value of w we re-initialized the random node locations, the feasible network E and production costs χ_v, and computed the optimal network configuration using (1) 200 times. Then we recorded the mean size of the connected components $< n_s >$ over the 200 optimal networks.

Figure 2 shows the resulting relationship between the edge construction cost, w and optimal component sizes, n_s. As one would expect, as network construction costs increase the optimal network sizes decrease. However, what is less obvious is that the change from large networks to small networks does not occur suddenly as does the first-order phase transition from a solid ice to liquid water. Instead, this transition occurs gradually over several orders of magnitude in w. In fact, fitting the data in Figure 2 to a power-law distribution indicates that mean component sizes fall as

$$< n_s > \sim w^{-0.648} \sim w^{-2/3}. \quad (2)$$

4 Optimal Infrastructure Network Design: Adding reliability

An obvious limitation in formulation (1) is the omission of reliability from the design criteria. Reliability has an enormous role in the design of infrastructure systems. In power systems, for example, electric utilities frequently argue in rate-case filings that the construction of a new transmission line is justified purely on reliability grounds. It is thus useful to understand what impact reliability considerations have on the topological structure of infrastructure networks.

In order to model the impact of reliability we add a third term to the original objective function (1a) to capture the cost of (un)reliability. Specifically, we measure the cost of not serving demand in response to a set of node or link outages (perturbations), P. These costs are assumed to be a linear function of the amount of unserved load over all perturbations. Unserved load is represented by a $n \times |P|$ matrix, $\Delta\mathbf{D}$ with non-positive elements. In addition to $\Delta\mathbf{D}$, we introduce matrices $\Delta\mathbf{G}$ and $\Delta\mathbf{F}$ into the set of decision variables to represent the changes in production and flow after each of the the possible outages in P. Introducing these three matrices allows the optimizer to determine how the network 'should' react to each of the disturbances modeled. As in (2), we separate the production cost term χ_v into separate terms for capital and operating costs, however in this case one will sometimes want to build surplus capacity to prepare for plausible failures. With these additions we get the following formulation.

$$\min_{\substack{\mathbf{f},\bar{\mathbf{f}},\mathbf{g},\bar{\mathbf{g}} \\ \Delta\{\mathbf{D},\mathbf{G},\mathbf{F}\}}} \quad \sum_{v\in V}(c_v g_v + \kappa_v \overline{g_v}) + w\sqrt{n}\sum_{e\in E}\overline{f_e} l_e - \frac{r}{n}\sum_{p\in P}\mathbf{1}^\intercal \Delta\mathbf{d}_p \quad (3a)$$

$$\text{s.t.} \quad 0 \le g_v \le \overline{g_v},\ \forall v \in V \quad (3b)$$

$$0 \le |f_e| \le \overline{f_e},\ \forall e \in E \quad (3c)$$

$$\mathbf{g} - \mathbf{d} = \mathbf{E}^\intercal \mathbf{f} \quad (3d)$$

$$\Delta g_{v,p_v} = -g_v,\ \forall v \in V \quad (3e)$$

$$\Delta f_{e,p_e} = -f_e,\ \forall e \in E \quad (3f)$$

$$|f_e + \Delta f_{e,p}| \le \overline{f_e},\ \forall p \in P,\ \forall e \in E \quad (3g)$$

$$0 \le g_v + \Delta g_{v,p} \le \overline{g_v},\ \forall p \in P,\ \forall v \in V \quad (3h)$$

$$-d_v \le \Delta d_{v,p} \le 0,\ \forall p \in P,\ \forall v \in V \quad (3i)$$

$$(\mathbf{g} + \Delta\mathbf{g}_p) - (\mathbf{d} + \Delta\mathbf{d}_p) = \mathbf{E}^\intercal(\mathbf{f} + \Delta\mathbf{f}_p),\ \forall p \in P \quad (3j)$$

In this formulation, (3a) is the modified objective, which now includes the reliability term, and r is a

