Reducing Emissions through Optimal Placement of Grid Storage Batteries in Texas

Charlie Cowen-Breen
MAT '22

Connor Larson
ORF '22

May 10, 2021

Abstract

Grid-energy storage using batteries is a proposed method to increase the reliability and efficiency of electrical infrastructure, especially as more renewable sources are added to the grid. This paper introduces a linear programming model to maximize efficiency and minimize fossil fuel usage in the grid, and then proposes a model for adding battery capacity to nodes. We show that, when placed optimally, grid-energy storage significantly increases the efficiency of electricity networks.

1 Motivation

Finding ways to reduce reliance on fossil fuels is crucial to combating climate change. Especially given this year's failures in the Texas power grid, improving the resilience of our electric infrastructure is more important than ever. One of the challenges to increasing our usage of renewable energy sources – such as solar and wind power – is these sources' increased variability in generational capacity when compared to traditional energy sources like coal and natural gas.

For this reason, interest in technology for short-term energy storage (particularly via large-scale lithium-ion battery installations[6]) has grown recently. One particular advantage of batteries (over pumped hydro, for example) is that they can be placed nearly anywhere. Our goal is to use the grid model to propose efficient locations for these batteries, and to determine the magnitude of their effect in reducing fossil fuel usage.

These efficient locations are determined by measuring a prescribed cost function, which minimizes both environmental and economic costs. Specifically, we penalize the network according to the amount of electricity generated by fossil fuels, labeled the generation cost.

This also indirectly penalizes routing inefficiency; as detailed in subsequent sections, electricity is lost in transmission via a proportional cost to distance traveled. Because they store electricity locally and, in theory, enable green energy as a viable alternative to fossil fuels, batteries are hypothesized to lower each of these costs.

An initial simplifying model is that generation cost is linear in the amount of electricity generated at fossil fuel plants, and transmission losses are linear with distance - more on this later. This enables us to determine the most efficient routing by solving a linear program, significantly accelerating the optimization process.

2 The Firehouse Problem

In this section, we provide theoretical context for finding efficient battery locations. With the linear assumptions above, deciding efficient battery locations is a special case of a classical NP-hard computer science program, called the vertex k-center problem (aka the *Firehouse Problem*). The general problem consists of choosing an optimal set of k locations to open facilities at, given a set of n houses which the facilities must service; here the facilities are batteries, but another concrete motivating example is firehouses.

One variant of this problem can be stated as follows: suppose that the n houses have locations $h_1, ..., h_n$. Then we seek choices of firehouse locations $f_1, ..., f_k$ to optimize the sum of distances from each house to its closest facility:

$$\min_{f_1,\ldots,f_k} \sum_{i \in |n|} \min_{j \ in|k|} d(h_i, f_j)$$

With $k = 1$, the problem becomes very familiar: it seeks the choice of f which maximizes closeness centrality.[1] (If the graph is complete, and d satisfies the triangle inequality, this is also the geometric median).

Since the Firehouse Problem is NP-hard, it is common to approximate it, particularly with greedy algorithms which optimize for quantities like closeness centrality. In this project, we hope to apply such algorithms on the US electric grid to find optimal locations for energy facilities (batteries), with a network-science approach: for instance, closeness centrality seems to be an appropriate heuristic for location optimality.

One note is that the k-vertex center problem typically assumes that d satisfies the triangle inequality. In our specific case, d denotes the cost to transfer energy between locations (with a potential additional cost due to fossil fuel generation). When d is assumed to be

[1] Here we define closeness centrality for weighted graphs with $d(x, y)$ = minimal sum of weights along paths $x- > y$.

linear in distance, the triangle inequality for d is then immediate from the triangle inequality on the surface of the earth.

We intend to demonstrate the benefits of optimal grid storage through the microcosm of the Texas grid. More broadly, grid energy storage is a well-studied field, and recent progress can be found at Breakthrough Energy publications linked below.[2]

3 Original Dataset

Our original idea for this project was based upon our discovery of a very cool model of the US electric grid, compiled and released open-source by a group of climate scientists and engineers at the Bill Gates-funded *Breakthrough Energy* project.[2] This model consists of both a dataset - with over 82,000 nodes and 104,000 links representing generation, demand, and transmission capabilities - and a suite of software libraries to simulate power flow over fixed periods given a time-series of grid demand, also included in the model.

