ELECTRICAL POWER

Data-driven modeling of solar-powered urban microgrids

10.1126/sciadv.1500700

Arda Halu,[1] Antonio Scala,[2,3] Abdulaziz Khiyami,[4] Marta C. González[1]*

Distributed generation takes center stage in today's rapidly changing energy landscape. Particularly, locally matching demand and generation in the form of microgrids is becoming a promising alternative to the central distribution paradigm. Infrastructure networks have long been a major focus of complex networks research with their spatial considerations. We present a systemic study of solar-powered microgrids in the urban context, obeying real hourly consumption patterns and spatial constraints of the city. We propose a microgrid model and study its citywide implementation, identifying the self-sufficiency and temporal properties of microgrids. Using a simple optimization scheme, we find microgrid configurations that result in increased resilience under cost constraints. We characterize load-related failures solving power flows in the networks, and we show the robustness behavior of urban microgrids with respect to optimization using percolation methods. Our findings hint at the existence of an optimal balance between cost and robustness in urban microgrids.

INTRODUCTION

As large-scale empirical data on technological and social systems have become more ubiquitous, the study of such systems in the form of complex networks (*1*–*3*) has gained considerable momentum. Critical infrastructures have traditionally been a major focus of networks research, owing to their vital importance to the functioning of modern societies. Among critical infrastructures, power grids have been studied extensively in the past 15 years in the context of complex networks. The high-voltage transmission portion of power grids has been fairly well studied in terms of structural characteristics (*4*, *5*), robustness against intentional attacks (*6*, *7*), self-organized criticality, and cascades of blackouts (*8*). More recently, they have been modeled as multiple layers of interdependent networks, and it has been shown that interdependency results in cascading failures (*9*, *10*). The effect of spatial embeddedness on complex networks with a focus on power systems has shown that spatially embedded interdependent networks undergo an abrupt collapse regardless of the coupling strength (*11*). Load-based cascading failures on power grids can be studied using simple load redistribution mechanisms with initially assigned capacities (*12*–*16*). The effect of load growth and power fluctuations on the abruptness of failures has also been investigated using direct current (DC) power flow (*15*), and flow optimization has been investigated on resistor networks (*16*). In this context, self-healing methods have been proposed for the quick recovery of distribution networks (*17*) and the effects of power-communications coupling have been studied. Overall, the new trends have clearly shown the need for introducing a certain level of realism in the statistical modeling of power grids by the inclusion of flow dynamics and conservation laws (*18*).

Although the existing electrical infrastructure has been well studied at the level of transmission grids, we need to take into account the ongoing departure from the present central generation paradigm toward distributed generation (DG), that is, moving toward the production of energy on a local scale using small resources instead of relying on large central facilities. As a response to the ever-increasing demand for electricity (*19*, *20*) and higher rates of urbanization (*21*), DG has emerged as a key player in battling the increasing strain on the existing grid infrastructure as well as the vulnerabilities caused by natural disasters and other unforeseen circumstances, while providing users with clean, affordable, and local energy. However, increased levels of DG penetration, coupled with the intermittency of renewable DG sources, introduce an additional level of instability and power quality problems, which is an imminent issue power grids are facing. Hence, the seamless integration of a large number of DG units in the existing infrastructure requires a systemic approach. This approach is embodied in the microgrid concept (*22*), which is based on the idea of an interconnected and coordinated group of buildings/loads equipped with DGs that aim to match consumption and production on a local level, attaining autonomy from the grid if the need arises (*23*).

The need to accommodate the rising urban demand in a self-sustainable way urges us to propose and study the implementation of urban microgrids. The study of urban microgrids differs from the previous studies concerning power grids in that (i) it involves the medium- and low-voltage distribution grid as the underlying network and (ii) it involves smaller networks (typically of the order of tens of nodes) compared to the larger transmission networks that are traditionally studied. Moreover, spatial heterogeneities throughout urban areas have to be taken into account because microgrids, as any other type of infrastructure network, are spatial networks (*24*) where nodes and links are embedded in a geometric space, thereby putting severe constraints on network topology and characteristics. The shape and efficiency of these spatial networks are also cost-driven (*25*–*27*). The traditional method used to study the robustness of networks, namely, percolation theory (*28*), is concerned with the thermodynamic limit ($N \to \infty$), whereas the robustness study of microgrids involves a regime of far smaller size. A very recent study has tackled the percolation properties of spatial micronetworks with microgrid applications in mind (*29*).

Here, we relate the specific case of urban environments to previous work on electrical infrastructure networks. In particular, we consider the space and demand heterogeneity in urban areas served by the

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Laboratory of Computational Social Science, Networks Department, IMT Lucca Institute for Advanced Studies, 55100 Lucca, Italy. [3]Institute for Complex Systems of the National Research Council, Physics Department, University of Rome "Sapienza," Piazzale Moro 5, 00185 Rome, Italy. [4]Center for Complex Engineering Systems at King Abdulaziz City for Science and Technology and MIT, Cambridge, MA 02139, USA.
*Corresponding author. E-mail: martag@mit.edu