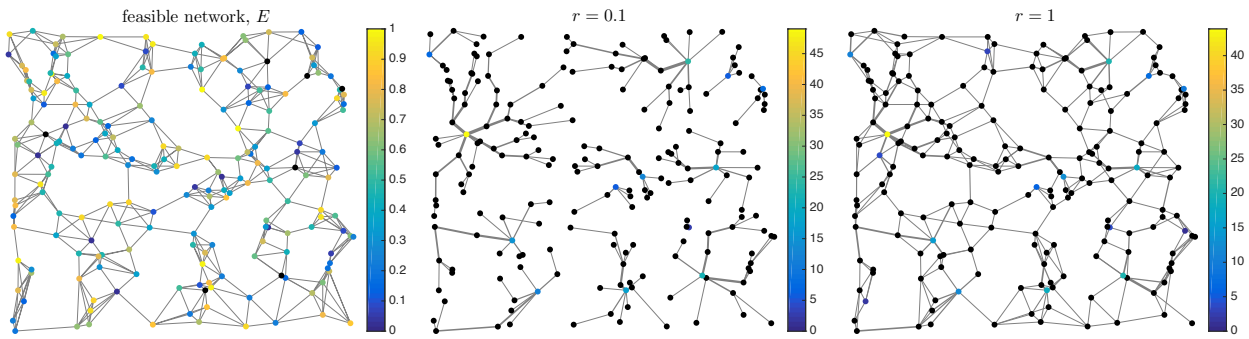

Fig. 3. Illustrating the impact of adding reliability to the optimal network construction formulation. Panel A shows the full feasible network E, with colors indicating production costs χ_v. Panel B shows the optimal network for $w = 0.01$ and $r = 0.1$, which is identical to the tree-like network that results from the simple model. Panel C shows the optimal network for $w = 0.01$ and $r = 1$, which shows the emergence of a meshed topology and substantial supply redundancy. Colors in panels B and C indicate the amount of production capacity at each node.

reliability cost parameter that allows us to adjust the relative importance of reliability. $\Delta\mathbf{d}_p$ is the change (loss) of demand that results from perturbation p, which is one of the set of all perturbations, P. As with with the second term in (1a), we divide the reliability term by n so that this term also roughly increases linearly with n. Eqs. (3b) and (3d) are equivalent to (1b)-(1e) in (1). Eqs. (3e) and (3f) cause the specific node, p_v, and edge, p_e, failures that together make up the set of all perturbations, P, by forcing the nodal production or edge flow to be zero for the a particular component outage in P. Eq. (3g) ensures that all flows are below edge capacities, after all perturbations. As a result there are $m(n + m)$ constraints of this type within the formulation. Similarly, (3h) constrains production at every node after each perturbation (a total of $n(m+n)$ constraints), to be below the chosen production capacities for each node. Eq. (3i) ensures that demand can only decrease, and only up to the total demand at node v, as a result of each p. Finally, (3j) enforces a nodal supply/demand balance after each perturbation. This forces the optimizer to compute production, demand loss, and flow patterns that obey Kirchhoff's Current Law for each disturbance in P.

This approach allows us to include a wide variety of disturbances in P. However, a limiting factor is the fact that the size of the linear program grows rapidly with both the number of disturbances included and with the size of the network. Specifically, the total number of constraints in (3) is $3n + 3m + 3|P|n + 2|P|m$, where $|P|$ is the size of the set P. If we start with a feasible network E with a fixed degree such that $m = kn/2$, and we include within P only the single edge and node outages, then the problem will have $3(0.5k + 1)n + (0.5k^2 + 2.5k + 3)n^2$ constraints. To illustrate, solving (3) for a network with $n = 200$ and $k = 5$ will require 1 122 100 constraints. Thus, in order to manage the computational complexity of this problem, we include in P only the single node and edge outages.[2]

As a whole this formulation allows us to observe how network size and structure changes as we increase the relative importance of reliability. If $r = 0$, demand losses are effectively deemed irrelevant, and the problem will produce results that are identical to those obtained from (1). On the other hand, as r increases we hypothesize that networks are likely to become more meshed (rather than tree-like) and more likely to include surplus production capacity. It is not obvious, *ex ante*, how r will impact optimal network sizes. On the one hand, small, local networks will be more robust to edge failures and thus may be more optimal when reliability is very important. On the other hand, large interconnected systems provide a high level of redundancy, which also has tremendous value. In the results that follow we explore this tradeoff.