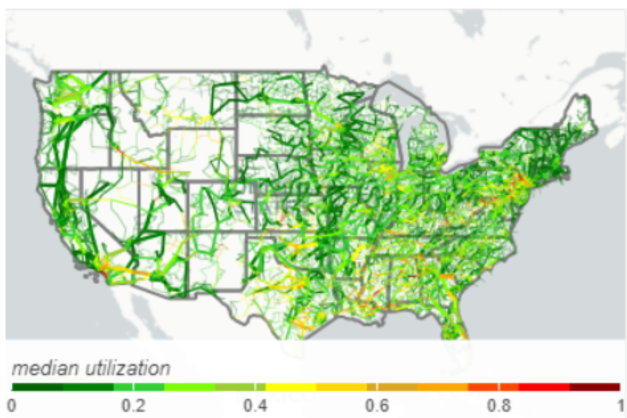

Figure 1: Example of Grid Model Shown on Breakthrough Energy Website

Upon initialization, the Breakthrough Energy framework allows for changes to the network in the form of methods to add (or remove) components such as new AC or DC transmission links, fossil fuel or renewable generator units, and - most importantly to us - batteries. However, we quickly discovered that the underlying simulation used by Breakthrough Energy is extremely complex.

Upon finalizing the grid structure input and running the model, at each point in time (defaulted to 1hr increments), the program solves a massive non-linear optimization program encompassing factors including elasticity of supply and demand to spot prices in the network, non-linear transmission losses in AC transmission lines, and historical weather conditions' effects on renewable capacities.

While the sophistication of this model is compelling - output statistics (such as total costs and losses) are highly robust, and thus the effect of changes to the initial network can be realistically modelled - this comes at great cost. For our project, we determined that using the Breakthrough Energy model would be impractical for two reasons: first, the enormous amount of memory required to the solver (2GB per hour of real-time simulated) was beyond the capabilities of our hardware, requiring 48GB of RAM per day simulated; second, the complexity of modelling transmission losses, coupled with the sheer size of the network, overly complicated our approach to calculating centrality measures for the network.[2]

To proceed, we decided we'd have to simplify our model, both in size and in algorithmic complexity.

4 Synthetic Dataset

The first step of tackling this challenge was finding a new dataset. Based on our original motivation - largely influenced by recent events in Texas[7] - and the fact that Texas uses its own subgrid largely separated from those of the rest of the US, we chose to use a model of the Texas grid compiled by Texas A&M University called ACTIVSg2000[3] (herein referred to as the "TAMU Model"). This dataset contains 2,000 nodes, connected by about 3,200 links.[3]

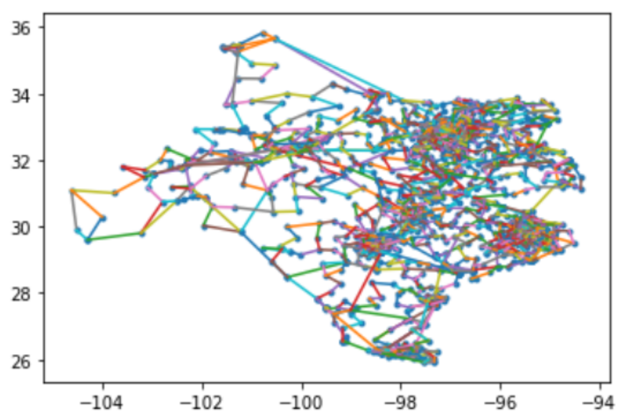

Figure 2: Map of Nodes and Links in TAMU Model

Even still, this dataset was provided with significantly more metrics than we intend to use. For instance, the model contains details on complex electrical properties of transmission

[2]Though we considered utilising the University's remote computation resources after trying (and failing) to get the base model running in a reasonable capacity on our own laptops, after much deliberation, we agreed that using a simpler model would likely be advantageous. We believe a simplified model will deliver similar insights at a more understandable and visualizable scale, while also allowing us total transparency and control over our optimization algorithm.

[3]Important note: this model is what's called a *Synthetic Grid Test Case*. This means that, while it is based on data from the real-world Texas grid, it is *much* smaller in size. It's purpose is to facilitate analyses similar to our project by providing a reasonable approximation of power generation and distribution capabilities in Texas.

lines (impedance, conductance, admittance, etc.) to allow non-linear loss modelling, as well as cost and efficiency data for specific generation facilities. Based on our cost function (described below), most of this data was unnecessary to our analysis. The metrics we retained (which can be found in our Github) are described below:

Links.csv: *Node1 and Node2*: endpoints of the link; *Index*: usually 1, may be 2 if there are multiple parallel links between nodes; *Limit*: maximum flow (in MW) allowable on link.

Nodes.csv: *Node*: number to index node; *Name*: name of the node (e.g. "Odessa"); *Latitude and Longitude*: geographical location of the node; *Limit*: maximum generation capacity (0 if a demand node); *Type*: type of generator (e.g. Wind, Solar, Natural Gas), blank if a demand node.

Loads.csv: *Time*: timestamp for the demands; *Node*: columns indexed by node numbers, containing a time-series of demand in MW at that node.

5 Modelling Approximations

In order to explore the effects of adding storage to the network, we first need a model to simulate the baseline case of the network. This simulation is fundamentally an optimization problem: at each step in time, generation must match demand while minimizing total generation and transmission costs. This problem is constrained by maximum power limits (in MW) for generators and links, as well as the requirement that flows in and out of every node must net to zero.