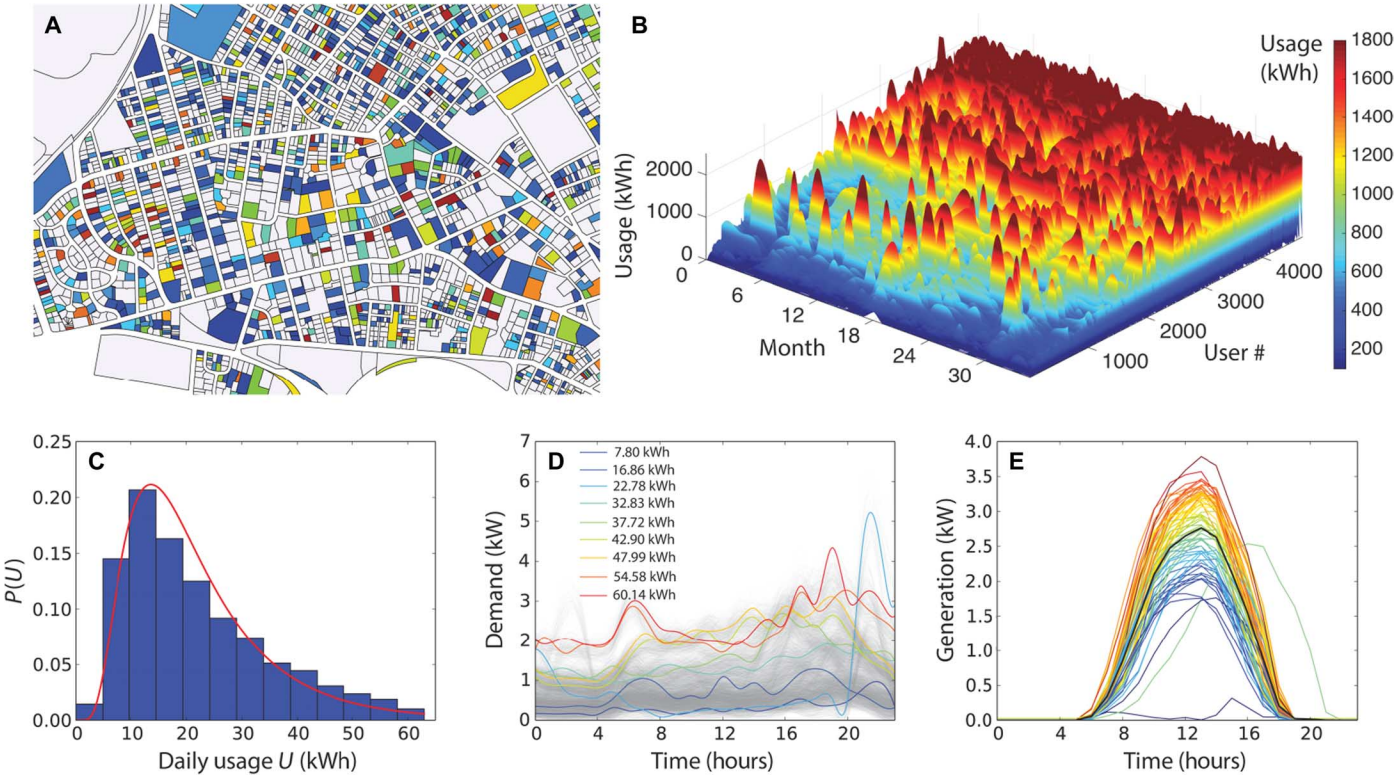

Fig. 1. Temporal patterns of electric energy demand. (**A**) Map of a portion of Cambridge, MA. The colors represent the monthly electricity consumption. (**B**) Monthly electricity consumption of 4683 users over the course of 3 years. (**C**) Distribution of the daily consumption for an average day in July. The solid red curve denotes the lognormal fit. (**D**) Hourly demand profiles for a typical day in July, with representative daily curves marked with colors and respective daily consumption values. (**E**) Hourly solar generation profiles for typical residential-size installations.

distribution grid and investigate the implementation and viability of coordinated networks of DG in the form of solar-powered microgrids. In doing so, we bring together the unique aspects of spatial networks, percolation theory, power flow dynamics, and the temporal evolution of real consumption data for a fundamental treatment of microgrids at the systems level, taking a cross-disciplinary approach between power systems engineering and the physics of complex networks. Remaining true to user demand, we regard the citywide implementation picture to quantify the self-sufficiency and flow dynamics of microgrids throughout the day. We propose optimal topologies that manage congestion and increase resilience while staying within realistic grid constraints, showing the competition between cost and resilience. Finally, we explore their robustness against load-based failures and find two distinct regimes as a function of an optimization parameter α. Our simulations thus suggest an optimal trade-off between cost and robustness in microgrids.

Description of data

To model the demand and generation profiles of urban microgrids, we use two sources of data. The first model is comprised of the monthly electric bills of 4683 Cambridge, MA accounts over the course of 36 months and is obtained from NSTAR, the electricity and gas utility in Cambridge (*30*). These data are geolocated by parcel centroids using the geographic information system (GIS) data of Cambridge. For the scope of this work, accounts are chosen such that they represent the single-family, residential portion of the user base where there is only one account per parcel (Fig. 1A). They subsequently fall within the residential/small-commercial monthly usage range of 200 to 2000 kWh (Fig. 1B). The distribution of these usage values, $P(U)$, for an average day in July is shown in Fig. 1C and follows a lognormal distribution

$$P(U) = \frac{1}{U\sigma\sqrt{2\pi}} e^{-(\ln U - \mu)^2/(2\sigma^2)}$$

with a mean (μ) of 2.94 and an SD (σ) of 0.57.

The second part of the data is the high–temporal resolution smart grid data provided by the Pecan Street Research Project (*31*). This data set consists of 17 months of aggregated and disaggregated usage and solar generation data collected from 70 participating homes, with 15-min-interval smart-meter readings. The hourly demand and generation profiles are calculated by averaging the 15-min-interval readings in 60-min windows. Similar to the user data set of Cambridge, the typical monthly usage range of Pecan Street users lies within the 200- to 2000-kWh range. To get an overview of the monthly user behavior within this range, we plot and superimpose the distributions of monthly usage of the Cambridge and Pecan Street users. Despite the different locations (Cambridge and Austin), we observe an agreement between the two distributions for nonsummer months, hinting at a similar aggregated usage behavior on average (see section S1).

This similarity in usage between the two data sets for residential users enables us to model the daily demand profiles of the 4683 Cambridge users after the Pecan Street users. To do this, we bin both data sets into monthly usage ranges, and then for each user in the

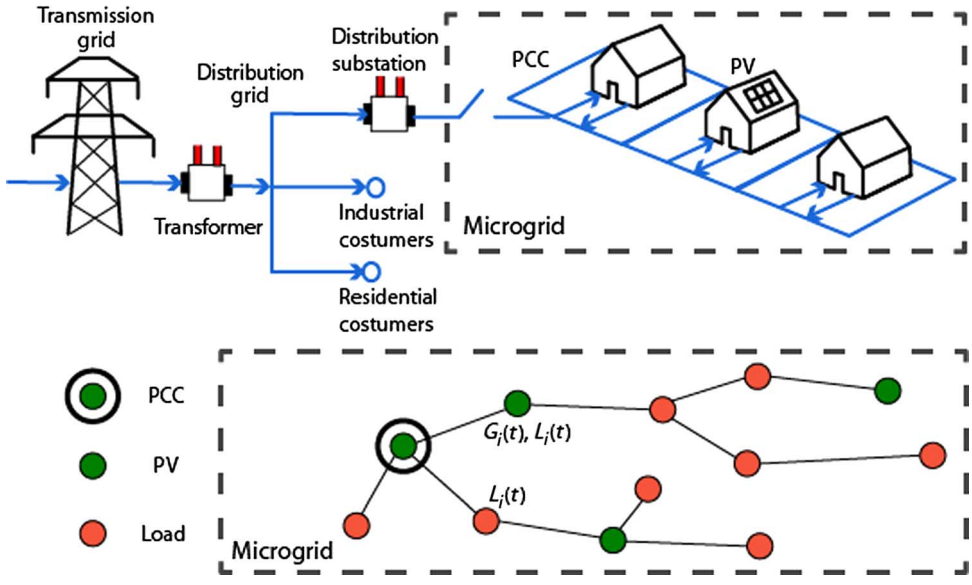

Fig. 2. Microgrid and its network representation. The microgrids are part of the distribution grid, downstream of distribution substations. Users with and without solar PVs are modeled as load and generator nodes, respectively, equipped with smart grid electronics to govern bidirectional flows and voltage fluctuations. The microgrid is connected to the distribution network at one point via the point of common coupling (PCC).