4.1 Reliability results

This section aims to understand how the results from the cost-only model (1) change after adding reliability costs as in (3). As in the cost-only model, we consider nodes scattered uniformly on a 2d plane. Also as before, we assume that each location has an overall production cost that is a uniform random variable in $[0, 1]$ and we assume that the production cost is split evenly between marginal and capital costs, $c_v = \kappa_v = 0.5\chi_v$. Here we restrict our attention to the case of networks with $n = 100$ nodes, since solving (3) for larger systems leads to prohibitively large solution times.

Figure 3 shows illustrative results for a $n = 50$ node network, which clearly show the importance of reliability to network structure. For small values of r the solutions are nearly identical to what we get

[2]Note that the majority of runs of the reliability model in this paper were run with $n = 100$ nodes. These runs required approximately 4-8 hours of computer time and 2-4 GB of memory using the CPLEX solver on a single Intel Xeon CPU.

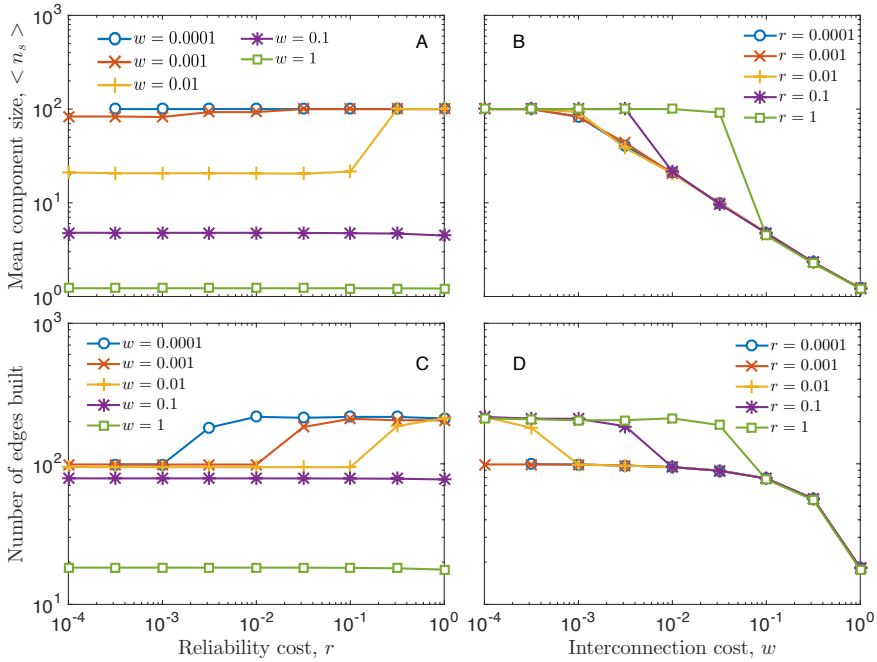

Fig. 4. Statistical results from the reliability model. Panels A and B show mean component sizes as a function of the reliability cost and interconnection cost parameters, r and w. Panels C and D show the average number of edges in the optimal network, also as a function of r and w.

from the simple model: tree-like networks that satisfy demand with no redundancy. However as r increases, we find a (rather sudden) transition to meshed networks that include substantial supply redundancy. For the example in Fig. 3, the system builds a network with total generation capacity equal to $\sum_{v \in V} \overline{g_v} = 82.2$, much more than what is needed to supply the 50 nodes in the system.

Next we computed optimal networks for several different values of the reliability parameter r and interconnection costs w for 100 node networks. For each value of w and r the random variables (x_v, y_v and c_v) were re-initialized 100 times in order to minimize variance.[3]

Figure 4 shows the resulting mean component sizes for various values of w and r. Fig. 4A describes the sensitivity of the mean component size to the relative value of reliability while holding the relative cost of interconnection fixed. For $w \leq 0.001$ and $w \geq 0.1$ the network sizes do not change substantially with r. In particular, for the smallest values of w the mean component size is at or near the size of the network, while for the larger values of w ($w \geq 0.1$) the size of the mean component is 5 nodes or fewer (for the case $w = 1$ it is hardly optimal to build any interconnection). For the intermediate case $w = 0.01$, however, we see a jump in optimal component size, to 100 nodes (the size of the whole network) once r becomes larger than 0.1.