When dealing with optimization problems, there are significant computational advantages to approximating all costs and constraints as linear. This will be further discussed in the next section. To do so, we make some critical approximations; the most prominent is the assumption of linear transmission costs.

While neither of us are Electrical Engineers, to our best knowledge, real-world transmission losses can be best expressed as a quadratic function of current. This is known as the "I-squared R Law" and states that power loss equal to $P = I^2 R$, where I is current and R is resistance. In a transmission link with fixed voltage, this means that losses would increase with the square of power transmitted.

The real-world case is even more complicated; because transmission lines use alternating current, there are additional losses from elements such as the Corona effect which vary with weather and impedance, for example. These are beyond the scope of our model.

Thus, for our project, we are choosing to approximate transmission losses as being linearly proportional to power and distance: $Loss = (p * distance) * Flow$, where p is some parameter. Based on our research of average losses in the electric grid, we use a value of 5% per 100 miles; as explained below, we do not believe the specific parameter value will be particularly important to our result.

Despite what seems like a large assumption breaking from reality, we do not believe this will invalidate the conclusions of our analysis. Our hypothesized reasoning for this is as follows. Our objective function is to minimize total cost: the sum of total generation at fossil-fuel generators, plus the sum of total losses on the grid. When p is much less than one, this fundamentally means that total losses will be much less than the total amount generated - say, 5% if the average transmission distance is 100 miles. Thus, we postulate that the optimization will essentially proceed in two steps: first, minimize the total number of fossil fuel generators used by maximizing generation from renewables, meaning these will always operate at full capacity; and then second, choose the routing that minimizes transmission losses.

Thus, because this is a secondary problem, the effect of small changes in the parameter p, while scaling the final cost statistic, will likely have little effect on the actual choice of variables. In turn, regardless of the specific choice of p, conclusions about the effects of added storage should be valid as long as p is not changed between test runs.

6 Linear Programming

Given a linear cost function and a set of linear constraints, this problem can be solved via the much-studied procedure of linear programming. The structure for our problem at each time-step is as follows:

Decision Variables: G is a vector of amounts generated at each node, indexed over the set i of generators; F is a vector of unidirectional (e.g. non-negative) flows along transmission links, indexed over the set (i, j) of connected nodes.

Cost Variables: C is a vector of generation cost coefficients, indexed over the set i of generators as above, with a cost of 1 for fossil-fuel plants and 0 for renewables; L is a vector of transmission cost coefficients, indexed over the set (i, j) of connected nodes as above, with values of $dist_{ij}p$ where p is the scalar loss parameter described above and distance is the length in miles of the transmission link.

Bounds: There are three sets of bounds. First, there are *generation* bounds which constrain each generation G_i to a maximum value. Second, there are *transmission* bounds which constrain each generation $F_{i,j}$ to a maximum value. Third, there are *flow* bounds which solve the equation $(1 - losses) * inflows - outflows + generation - demand = 0$ for every node, given the time-series data D, a vector indexed over i which gives the exogenous demand at the point in time t for each node.

Given all of these constraints, the model minimizes the objective function: $\zeta = C^T G$, which is the sum of all energy generated at fossil fuel plants. Using interior-point methods, this problem can be solved in polynomial time. A specific formulation of the linear program is below:

$$
\begin{aligned}
\text{minimize} \quad & \sum_i C_i G_i \\
\text{subject to} \quad & G_i - D_i + \sum_j (1 - L_{ij}) F_{j \to i} - \sum_j F_{i \to j} = 0 && \forall i, \\
& G_i \leq G_{max,i} && \forall i, \\
& F_{i \to j} \leq F_{max,i \to j} && \forall i, j, \\
& F_{i \to j}, G_i \geq 0 && \forall i, j
\end{aligned}
$$

Finally, by rearranging the problem, we can see that this is equivalent to maximizing the efficiency of the network. Rearranging the first constraint and taking the sum over i yields an equivalent objective:

$$\min \sum_{i,j} F_{i \to j} - \sum_{i,j} (1 - L_{ij}) F_{j \to i}$$

Where exogenous variables have been omitted. Rearranging, see:

$$\min \sum_{i,j} L_{ij} F_{j \to i}$$

Subject to the set of constraints as above, this equivalent problem minimizes the amount of energy lost in transmission while ensuring all demands in the network are met.

7 Grid Storage Batteries

Given a baseline efficiency from the optimized network problem solved above, our aim is to explore the effects of adding batteries to nodes on the grid. The goal of batteries is to reduce the reliance on fossil fuels by allowing usage of renewables even in times where they are less available, such as after daylight hours. Another potential benefit of batteries is reducing transmission costs, by banking energy from generators near cities at time of low demand, and supplementing generation during high-demand periods, reducing reliance on far-away generators to meet excess demand.