Cambridge data set, we randomly pick one of the users in the Pecan Street in the same usage range and assign the hourly demand of that Pecan Street user to the Cambridge user, scaling up or down by a constant factor to match the exact monthly usage of the Cambridge user. To introduce additional stochasticity to the assignment, we add noise in the form of $d_i(t) = d_i^0(t)(1 + \beta\xi_i)$, where $d_i^0(t)$ is the initial demand of user i at time t, $d_i(t)$ is the demand of user i at time t with added noise, where ξ_i is a random variable uniformly distributed in [−1, 1], and $\beta = 0.2$ to keep the Cambridge demands within 20% error relative to the original Pecan Street user demands. This results in a general picture of Cambridge average hourly demands where the average hourly demand of Pecan Street is conserved and variability is introduced between users (Fig. 1D). The hourly demands of Cambridge accounts used in the power flow calculations of microgrids are sampled from these profiles throughout the paper. The solar generations are similarly sampled from Pecan Street generations (Fig. 1E); however, in this case, there is much less variability between users' generation profiles as the typical residential solar photovoltaic (PV) system installation sizes do not show much variation.

Microgrid model

In our framework, microgrids are modeled as part of the existing electricity distribution network, downstream of distribution substations (Fig. 2). We assume the microgrids to be equipped with solar PV as distributed energy resources and with a smart grid architecture that is capable of managing bidirectional flows in the microgrid. To formalize our model, we follow the convention of representing buses as nodes and distribution lines connecting the loads as links. Each node i is either a load bus or a generator bus depending on whether it has generation capability. We represent the net real power injected by node i at time t as $P_i(t) = G_i(t) - L_i(t)$, where $G_i(t)$ and $L_i(t)$ are the generation and loads at time t, respectively. The power flow F_{ij} on each link $ij \in E$ is given by DC power flow calculations (see Materials and Methods). The microgrid is connected to the utility grid at one point, the PCC. For the purposes of our simulations, where the aggregate demand on the PCC is considered, we randomly place the PCC in the network. However, we note that a detailed knowledge of the actual distribution network topology can inform one's decision in PCC site selection for more realistic results. We assume that the voltage-related assumptions of DC power flow (small voltage angle differences, flat voltage profile) are ensured by centralized or decentralized control strategies (*32*–*34*) that are able to react to system conditions in real time, either by voltage control devices at the PCC or within the microgrid by the power electronics on each node. Here, we can assume that the reactance-to-resistance (X/R) ratios in the microgrid infrastructure are high enough for the DC approximation (*35*–*38*) to be valid. See section S2 for a sensitivity analysis of the alternating current (AC) versus DC comparisons on a realistic building microgrid configuration with real demand and generation values.

When building the microgrids, we respect the low- and medium-voltage distribution network topologies, which are radial, or open- or closed-loop networks with laterals, that is, tree networks with or without a small number of meshed connections that can be switched on and off during emergencies. The simplest topology of distribution networks is the radial topology. Although it is the least costly topology, it is also the least likely to work reliably with the increase of the share of DG (*39*). Closed-loop topologies, on the other hand, offer a good compromise between cost and reliability by offering a few small-world links that go into effect when a part of the network becomes disconnected.

RESULTS

Citywide implementation of microgrids

For a citywide implementation of microgrids, we partition the single-account parcels into microgrid neighborhoods. In the partitioning, we assume that the cost related to line lengths is the most dominant factor

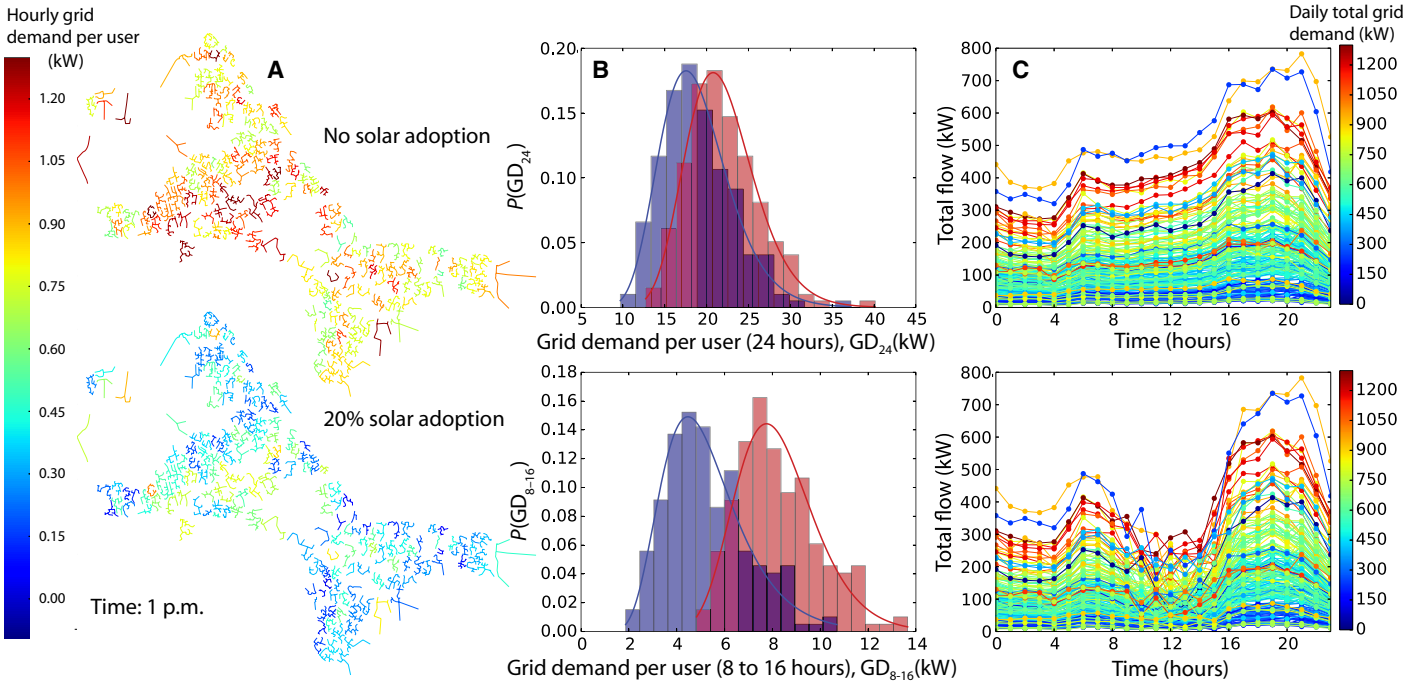

Fig. 3. The role of microgrids in consumption of electric energy. (**A**) Proposed microgrids in Cambridge. The colors denote the grid demand per user in the microgrid at 1 p.m. for the no–solar adoption (top) and the 20% solar adoption (bottom) case. (**B**) Distributions of grid demand for the 200 microgrids for the whole day (top) and daylight hours (bottom) for the no-adoption (red) and the 20% adoption (blue) case. The red and blue solid lines indicate the lognormal fits for the no-adoption and 20% adoption cases, respectively. (**C**) Total flow in the microgrid as a function of time of day for no solar adoption (top) and 20% solar adoption (bottom). The colors denote the total daily grid demand of each microgrid.