We do, however, see a more sensitive relationship between component sizes $< n_s >$ and w when r is held fixed. Fig. 4B shows that for any value of r, a very small cost of interconnection yields a complete network (mean component size = 100), which is similar to the results in Fig. 4A. As r increases beyond 0.001, the mean component size begins to decline as the interconnection cost w becomes larger than 0.001. The transition point and rate of decline is identical for $r \leq 0.01$, and the rate of decline is also identical for all values of r once w exceeds 0.1 (where the mean component size declines to $< n_s > \cong 4$, indicating that little interconnection is built). For $r \geq 0.1$ we observe that an order of magnitude increase in r (from 0.1 to 1) corresponds to an order of magnitude increase in the value of w (roughly 0.015 to 0.15) at which the transition occurs.

Not only do the optimal network sizes change, but the level of redundancy also changes with r and w. One way to measure the level of redundancy is by the number of edges constructed in the optimal network. In the tree-like networks that result from the simple model there are always fewer than n edges. But, as shown in Figs. 4C and 4D, as r increases the number of edges in the optimal networks also increase, increasing redundancy. Unlike with the relationship between w and $< n_s >$, this increase becomes quite abrupt for

[3]A few of these cases failed to solve, which means that a few of the results are averaged over fewer than 100 trials.

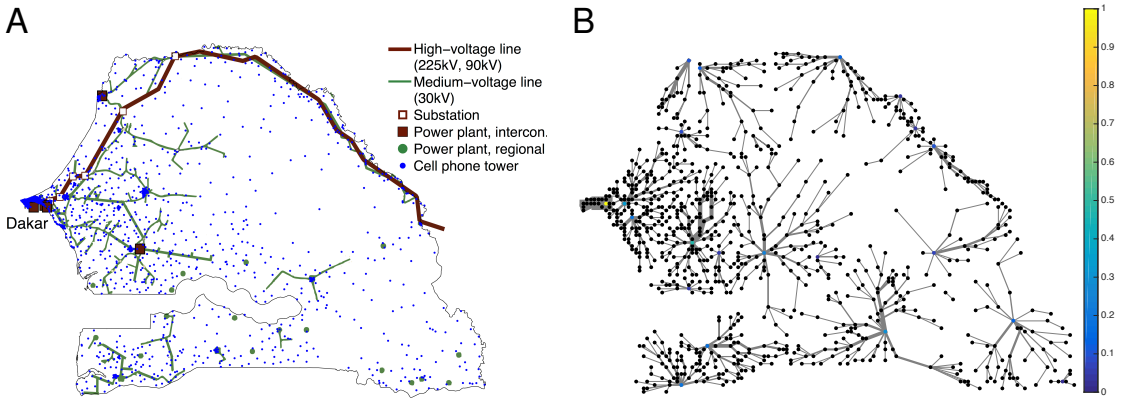

Fig. 5. Application of the optimal network design model to the geographic distribution of electricity demand centers in Senegal. Panel A shows the existing high and medium voltage network together with the geographic distribution of the mobile phone towers (adopted from [32]). Panel B shows the optimal network for $w = 0.01$, given random production costs and ignoring reliability.

sufficiently small values of w.

Taken together, Fig. 4A-D suggest several lessons concerning the scope of optimal infrastructure networks. First, very small interconnection costs yield highly connected networks; this is invariant to the value of reliability. Second, even when the value of reliability is large in our simulations, the construction of additional local supply has lower costs than increased interconnection except when the interconnection cost is orders of magnitude smaller than the value of reliability. Third, when the value of reliability is high relative to the cost of interconnection the extent of interconnection redundancy increases (Fig. 4C) even in highly connected networks. Fourth, only when the value of reliability and the cost of interconnection are both large is it sensible to build redundancy through additional local supply rather than through interconnection (Fig. 4B and D).

5 Senegal application

As a real-world case study, we applied the infrastructure network design model to the geographic distribution of cities and rural towns in Senegal. About half of the countries' population still has no access to electricity, and the electrification rate in rural areas is as low as 28% [32]. In contrast to Senegal's electric power grid, the mobile communication infrastructure is highly developed, with 1666 mobile phone towers distributed across the country and a mobile phone penetration rate of almost 100%. This allows for the use of data from the mobile communication system to accurately predict the geographic distribution of electricity demand (see [32]). Data on the mobile phone infrastructure has been made available by ORANGE / SONATEL within the framework of the D4D Challenge [33]. Figure 5A depicts both the existing electricity infrastructure and the location of the mobile phone towers.