Clearly, a large question is when to charge the batteries and when to discharge them for optimal reductions in costs. One solution for this would be to expand our linear program to not just solve each time step individually, but rather solve a whole day (for example) by further indexing each variable by times t. However, we decided against this for two reasons: first, expanding the model means the optimization will take proportionally longer and add complexity; second, an optimization over a long time-span allows changes to the zeroth hour based on future demands, for example, which is akin to "seeing into the future."

Therefore, we decided that it makes more sense for us to introduce another simplifying assumption, based on average demand data. Using historical data on times of high and low net demand, we chose to enforce specific times for batteries to charge during times of low demand, and discharge during periods of high demand. This has the additional advantage that total flows in and out of batteries net to zero over a 24-hour period; this allows direct comparison to total energy usage in the base-case when evaluated over one-day intervals.

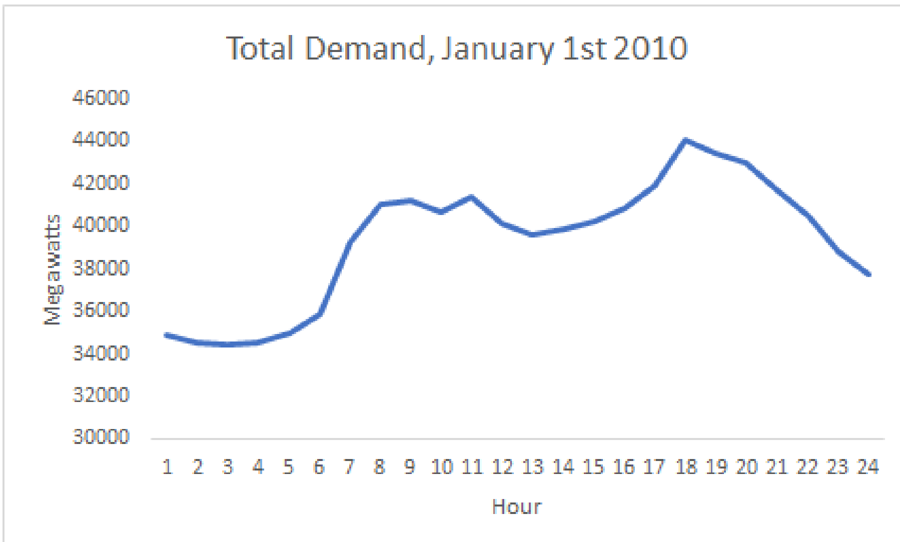

Figure 3: Example of Daily Power Demand Fluctuations

Given the typical demand shown in Figure 3, we chose to exogenously charge batteries at rate $Capacity/8$ from 11pm to 7am, and discharge at rate $Capacity/16$ from 7am to 11pm.

8 Node Centrality Measures

In this section we describe some specifics about the process of computing grid centrality to determine optimal battery locations. One challenge is that the network immediately constructed from the dataset is not connected, and thus closeness centrality cannot be defined. A solution, which we implement here, is to use the closely related notion of harmonic centrality—it seems that experts in the field favor this centrality measure for disconnected graphs.[4]

Formally, suppose we are given a positively weighted directed graph $G = (V, E, w)$, where V denotes the set of vertices, $E \subset V \times V$ denotes the set of directed edges, and

$w : E \longrightarrow \mathbb{R}^{>0}$ denotes the positive weight on each edge. We may then define a distance

$$d(i,j) = \underset{i=v_0,\ldots v_n=j}{\arg\min} \sum_{k=0}^{n-1} w(v_k, v_{k+1})$$

where the $v_0, \ldots v_n$ runs over all paths $i \rightsquigarrow j$.

Recall that closeness centrality is defined, for a connected graph G, by

$$c(i) = \left(\sum_{j \neq i} d(i,j) \right)^{-1}$$

If G is disconnected, it is natural to define $d(i,j) := \infty$ when i and j lie in different connected components. But then the above formula evaluates to 0 for all i, rendering closeness centrality useless.

This motivates the definition of harmonic centrality, for a general graph G, to be

$$h(i) = \left(\sum_{j \neq i} d(i,j)^{-1} \right)$$

Note that the two definitions are related by the HM-GM-AM-QM inequalities, which show that $h(i) \geq (|V| - 1)^2 c(i)$.

In Figure 4, we display the harmonic centrality of nodes in a network derived from the TAMU model as a heatmap. For sake of completeness, we construct this network in two ways. On the left, we construct a weighted graph on the set of Nodes, with weights corresponding to Euclidean distance (hence linear cost). On the right, we contrast this with the corresponding unweighted graph (i.e. constant cost)–this is more consistent with the classical definitions of centrality on unweighted graphs. Finally, on the bottom, we display the heatmap derived from the weighted graph, overlaid with the set of Links.

9 Objective and Hypothesis

Given all of the setup above, we are finally ready and able to conduct our study. Our goal, as outlined in our methodology, is to find locations to add batteries in power networks to best minimize total costs. Our hypothesis is as follows:

We hypothesize that adding batteries to the grid will lower total fossil fuel usage. Furthermore, given some number N batteries available to add to the grid, we hypothesize that choosing nodes with the largest weighted closeness centrality measures will most effectively minimize total fossil fuel usage.