driving network topologies (*27*). Hence, we use a *k*-means algorithm to cluster parcels into microgrids spatially, which provides a good compromise between network size (in number of nodes) and spatial size (in meters). In our *k*-means partitioning, we set the number of clusters to 200, which results in a microgrid size distribution centered around 25 users (see section S3). The clustering is run 1000 times, and the partition with the lowest distortion, defined as the sum of the squared differences between the observations and the corresponding parcel centroid, is returned. Once the nodes in a microgrid are determined with *k*-means, we construct minimum-cost distribution networks between them. To have a realistic approximation of distribution networks, in agreement with the open- or closed-loop radial networks with laterals, we construct tree topologies. The tree topologies are constructed as Euclidean minimum spanning trees (EMSTs); that is, first, the Delaunay triangulation is defined as the underlying spatial topology between nodes and then the minimum spanning tree that minimizes the total distance in the network is calculated from this network. This gives us an approximation of small microgrid topologies that are part of the distribution grid (Fig. 3A). We sample the hourly demands of each node from the Cambridge demand profiles modeled after smart grid data. Moreover, we define a solar PV adoption rate of 20% and choose generator nodes independently according to this adoption rate. The generator nodes have their solar generation profiles sampled from Pecan Street solar generation data. We assume that the typical solar PV installation size of the homes in Cambridge urban microgrids is similar to that of houses in the Pecan Street range, in the typical residential range of 3 to 5 kW.

Although microgrids, by definition, can operate in both grid-connected and off-grid modes, because we consider only solar generation without grid storage for the scope of this work, we assume the microgrid to be in grid-connected mode at all times. This means that depending on the solar adoption ratio of a microgrid, and depending on the hour of the day, the microgrid can either import power from the grid or export power to the grid. As one of the main objectives of microgrids is self-sufficiency and cost savings for the user, we quantify the self-sufficiency of the microgrid in terms of the grid demand per user as a function of time

$$\mathrm{GD}^M(t) = \frac{1}{N_M} P^M_{\mathrm{PCC}}(t) \quad (1)$$

where N_M is the number of users in microgrid M and

$$P^M_{\mathrm{PCC}}(t) = \sum_{i \in M} P_i(t) \quad (2)$$

is the total grid demand of microgrid M at time t. The reasoning behind looking at the grid demand per user is that (i) because the generation is shared in each microgrid in a collaborative manner, the collective benefit of its users must be considered, and (ii) because microgrids have a different number of users, a normalized metric is needed.

In Fig. 3A, we show the hourly grid demand per user for all Cambridge microgrids at 1 p.m., which is the peak solar production time of the day. We use the no–solar adoption case as a baseline for comparison with the 20% solar adoption rate. We observe that, overall, the city shows a wide variability in grid demand of microgrids. Moreover, under 20% adoption, most microgrids are able to reduce their grid demand to almost zero, attaining self-sufficiency at midday. That said, we also observe some microgrids that remain dependent on the grid

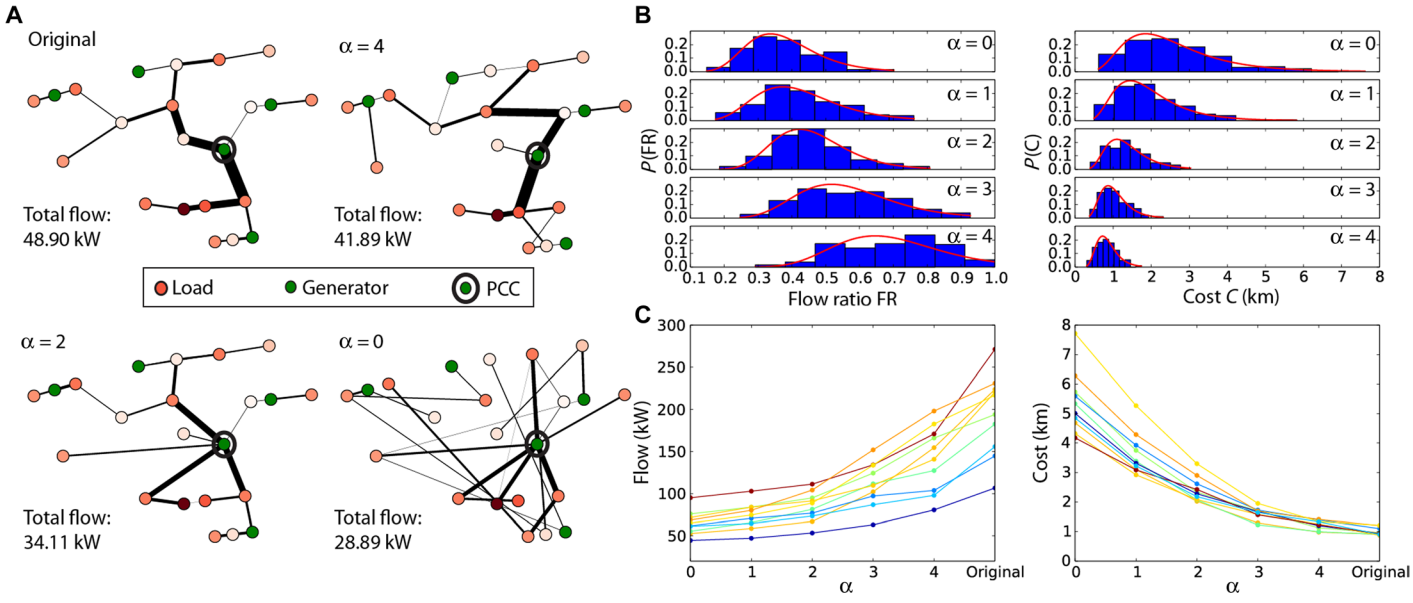

Fig. 4. Effects of microgrid configurations in costs and energy flow. (**A**) Example microgrid and the resulting optimized topologies for different rewiring parameters α. The darker shade of red denotes higher load. (**B**) Distributions of flow ratio and cost for different values of α over all microgrids in Cambridge. The solid red lines indicate the lognormal fits. (**C**) Total flow and cost with respect to different values of α for the 10 largest microgrids in Cambridge.

despite their solar production, mainly attributed to their high levels of consumption. To measure the statistics of the grid demand per user, we plot in Fig. 3B the distributions over all microgrids for no-solar and 20% solar scenarios, for the whole day (24 hours) and during daylight hours (8 to 16 hours). We see that both distributions are lognormal, and as can be expected, the difference between the distributions of no solar and 20% solar becomes more pronounced during daylight hours. Nevertheless, the overall daily savings is still considerable between the two cases, amounting to nearly 5 kW of daily savings on average for all microgrids.