In order to use data from the communication system to model demand for electricity, we first partitioned the country into a rectangular grid with cell size 5km $\times$ 5km. Following the approach in [32] we then used the number of cell phone towers that are located in each grid cell as a proxy for the relative electricity demand within that cell, d_v. Note that for the purpose of our analysis we are not interested in estimating absolute demand, but rather the relative amount of electricity that might be consumed in a particular location. The center points of the grid cells were used as locations x_v, y_v for the load nodes. As with the uniformly distributed vertices, we randomly assigned production costs to each node in the network, using the load locations described above, using uniform random variables over $[0, 1]$.

Figure 5B shows the result of applying the basic optimization model (1) for the case of $w = 0.01$. Without reliability constraints, our model produces tree-like networks that are similar in structure to what we found with randomly distributed vertices in Sec 3.2. The network realization resembles the spatial topology of the existing electricity grid in Senegal, indicating low reliability standards. Moreover, we find that the optimal component sizes for this real-world system (Fig. 6) scale in the same way as was found for synthetically generated load points in Eq. (2). This suggests that the power law decay in optimal network sizes also holds for real-world settings where the geographic location of load centers is far from being random.

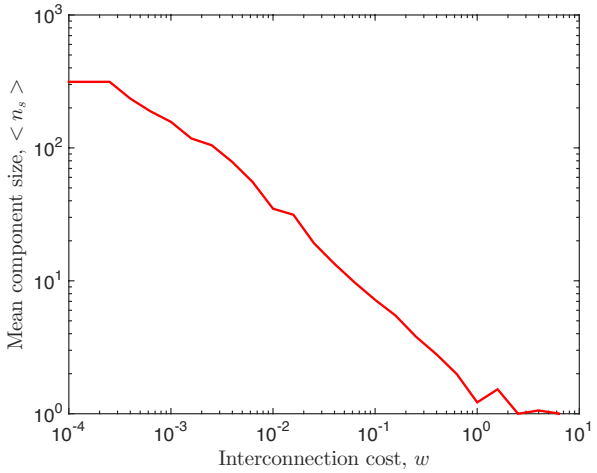

Fig. 6. The mean size of network components for Senegal.

Application of the reliability model (3) to the Senegal system remains for future work.

6 Discussion and Conclusions

This paper presents results from a model of optimal infrastructure network design, with which we aim to better understand the conditions under which small decentralized infrastructure networks are preferable to large systems that depend on a centralized architecture. Several interesting observations emerge from this work. First we find, unsurprisingly, that as network construction costs increase the optimal size of infrastructure networks decreases and the local provision of services becomes preferable. In the case where cost is the primary network design objective and reliability is not important, we unsurprisingly find that optimal topologies always have a tree structure. More surprising, however, is that the decrease in optimal network size occurs gradually, over several orders of magnitude in the relative cost of network infrastructure, w. More specifically, we find that optimal network sizes decrease with the power-law $\sim w^{-2/3}$. This same scaling property appears both in the random graphs that we generated for simulation purposes and when we apply our infrastructure design model to a spatial distribution of demand centers taken from data from the country of Senegal. This suggests that when cost is the primary design criterion there is no single optimal size for infrastructure networks; different sizes are likely to be optimal for different locations. Whether local or global networks are preferable depends critically on local conditions.

We do find that this gradual scaling becomes a more sudden transition once reliability is added to the design objectives. When the failure to supply demand after vertex or edge outages is deemed costly (large r), the optimal network is a single interconnected system that spans all nodes for a wide range of values for infrastructure costs w. However, in this case small increases in network costs w can cause a rapid return to small networks being optimal. Also, as the importance of reliability increases, the optimal network topology transitions from being a tree, in which there are no duplicate paths, to a meshed system with substantial redundancy.

Clearly, the model that we used to reach these conclusions is simple and the results leave open a number of questions to be addressed in future work. This model neglects many of the engineering details that are important to particular networks, such as voltages in power networks and the non-linear relationships between demand and flow in traffic networks. Our reliability formulation models only single component outages and does not include a model of restoration, whereas both multiple outages and restoration are important components of resilience—an increasingly important design criteria in modern infrastructure systems. Expanding the model beyond single-component outages would also be important to consider how the possibility of cyber-attacks would change design decisions.