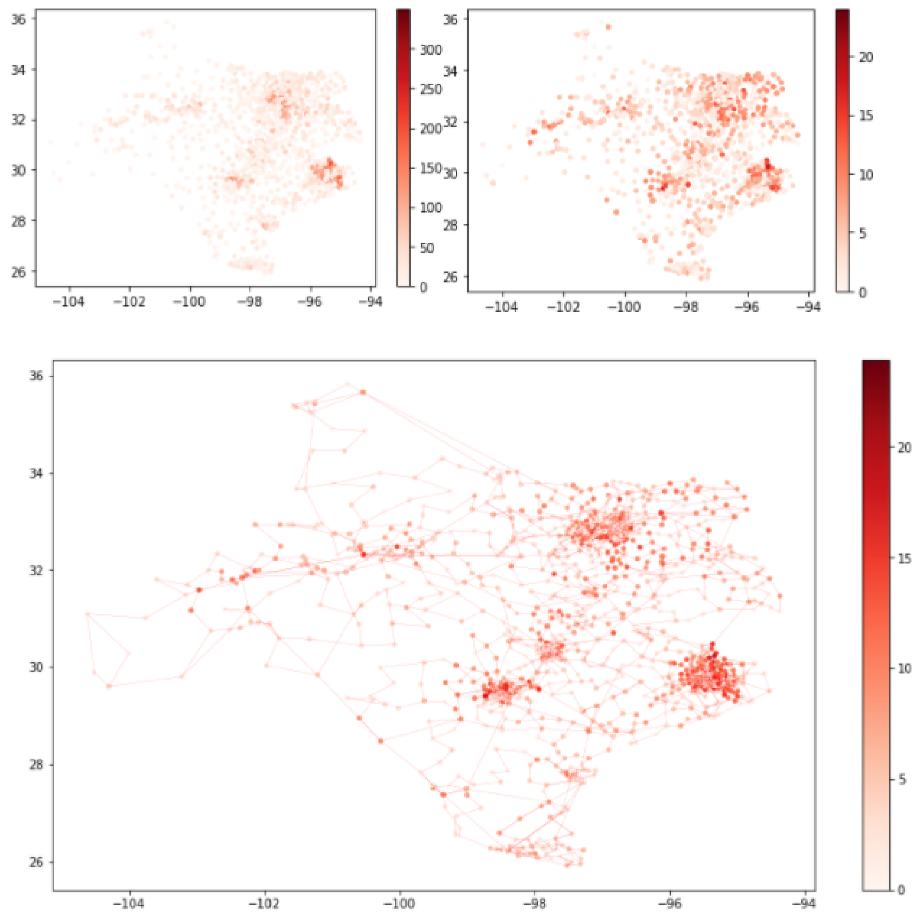

Figure 4: Centralities. Top Left: Unweighted, Top Right: Weighted, Bottom: Weighted, with Links Shown

To test this, we will first run the base network as described above, for a number of time periods $T = 24$, or one day, obtaining a total cost at each time period; summing these costs over t will yield a baseline value of total fossil fuels used.

Next, we will add N batteries at the nodes with the highest closeness centralities, and obtain a new total cost; we hypothesize that this will be lower than the baseline total cost, as the batteries reduce fossil fuel generation and transmission losses in the network. For comparison, we will also observe the results of adding N batteries at the nodes with the lowest closeness centralities.

Finally, we will run many trials adding N batteries at random nodes; this will serve as a control, to ensure that the closeness centrality of placement nodes has a positive reductive effect on total cost. We hypothesize that the control total cost will be lower than the base cost, but higher than the cost obtained via placement at nodes with high closeness centrality.

10 Results and Analysis

10.1 Base Case

Using the linear program solver Gurobi, the base case of our network was efficiently solved in modest time. As shown for a single time-step in Figure 5, the efficient solution results in a directed acyclic graph; that is, there are no directed cycles. Though this digraph uses significantly fewer links that available, it is not a polytree, as the capacity constraints on links result in some redundent connections between nodes. The digraph is also not completely connected; while there appears to be one giant connected component, there are also several smaller components. This is most obvious in the *eastern* portion of the state: observe an example of a small connected component north of Houston.