As an important proxy of their resilience, we calculate the total power flow in each microgrid given the real demands and generations of users in them, as a function of time (Fig. 3C). This enables us to monitor the time evolution of each microgrid's total flow and gives us an insight as to what level of flow they are expected to sustain on a typical day. As we expect, the total flows in the no-solar scenario roughly follow the average daily trend of individual demands, with two major increases throughout the day in the morning and evening hours, corresponding to when people wake up and come home from work. In the 20% solar adoption case, however, we see the alleviating effect of solar production on the flows during daylight hours, considerably decreasing the total power flow in the microgrids. This is because the solar production of homes with PV first meets their own demands locally, decreasing their inflows from their network neighbors in the microgrid. Unlike the curves in the no-solar case, the flows in the 20% adoption case have a more erratic trend during daylight hours, where some microgrids are able to decrease their total flow considerably, whereas others have a smaller flow decrease because of their solar production. Furthermore, we color code the microgrids according to their daily total demand from the grid. Here, we observe that although most microgrids' total power flow is proportional to their grid demand, there are also some outlier microgrids that are able to sustain lower levels of total flow while being highly dependent on the grid and some that are more self-sufficient, yet have higher power flows.

Optimizing microgrids for resilience and congestion mitigation

Congestion in power systems poses one of the greatest challenges to power infrastructures as the demand and the share of renewables in the grid increases. Although the term "congestion" is more generally used for the transmission grid, it also applies to the distribution grid. The two main approaches in congestion management are (i) the investment in additional grid infrastructure and (ii) demand side management. The congestion mitigation strategy considered in this work for microgrids is the increase of transfer capacity by new cables, by a larger transformer, or by reconfiguring the topology of the grid. Although demand response is currently a very hot topic and has several applications, coming from the complex networks perspective, here we take the latter approach. The modification of the distribution grid can be done permanently either by the addition of lines (overhead or underground) or with a more modern approach, by automatic breakers that allow remote and automated reconfiguration (*40*). The existing distribution grid topologies are far from optimal, considering their historical evolution and the exponential growth of demand.

Our aim is to propose topologies that mitigate congestion in microgrids by minimizing the total absolute power flow in the network by making minimal adjustments to the topologies given the demand and generation on each node; that is, we seek

$$\min \sum_{ij \in E} |F_{ij}| \qquad (3)$$

We propose a network rewiring scheme that uses a limited "budget" (number of extra lines to add) and a single parameter α that modulates the dependence of the probability of making a connection between

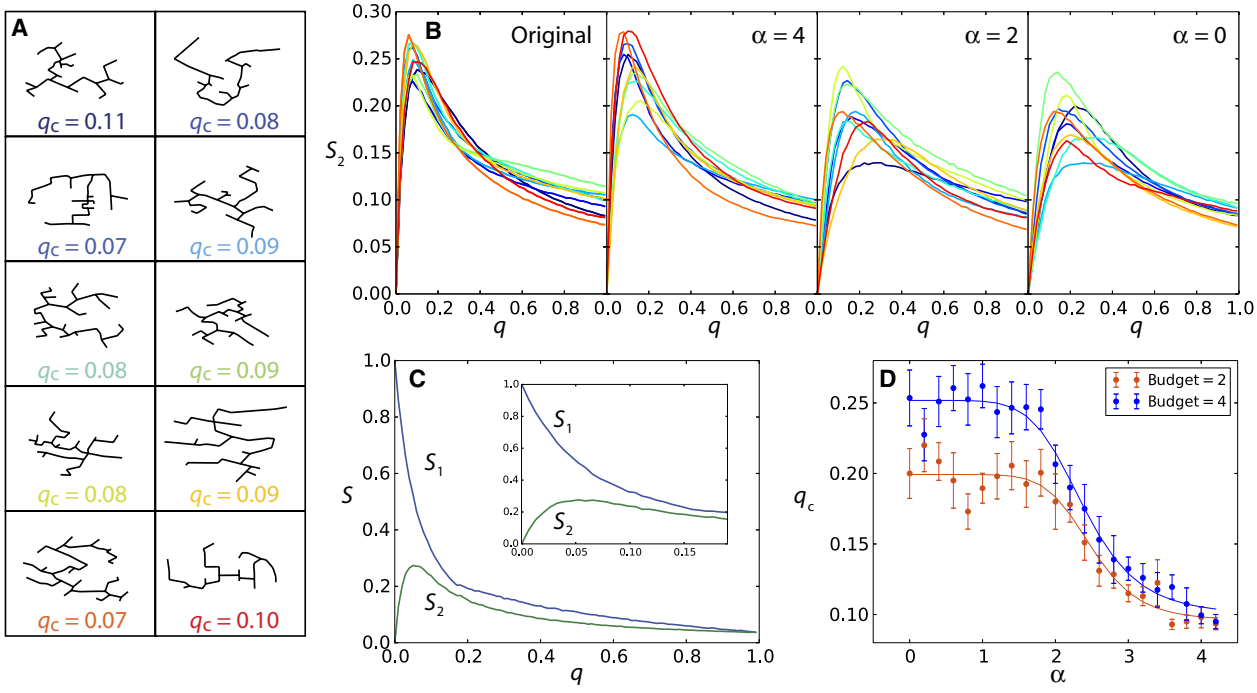

Fig. 5. Effects of microgrid configurations on their resilience. (**A**) Topologies and q_c values of the 10 largest microgrids in Cambridge, MA. (**B**) Size of the second largest connected component S_2 with respect to q for different values of α. The microgrids are color-coded according to (A). (**C**) Size of the largest and second largest connected component, S_1 and S_2, respectively, as a function of q for a sample synthetic microgrid of size $N = 50$. (**D**) Percolation threshold q_c as a function of α for ensembles of synthetic microgrids of size $N = 50$, averaged over 20 realizations. The solid lines follow Eq. 6.

nodes i and j on the distance between them such that $p(e_{ij}) = d^{-\alpha}_{ij}$, where e_{ij} is the edge between i and j, and d_{ij} is the Euclidean distance between i and j, similar to Li *et al.* (*41*). This scheme enables us to account for the spatial constraints in building new infrastructure. A high value of α strictly prohibits the addition of long-range links, resulting in a spatial network, whereas an α close to zero effectively allows for any connection regardless of the distance, resulting in a random-like network.

In our simulations, we set our budget to 2; that is, we add two lines at the beginning of the simulation. The optimization problem is solved using simulated annealing where topologies minimizing the total absolute power flow are sought, given a certain α (see Materials and Methods). For each optimized topology, we measure the cost C of building the network in terms of the total length of links and the congestion mitigation as the flow ratio FR, defined as the ratio of the final total power flow to the initial one; that is

$$C = \sum_{ij \in E} d_{ij}, \quad \mathrm{FR} = \frac{\sum_{ij \in E} |F_{ij}|^{\mathrm{final}}}{\sum_{ij \in E} |F_{ij}|^{\mathrm{init}}} \tag{4}$$

In Fig. 4A, we show the results of this optimization on a sample microgrid for different values of α, compared to the original topology. The effect of lowering α is manifested as a less spatially constrained network. This effect is accompanied by the observation that, with lower α, demand nodes are more free to connect directly to generation nodes. We see that this effect is the strongest in the dependency to connect to the main supplier, which is the PCC at 1 p.m. for the specific microgrid represented in the figure. For solar microgrids without storage that are considered in the network, the microgrids usually have to make up for their deficit from the PCC, which results in a more central topology as shown here. However, we note that this need not be the case where demand and generation are matched more evenly and the demand nodes are free to connect to closer generation nodes.