Yet, a key characteristic of complex systems research is to identify simple models that can be studied statistically to reveal useful insight about challenging problems. Hopefully the simple models presented here can be improved in future work to reveal additional properties about the conditions under which different network structures are preferable. This approach might be modified to understand when it would be preferable to operate certain portions of a network as independent microgrids rather than as large interconnected networks. Some version of this approach might also help one to understand why existing large-scale networks have the structures that they do. Assuming that larger networks can be synthesized using this method, it would be valuable to compare the statistical characteristics of the networks resulting from our model to those of actual large-scale power systems, such as the US Eastern Interconnect. Also, it would be valuable in future work to relax our assumption that demand does not vary in time in order to understand how time-varying demand impacts optimal infrastructure network designs. A particularly important goal for future work is to provide analytical results that better explain the behaviors that we see, such as the power-law relationship between transmission cost, w, and network component size, $< n_s >$. Another goal is to modify the model to help us to understand when it would be preferable to operate certain portions of a network as an independent system (such as a microgrid) rather than as a large interconnected network.

One of our original motivations for this work was to explain why different infrastructure networks, such as water and power, often have dramatically different structures. While far from conclusive these

results provide some insights regarding the optimality of different topological configurations, under different cost and reliability conditions. For example, our model suggests that building a highly connected power grid is an optimal design choice only when transmission costs are very small relative to the cost of local power production, or if the value of reliability is very high. The same logic would hold for water distribution networks, in which the value of reliability is also high (but outage probabilities are low), and transportation is very expensive (because water typically needs to be pumped to be transported over long distances). Such networks thus exhibit predominantly local scope, except in extreme cases like Southern California, where local water supplies are scarce, increasing the value of interconnection. The tendency for high cost areas to build long connections to remote regions also appears in our model (see Figure 1). A detailed comparison of the structure of existing water and power networks to the statistics of our model remains for future work.

References

[1] "Factoids: Drinking water and ground water statistics for 2007," tech. rep., US Environmental Protection Agency, March 2008.

[2] "About u.s. natural gas pipelines - transporting natural gas," tech. rep., U.S. Energy Information Agency, 2007.

[3] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure," *IEEE Systems Journal*, vol. 6, no. 4, 2012.

[4] R. Lasseter, "Smart distribution: Coupled microgrids," *Proceedings of the IEEE*, vol. 99, pp. 1074 – 1082, June 2011.

[5] "Distributed generation energy technology capital costs," tech. rep., National Energy Technology Laboratory, 2013.

[6] H. Zerriffi, H. Dowlatabadi, and A. Farrell, "Incorporating stress in electric power systems reliability models," *Energy Policy*, vol. 35, no. 1, pp. 61–75, 2007.

[7] A. Zvoleff, A. S. Kocaman, W. T. Huh, and V. Modi, "The impact of geography on energy infrastructure costs," *Energy Policy*, vol. 37, no. 10, pp. 4066–4078, 2009.

[8] H. Zerriffi, *Rural electrification: Strategies for distributed generation*. Springer Science & Business Media, 2010.

[9] G. Zorpette, "Re-engineering Afghanistan," *IEEE Spectrum*, vol. 48, pp. 30–43, 30 September 2011.

[10] A. Fabrikant, A. Luthra, E. Maneva, C. Papadimitriou, and S. Shenker, "On a network creation game," in *Proceedings of the Twenty-second Annual Symposium on Principles of Distributed Computing*, PODC '03, pp. 347–351, 2003.

[11] Z. Wang, A. Scaglione, and R. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE Trans. Smart Grid*, vol. 1, pp. 28–39, June 2010.

[12] K. Kanskey, *Structure of transportation networks: relationships between network geometry and regional characteristics*. University of Chicago Press, 1960.

[13] P. Haggett and R. Chorley, *Network analysis in geography*. St. Martin's Press, 1969.

[14] M. T. Gastner and M. E. Newman, "The spatial structure of networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 49, pp. 247–252, Jan 2006.