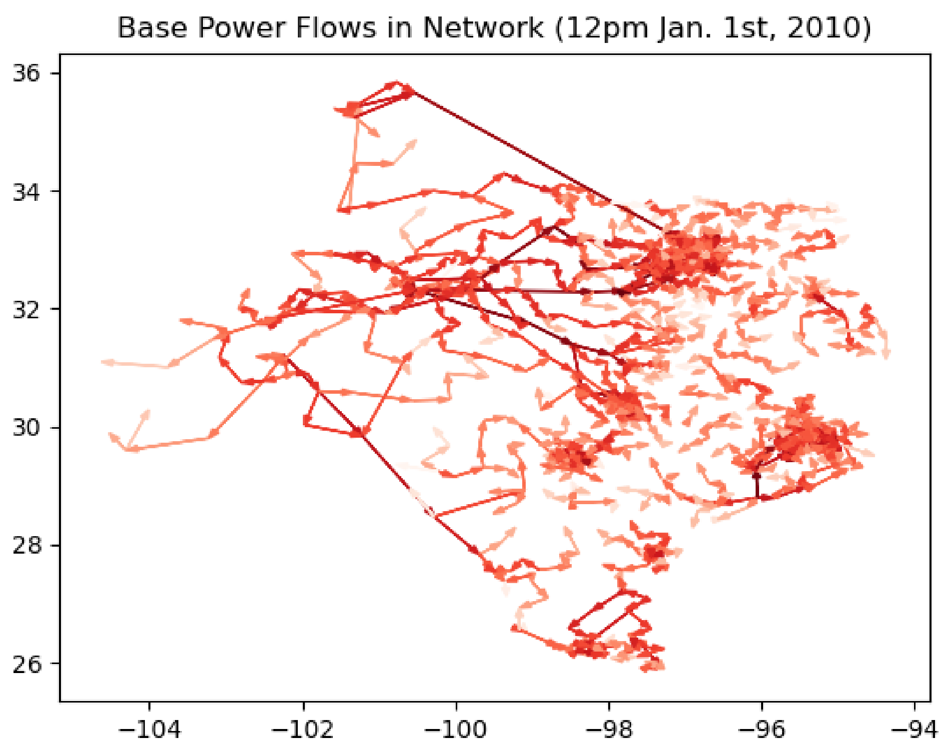

Figure 5: Power Flow in the Base Case, Solved via Linear Programming

In the 24-hour period starting on January 1st, 2010, the total amount of electricity generated from fossil fuel sources in the base case was 522.25 GWh. This resulted in a grid efficiency of 96.4% (where efficiency is defined as total generated divided by total used at demand nodes).

Finally, we observe that in the base case of the network (and in all subsequent runs with batteries), all sources of renewable energy operate at their maximum capacity limits in every time step. This matches our expectation; this also means that in subsequent stages of our analysis, all reductions in fossil fuel usage are a result of increased transmission efficiency (as opposed to an increase in renewables output, which could only happen if renewables were not already at their upper bounds).

10.2 Battery Strategies

Adding batteries to the network significantly reduced fossil fuel usage, confirming our hypothesis. Because the batteries charge and discharge equal amounts over our 24-hour test period, net demand does not change from the base case; thus, all observed decreases in fossil fuel usage are a result of increased transmission efficiency.

As planned, we evaluated three methods for choosing nodes at which to place batteries: first, our hypothesized solution of the N nodes with the *highest* closeness centralities; second, a control case of N *random* nodes selected via 20 Monte Carlo simulations; and third, a set of N nodes with the *lowest* closeness centralities, for comparison. We also added a fourth method for evaluation: the N nodes with the highest net load over the 24-hour test period.

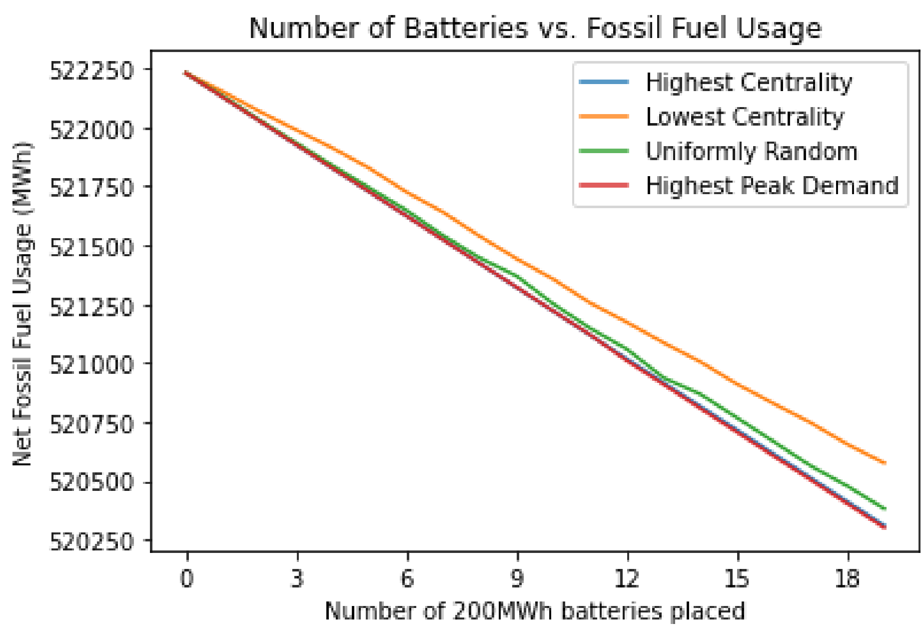

Figure 6: Number of 200MWh Batteries vs. Fossil Fuel Usage with Different Placement Strategies