To show the competing effect of cost and total flow as a function of α, we plot the distributions of these metrics over all 200 microgrids in Cambridge for different α values in Fig. 4B. It is clear from the figures that decreasing α to zero results in a considerable shift of the flow ratio toward zero, meaning that as the spatial constraints are relaxed in the optimization, congestion levels in the network are reduced considerably, resulting in a more resilient setup. On the other hand, this effect is countered by an increase in the building cost of the microgrid. For α = 4, which is similar to the initial topology, the building costs have a mean of less than 1 km with little spread, whereas α = 0 results in microgrids that have a mean around 2 km with a much greater spread. Both cost and flow ratio distributions are well approximated by lognormal distributions.

For a closer look into how individual microgrids behave in terms of flow ratio and cost as a function of α, we plot these values for the 10 largest microgrids in our microgrid ensemble in Fig. 4C. Here, we clearly see the increase in flows and decrease in costs as α is increased. This points to the fact that microgrids constructed in this manner are amenable to optimization with respect to cost and total flow.

Robustness of microgrids against overload failures

Resilience refers to the systems' ability to recover from failures, whereas robustness is usually used to mean their ability to survive failures. In the previous section, we have assessed the resilience of microgrids in terms

of their total power flow. In this section, we study the robustness of microgrids to overload failures. The robustness of small spatial networks such as our proposed microgrids has recently been investigated from the perspective of physical distances, under random failures governed by a bond percolation process (*29*). Here, we follow a similar approach, but rather than adopting the "falling tree" approach and assuming a link failure probability proportional to the link length, we define the link failure probability Q_{ij} that is dependent on the flow F_{ij} on the link such that

$$Q_{ij} = q\frac{F_{ij}}{\langle F_{ij}\rangle} \tag{5}$$

where $\langle F_{ij}\rangle = \Sigma_{ij\in E}F_{ij}/M$, M is the number of edges, and q is the traditional bond percolation parameter denoting the fraction of nodes that fail on average. That is, link failures are attributed to overloaded lines (*15*). We use the size of the second largest connected component S_2 to estimate the percolation threshold q_c. In particular, the q corresponding to the maximum of S_2 is a proxy for the percolation threshold q_c. Here, we note that although, even in the DC approximation, percolation analysis for robustness offers a lower bound on the actual network damage and in reality the purely structural removal of links can potentially cause further network-wide cascades of overload failures, it nevertheless provides a rich quantitative picture on the microgrid perspective.

In Fig. 5A, we show the topologies of the 10 largest microgrids in Cambridge as visual cues, along with their percolation threshold q_c as estimated from S_2. We see that these microgrids have percolation thresholds in the narrow range 0.07 to 0.11. This finding shows that the original configurations of these microgrids are highly vulnerable to overload failures, as they are both spatial and their topologies are not optimized for lower levels of flow. We therefore optimize the 10 largest microgrids for each α and perform simulations on the resulting networks to see how our optimization scheme minimizing the total flow helps with microgrids' robustness. We plot S_2 as a function of q for different α values in Fig. 5B. We see that as α is decreased to zero, the percolation threshold q_c increases, resulting in more robust configurations. We remark that the increase in the robustness is notable between α = 2 and α = 4. We also observe some variability in their capability to increase their robustness with α. This is likely due to the specific topologies and demand/generation distributions of each microgrid, although the behavior with respect to α follows the same pattern. Next, to serve as a baseline for comparison with the real microgrids in Cambridge, we construct synthetic microgrids of size $N = 50$. We uniformly place the 50 nodes randomly over an area compatible in size with the Cambridge microgrids and construct the Delaunay triangulation and calculate the EMST as before. We then sample the demands and generations from the demand and generation distributions of Cambridge. For the original (unoptimized) topologies of these synthetic microgrids, we plot the size of the largest connected component S_1 and the second largest connected component S_2 as a function of q in Fig. 5C. We see that these networks are again highly fragile because of their spatial structure. For the microgrid shown in the figure, $q_c = 0.06$, which is slightly lower than the 10 largest microgrids in Cambridge but generally in agreement with their range. Moreover, as noted by McAndrew *et al.* (*29*), we see a departure from the q_c value predicted by the theory for random graphs with given $\langle k\rangle$ and $\langle k^2\rangle$, that is, the value when $\langle k^2\rangle/\langle k\rangle = 2$ (*42*). Over the entire range of α considered here, the theoretical q_c values run higher (between 0.20 and 0.40 on average, see section S4) than the ones approximated by S_2. S_2 is a good estimator for large networks, whereas our microgrids are much smaller in size, resulting in finite-size corrections. This is further compounded by the fact that these networks are built under the conditions of the optimization and are not connected at random as in the theory. Finally, we plot q_c as a function of α for the synthetic microgrids in Fig. 5D to see the detailed behavior of robustness with respect to optimization. We see that as we lower α from 4 to 0, the robustness reaches a plateau after 2, pointing to the fact that even though lower α values favor lower total power flows and therefore fewer failed lines, there is not much difference in robustness between 2 and 0, although we had shown that the cost of building the microgrid continuously increases. That is, α = 2 provides an optimal trade-off between robustness and cost. The data can be fit with a four-parameter logistic function in the form

$$q_c(\alpha) = \frac{q_c^{\text{nonspatial}} - q_c^{\text{spatial}}}{1 + (\alpha/A)^B} \tag{6}$$

where $q_c^{\text{nonspatial}} = 0.199$ is the maximum asymptote, $q_c^{\text{spatial}} = 0.096$ is the minimum asymptote, $A = 2.497$ is the inflection point, and $B = 7.720$ is the steepness, for a budget of two lines. The same behavior persists for different budget values, albeit with higher maximum asymptotes. Moreover, we observe that this change of regime near α = 2 coincides with the border between long-range and short-range behavior. In particular, the average distance $\langle r\rangle$ reached by a link in our case is proportional to $\int r^{-\alpha+1}dr$, which is finite for α > 2 and divergent for α ≤ 2. Thus, α > 2 signifies local (short-range) behavior with lower robustness, whereas α ≤ 2 means a more small-world (long-range) behavior with higher robustness.

DISCUSSION

Here, we have presented a comprehensive study on the feasibility of solar-powered urban microgrids. By using the citywide demand and generation profiles modeled after smart grid data, we have extended the implementation of microgrids to the entire city to study self-sufficiency and identify the savings in the proposed microgrids. We have offered insights into optimal microgrid topologies that help mitigate congestion and quantified the competing behavior of cost and resilience with respect to the optimization parameter. Furthermore, we have investigated the robustness of these microgrids to load-based failures using percolation theory. We have shown the existence of two regimes in the robustness of microgrids as a function of the optimization parameter, pointing toward a switch from short-range to long-range behavior and an optimal balance between cost and robustness.

The study of microgrids is new to network science as it involves small networks and takes place on the distribution grid. The proposed modeling framework captures the essential requirements and characteristics of microgrids such as spatial constraints, power flow equations, and realistic topologies while building on previous work in network science on power systems such as percolation and congestion mitigation. Hence, our aim is to bridge the gap between the large-scale systems approach of complex networks and detailed approaches of engineering.