[15] W. Garrison, "Connectivity in the interstate highway system," *Papers in Regional Science*, vol. 6, no. 1, pp. 121–137, 1960.

[16] M. Kurant and P. Thiran, "Extraction and analysis of traffic and topologies of transportation networks," *Phys. Rev. E*, vol. 74, p. 036114, Sep 2006.

[17] S. Chan, R. Donner, and S. Lämmer, "Urban road networks – spatial networks with universal geometric features?," *The European Physical Journal B*, vol. 84, no. 4, pp. 563–577, 2011.

[18] S. Wilkinson, S. Dunn, and M. Shu, "The vulnerability of the European air traffic network to spatial hazards," *Natural Hazards*, vol. 60, no. 3, p. 036114, 2012.

[19] R. Lambiotte, V. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. V. Dooren, "Geographical dispersal of mobile communication networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, pp. 5317 – 5325, 2008.

[20] M. Barthelemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1–3, pp. 1 – 101, 2011.

[21] A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, "The extreme vulnerability of interdependent spatially embedded networks," *Nature Physics*, vol. 9, pp. 667–672, 10 2013.

[22] B. A. Carreras, D. E. Newman, and I. Dobson, "Does size matter?," *Chaos*, vol. 24, no. 2, pp. –, 2014.

[23] M. Barthelemy, "Spatial networks," *Physics Reports*, vol. 499, pp. 1–101, Apr 2011.

[24] T. C. McAndre, C. M. Danforth, and J. P. Bagrow, "Robustness of spatial micronetworks," *Phys. Rev. E*, vol. 91, p. 042813, Apr 2015.

[25] C. Unsihuay-Vila, J. Marangon-Lima, A. de Souza, J. Perez-Arriaga, and P. Balestrassi, "A model to long-yerm, multiarea, multistage, and integrated expansion planning of electricity and natural gas systems," *IEEE Trans. Power Systems*, vol. 25, no. 2, pp. 1154–1168, 2010.

[26] J. Choi, T. Tran, A. El-Keib, R. Thomas, H. Oh, and R. Billinton, "A method for transmission system expansion planning considering probabilistic reliability criteria," *IEEE Transactions on Power Systems*, vol. 20, pp. 1606–1615, Aug 2005.

[27] C. B. Sanchez, R. Bent, S. Backhaus, S. Blumsack, H. Hijazi, and P. van Hentenryck, "Convex optimization for joint expansion planning of natural gas and power systems," in *49th Hawai'i Int'l. Conf. Sys. Sci.*, pp. 1–10, Jan 2016.

[28] E. Fisher and a. M. F. R. O'Neill, "Optimal transmission switching," *IEEE Trans. Power Systems*, vol. 23, pp. 1346–1355, Aug 2008.

[29] K. Hedman, M. Ferris, R. O'Neill, E. Fisher, and S. Oren, "Co-optimization of generation unit commitment and transmission switching with n-1 reliability," *IEEE Trans. Power Systems*, vol. 25, pp. 1052–1063, May 2010.

[30] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, "Rules for biologically inspired adaptive network design," *Science*, vol. 347, pp. 439–442, 2010.

[31] J. Dall and M. Christensen, "Random geometric graphs," *Physical Review E*, vol. 66, no. 1, p. 016121, 2002.

[32] E. Martinez-Cesena, P. Mancarella, M. Ndiaye, and M. Schläpfer, "Using mobile phone data for electricity infrastructure planning," *arXiv preprint arXiv:1504.03899*, 2015.

[33] "Data For Development (D4D) Senegal Challenge." http://www.d4d.orange.com. Accessed: 2015-08-30.

Acknowledgement

The authors gratefully acknowledge the hospitality of the Santa Fe Institute in Santa Fe, NM, USA, where Hines and Blumsack were sabbatical visitors in 2014-2015 and where much of this work was completed. The authors would also like to acknowledge Christa Brelsford, Luis Bettencourt, and participants at the Santa Fe Institute workshop, "Reinventing the Grid" for helpful suggestions and discussions. P.H. was supported by NSF awards ECCS-1254549 and DGE-1144388, and DTRA award HDTRA110-1-0088. S.B. acknowledges support from NSF award CNS-1331761. M.S. was supported by the Army Research Office Minerva Programme (grant no. W911NF-12-1-0097).