As shown in Figure 6, all placement strategies had a significant reductive effect on total generation from fossil fuel sources. The least effective strategy, as expected, was placing batteries at the nodes with lowest centrality. The most effective strategy was placing batteries at the nodes with highest demand, though this was only marginally more effective than our

Placement strategy	MWh saved per battery	MWh saved per MWh capacity
Highest peak demand	96.35	0.482
Highest centrality	95.90	0.480
Random choice	92.33	0.462
Lowest centrality	82.65	0.413

Table 1: Efficiency comparison of battery placement strategies (200Mwh batteries, 24hr savings)

hypothesized strategy of choosing nodes with the highest centrality. These differences are more easily observed in Figure 7, which shows the effect of each strategy normalized against our control case of 20 Monte Carlo simulations with random nodes.

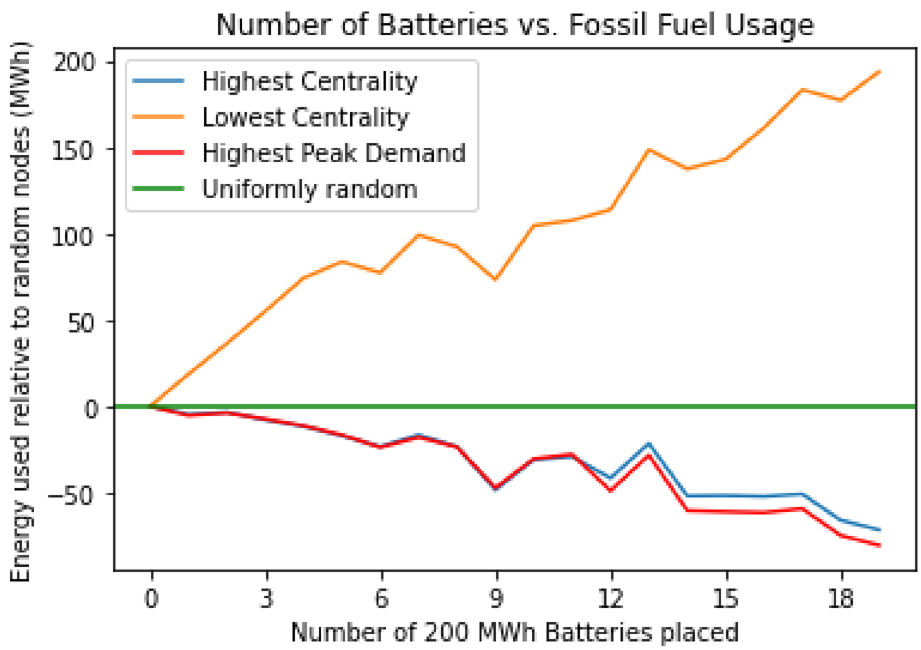

Figure 7: Comparison of Different Strategies with Control Case

These results demonstrate our hypothesis that some placement strategies are more efficient than others. To quantify this observation, we compare the average amounts of energy saved per battery according to different placement strategies in Table 1. Note that, as shown in Figure 6, the total amount of energy saved appears to be roughly linear in the number of batteries used, across all placement strategies.

Of the strategies listed, we find that placing batteries according to highest peak demand is the most efficient, but only marginally more so than according to highest centrality. Therefore, in the case that the highest peak demand nodes are difficult to place batteries at—for instance, if they coincide with the most urban nodes—placing batteries according to highest centrality is a viable alternative.

Both strategies are more efficient than placing randomly, which is in turn much more efficient than placing according to the lowest centrality nodes. In other words, placing randomly is less efficient than placing by highest centrality, but much more efficient than placing by lowest centrality. One intuition for this disparity is that the higher centrality nodes are more likely to be randomly sampled (this is a form of the friendship paradox), and so the effects of choosing a low centrality node intentionally are particularly catastrophic. Experimenting with other, fairer, random sampling schemes—perhaps uniformly over area instead of uniformly over all nodes—is an interesting area of future study.

10.3 Efficiency

Finally, given the most effective strategy of placing batteries at nodes with high demand, we ran additional additional tests to determine the relationship between total battery capacity added to the grid, and net efficiency of electricity transmission in the network. Our analysis also reveals a likely mechanism of action for the efficiency improvements of the network.