There are many ways this work can be extended. When building microgrids, the partition of the entire city into microgrids can not

only be made spatially but also based on demand. Data on actual distribution grid topologies can easily be incorporated. Furthermore, we limit ourselves to only solar generation within the scope of this work, whereas solar generation, because of its intermittent nature, should always be coupled with electrical storage within the microgrid context if off-grid operation is desired. Further models of solar-powered urban microgrids can incorporate grid storage elements. The detrimental effects of excessive grid export can be explored in the high renewable penetration regime. Finally, as smart grid projects become more widespread, real-time demand data can be used to propose dynamical models of urban-scale microgrids, paving the way for recommendation systems for consumers, utilities, and regulators alike.

MATERIALS AND METHODS

DC power flow

In our microgrid simulations, we used the DC power flow approximation to calculate power flows between consumer and generator nodes. In complex networks terminology, this corresponds to the flow dynamics between sources and sinks. The DC power flows between nodes i and j at time t

$$F_{ij} = \frac{\theta_j(t) - \theta_i(t)}{x_{ij}} \tag{7}$$

were calculated according to

$$P^{\mathrm{DC}}(t) = B\theta(t) \tag{8}$$

where $P_i^{\mathrm{DC}}(t) = \sum_j F_{ij}(t)$ is the power on node i at time t, θ_i is the voltage phase of node i at time t, and B is an $N \times N$ matrix whose elements are determined by the line reactances x_{ij} such that

$$B_{ij} = -1/x_{ij}, B_{ii} = \sum_k 1/x_{ik} \tag{9}$$

Each microgrid was assumed to be connected to the utility grid at one point, namely, the PCC. We envisaged a dynamic control system that can govern, using smart switches, the bidirectional flow between the two channels: the microgrid and the utility (distribution) grid, while obeying voltage and frequency regulation concerns. This inherent interaction of the microgrid and the central grid comes into play in the load-balancing part of the DC power flow calculation, where load and generation are matched in every connected component so that

$$\sum_i G_i - L_i = \sum_i P_i = 0 \tag{10}$$

which is the criterion that ensures that the DC power flow equations have a unique solution. This requirement of DC power flow forms the basis of the interactions between the microgrid and the distribution grid. At any time t in the day, before calculating the DC power flow on each line in the network, we balanced load and generation in the network in two ways:

(i) If $\sum_i L_i(t) > \sum_i G_i(t)$, feed from the utility grid. In this case, the available generation is used up by the microgrid and the PCC acts as a generator to make up for the deficiency in generation. Aside from its own load $L_{\mathrm{PCC}}(t)$ and generation $G_{\mathrm{PCC}}(t)$ (if available) at time t, it draws an amount equal to the difference

$$P_{\mathrm{PCC}}(t) = \sum_i L_i(t) - \sum_i G_i(t) \tag{11}$$

from the utility grid.

(ii) If $\sum_i G_i(t) > \sum_i L_i(t)$, feed into the utility grid. In this case, the existing load is completely met by the microgrid and the PCC acts as an extra load to relieve the microgrid of the excess in generation. Aside from its own load $L_{\mathrm{PCC}}(t)$ and generation $G_{\mathrm{PCC}}(t)$ (if available) at time t, it feeds an amount equal to the difference

$$P_{\mathrm{PCC}}(t) = \sum_i G_i(t) - \sum_i L_i(t) \tag{12}$$

into the utility grid.

Hence, at each time t, that is, on an hourly basis, we assumed a basic central control mechanism, compatible with the current research on microgrid control, where a central microgrid switch can collect the total load and generation information from the microgrid and adjust the grid interaction of the PCC with the microgrid according to the simple set of rules above.

Simulated annealing and rewiring scheme

The initial topologies of the microgrids were constructed according to the rules described in the respective sections. After that, optimized topologies were sought by the limited addition and subsequent rewiring of links. The number of links to be added was constrained by the budget, which was fixed as 2 in our simulations. Both the additions and the rewirings were made with probability proportional to the Euclidean distance between the nodes, such that $p(e_{ij}) = d_{ij}^{-\alpha}$, where the parameter α modulates the dependence of the addition/rewiring on distance. Nodes i and j were chosen to avoid self-loops and multiple edges. In the rewiring, it was ensured that the graph remained connected for the DC power flow equations to have a solution; that is, links that left the graph disconnected were not rewired. We note that the initial addition of lines within the budget is required in tree topologies as rewiring only is more likely to leave these networks disconnected.

The optimization method used in searching for feasible topologies was simulated annealing. In our implementation of simulated annealing, the Hamiltonian we sought to minimize was the amount of total absolute power flow $|F_{ij}|$ over all the lines $ij \in \mathrm{E}$ in the network

$$H = \sum_{ij \in E} |F_{ij}| \tag{13}$$

while keeping the original demand $L_i(t)$ and generation $G_i(t)$ values of the nodes. We sought to find configurations where congestion was minimized, and hence, the resilience of the networks was improved. Our simulated annealing schedule consisted of one addition or rewiring at each step. Starting from a temperature $T = 1$, the system was cooled down exponentially such that $T(t + 1) = T(t) \times 0.95$ until $T = 0.01$, with 10^3 Monte Carlo steps at each temperature. The link addition or rewiring was accepted if it lowered H, but if it did not, it was accepted with probability $e^{-\Delta H/kT}$ and rejected otherwise, according to the Boltzmann criterion. The network configuration with the lowest H was retained at the end of the simulated annealing process, which,

in most cases, was the same as the final configuration reached at the end of the cooling process (after 9×10^4 steps); that is, our simulated annealing algorithm was able to find the global minimum.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/1/e1500700/DC1

S1. Cambridge and Pecan Street data set selection

S2. DC versus AC power flow

S3. Microgrid size distributions

S4. Theoretical q_c as a function of α

Fig. S1. The monthly usage distributions of Cambridge for January (triangles) and July (circles).

Fig. S2. The hourly demand profiles of Pecan Street users for 17 months from December 2012 to April 2014.

Fig. S3. The hourly solar PV generation profiles of Pecan Street users for 17 months from December 2012 to April 2014.

Fig. S4. The monthly usage distributions of Cambridge in July (circles) and Pecan Street for 17 months (squares) from December 2012 to April 2014.

Fig. S5. The topology of the building microgrid used in the power flow sensitivity analysis.

Fig. S6. The % error (P_{err}) between AC and DC power flow as a function of the X/R ratio for the proposed topology for different values of R.

Fig. S7. Size distributions of the microgrids in Cambridge partitioned using k-means.

Fig. S8. Theoretical q_c calculated over the whole range of α values, averaged over 10 realizations.