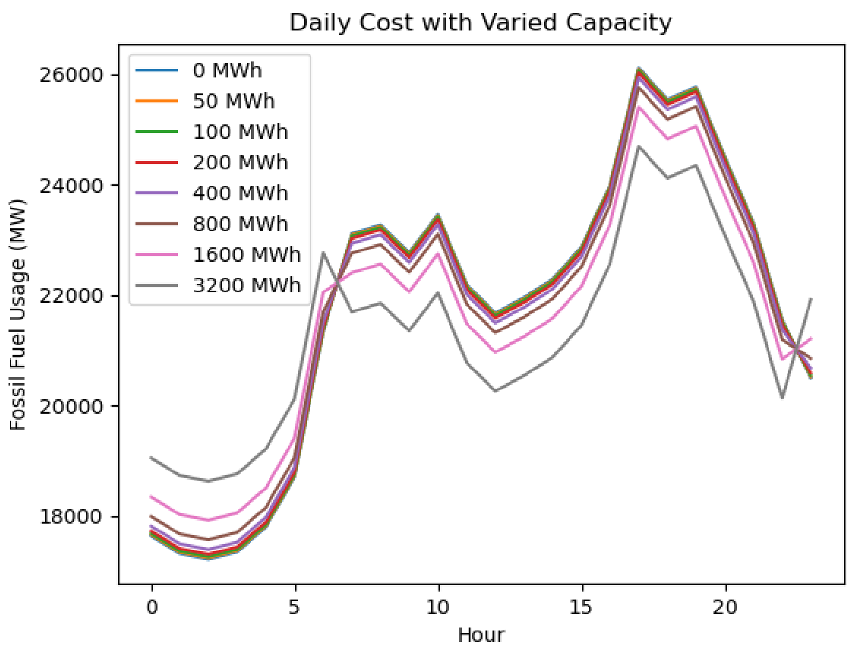

Figure 8: Hourly Fossil Fuel Usage with Varying Battery Capacities

As shown in Figure 8, increasing battery capacity effectively smooths daily demand, reducing the peak load on the network and increasing the minimum load on the network. As more fossil fuel plants come online to provide energy during times of peak demand, the average distance electricity is transmitted through the network increases; the fossil fuel plants near cities reach capacity, and transmission links bind forcing energy to take longer paths

Total battery capacity	Overall transmission efficiency
0 MWh	96.4%
4,000 MWh	96.6%
8,000 MWh	96.8%
12,000 MWh	97.0%
16,000 MWh	97.2%
20,000 MWh	97.4%

Table 2: Improvements in Efficiency from Increasing Net Capacity

from generation to usage. By charging during times of lower demand when fewer link and generation constraints are binding, and then discharging directly to fulfill demand (especially when batteries are placed directly at nodes with high demand), overall transmission distance in the network is reduced and efficiency is improved.

Lastly, we observe that efficiency savings are linear in proportion to capacity added, provided capacity is added via the high-demand node strategy.

11 Discussion and Conclusions

Despite the large number of assumptions and approximation discussed throughout this project—the primary ones being: a synthetic (scaled-down) network, strictly linear costs, always-on wind and full-capacity solar during daylight hours, and fixed charge/discharge cycles for batteries—we believe that these results are generalizable to more complex models of real-life energy grids. But even in our simplified model, some of the conclusions were not at all obvious to the authors—for example, that low centrality nodes are particularly inefficient locations at which to place batteries. For instance, it could well be that the most efficient battery locations are geographically dispersed; on the contrary, here we find that the most efficient locations are clustered together.

Furthermore, despite these assumptions, our model also gives what we believe to be a reasonable sense of scale to the potential benefit of adding batteries to our current grid. Based on our observation of a 0.482 MWh daily energy savings from fossil fuel sources when added using the optimal strategy, each MWh of battery capacity introduced to the grid has the potential to reduce yearly energy generation by nearly 176 MWh. At a cost of $100 per MWh from fossil fuel sources, this amounts to $17,600 in savings per year, per MWh capacity. [1] Based on the latest projection's of a $200,000 per MWh cost for Tesla's Megapack, these batteries could save enough to cover their capital costs in as little as 12 years. [5] While this does not consider inflation, it does suggest that with falling battery prices, installing grid capacity may likely be economically feasible in the near future.

Ultimately, we hope that these insights may be further applicable to broader problems outside electrical delivery; some models that come to mind include reservoirs in water delivery systems, parking lots in autonomous taxi systems, and data centers in internet systems.

References

[1] Cost of electricity by source, May 2021. Page Version ID: 1021914180.

[2] Breakthrough Energy. Breakthrough Energy Sciences.

[3] Hanyue Li, Ju Hee Yeo, Ashly L. Bornsheuer, and Thomas J. Overbye. The Creation and Validation of Load Time Series for Synthetic Electric Power Systems. *IEEE Transactions on Power Systems*, 36(2):961–969, March 2021.

[4] Massimo Marchiori and Vito Latora. Harmony in the small-world. *Physica A*, 285, 2000.

[5] Zachary Shahan. Tesla Megapack, Powerpack, & Powerwall Battery Storage Prices Per kWh — Exclusive, October 2020.

[6] Tesla, Inc. Introducing Megapack: Utility-Scale Energy Storage, July 2019.

[7] Wikipedia Contributors. 2021 Texas power crisis, May 2021. Page Version ID: 1021838809.