REFERENCES AND NOTES

1. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
2. M. E. J. Newman, The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
3. S. H. Strogatz, Exploring complex networks. *Nature* **410**, 268–276 (2001).
4. L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, Classes of small-world networks. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149–11152 (2000).
5. R. Albert, I. Albert, G. L. Nakarado, Structural vulnerability of the North American power grid. *Phys. Rev. E* **69**, 025103(R) (2004).
6. R. V. Solé, M. Rosas-Casals, B. Corominas-Murtra, S. Valverde, Robustness of the European power grids under intentional attack. *Phys. Rev. E* **77**, 026102 (2008).
7. C. M. Schneider, A. A. Moreira, J. S. Andrade Jr., S. Havlin, H. J. Herrmann, Mitigation of malicious attacks on networks. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3838–3841 (2011).
8. I. Dobson, B. A. Carreras, V. E. Lynch, D. E. Newman, Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos* **17**, 026103 (2007).
9. S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
10. C. D. Brummitt, R. M. D'Souza, E. A. Leicht, Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E680–E689 (2012).
11. A. Bashan, Y. Berezin, S. V. Buldyrev, S. Havlin, The extreme vulnerability of interdependent spatially embedded networks. *Nat. Phys.* **9**, 667–672 (2013).
12. A. Asztalos, S. Sreenivasan, B. K. Szymanski, G. Korniss, Cascading failures in spatially-embedded random networks. *PLOS One* **9**, e84563 (2014).
13. A. E. Motter, Y.-C. Lai, Cascade-based attacks on complex networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **66**, 065102 (2002).
14. R. G. Morris, M. Barthelemy, Interdependent networks: The fragility of control. *Sci. Rep.* **3**, 2764 (2013).
15. S. Pahwa, C. Scoglio, A. Scala, Abruptness of cascade failures in power grids. *Sci. Rep.* **4**, 3694 (2014).
16. A. Asztalos, S. Sreenivasan, B. K. Szymanski, G. Korniss, Distributed flow optimization and cascading effects in weighted complex networks. *Eur. Phys. J. B* **85**, 288 (2012).
17. W. Quattrociocchi, G. Caldarelli, A. Scala, Self-healing networks: Redundancy and structure. *PLOS One* **9**, e87986 (2014).
18. M. Korkali, J. G. Veneman, B. F. Tivnan, P. D. H. Hines, Reducing cascading failure risk by increasing infrastructure network interdependency. arXiv:1410.6836v3 Preprint (2014).
19. International Energy Agency, World Energy Outlook 2014, Released November 2014.
20. U.S. Energy Information Administration, Annual Energy Outlook 2014, Released May 2014.
21. United Nations Department of Economic and Social Affairs, Population Division, "World Urbanization Prospects," 2014 Revision.
22. N. Hatziargyriou, H. Asano, R. Iravani, C. Marnay, Microgrids: An overview of ongoing research, development, and demonstration projects. *IEEE Power Energy Mag.* **5**, 78–94 (2007).
23. CIGRÉ, WG C6.22 Microgrids Working Group.
24. M. Barthélemy, Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
25. M. T. Gastner, M. E. J. Newman, Shape and efficiency in spatial distribution networks. *J. Stat. Mech. Theory Exp.* P01015 (2006).
26. M. T. Gastner, M. E. J. Newman, Optimal design of spatial distribution networks. *Phys. Rev. E* **74**, 016117 (2006).
27. M. T. Gastner, M. E. J. Newman, The spatial structure of networks. *Eur. Phys. J. B* **49**, 247–252 (2006).
28. D. Stauffer, A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1991).
29. T. C. McAndrew, C. M. Danforth, J. P. Bagrow, Robustness of spatial micronetworks. *Phys. Rev. E* **91**, 042813 (2015).
30. J. Z. Kolter, J. Ferreira Jr., A large-scale study on predicting and contextualizing building energy usage, *Proceedings of the Conference on Artificial Intelligence (AAAI), Special Track on Computational Sustainability and AI*, 2011.
31. https://dataport.pecanstreet.org/
32. J. Shah, B. F. Wollenberg, N. Mohan, Decentralized power flow control for a smart micro-grid. *IEEE Power Energy Soc. Gen. Meet.* 1–6 (2011).
33. S. Bolognani, S. Zampieri, Distributed control for optimal reactive power compensation in smart microgrids, 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC 2011), Orlando, FL, 2011, pp. 2818–2833.
34. A. G. Tsikalakis, N. D. Hatziargyriou, Centralized control for optimizing microgrids operation. *IEEE Trans. Energy Convers.* **23**, 241–248 (2008).
35. K. Purchala, I. Meeus, D. Van Dommelen, R. Belmans, Usefulness of DC power flow for active power flow analysis. *IEEE Power Syst. Eng. Soc. Gen. Meet.* **1**, 454–459 (2005).
36. H. Kakigano, Y. Miura, T. Ise, Low-voltage bipolar-type DC microgrid for super high quality distribution. *IEEE Trans. Power Electron.* **25**, 3066–3075 (2010).
37. B. Stott, J. Jardim, O. Alsac, DC power flow revisited. *IEEE Trans. Power Syst.* **24**, 1290–1300 (2009).
38. A. Scala, S. Pahwa, C. Scoglio, Cascade failures and distributed generation in power grids. *Int. J. Crit. Infrastruct.* **11**, 27–35 (2015).
39. D. Bakken, *Smart Grids: Clouds, Communications, Open Source, and Automation* (CRC Press, Boca Raton, FL, 2014).
40. L. Holmberg Rasmussen, C. Bang, M. Togeby, *Managing Congestion in Distribution Grids—Market Design Consideration. How Heat Pumps Can Deliver Flexibility Through Well- Designed Markets and Virtual Power Plant Technology* (Ea Energy Analyses, Copenhagen, Denmark, NEI-DK–5788, 2012).
41. G. Li, S. D. S. Reis, A. A. Moreira, S. Havlin, H. E. Stanley, J. S. Andrade Jr., Towards design principles for optimal transport networks. *Phys. Rev. Lett.* **104**, 018701 (2010).
42. R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).

Acknowledgments: We acknowledge A. Alfaris and other members of the Center for Complex Engineering Systems (CCES) at King Abdul Aziz City for Science and Technology (KACST) and Massachusets Institute of Technology (MIT) for important comments and suggestions. **Funding:** The research was partly funded by the MIT Portugal Program and the CCES at KACST-MIT. A.S. thanks CNR-PNR National Project "Crisis-Lab," EU FET project DOLFINS no. 640772, EU FET project MULTIPLEX no. 317532, and EU HOME/2013/CIPS/AG/4000005013 project CI2C for support. The contents of the paper do not necessarily reflect the position or the policy of funding parties. **Author contributions:** A.H. performed the analysis and wrote the paper. A.K. made the AC versus DC error tests and provided advice on the power grid literature. A.S. provided general advice and helped in the analysis. M.C.G. designed and guided the research and wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 3 June 2015
Accepted 7 November 2015
Published 15 January 2016
10.1126/sciadv.1500700

Citation: A. Halu, A. Scala, A. Khiyami, M. C. González, Data-driven modeling of solar-powered urban microgrids. *Sci. Adv.* **2**, e1500700 (2016).

ScienceAdvances
AAAS

Data-driven modeling of solar-powered urban microgrids
Arda Halu, Antonio Scala, Abdulaziz Khiyami and Marta C. González (January 15, 2016)
Sci Adv 2016, 2:.
doi: 10.1126/sciadv.1500700