A Statistical Method for Synthetic Power Grid Generation based on the U.S. Western Interconnection

Saleh Soltan, Gil Zussman
Columbia University
New York, NY
Failures in Power Grids

- One of the most essential infrastructures in modern life
- Rely on physical components → Vulnerable to physical attacks/failures
- Failures may cascade → Blackouts (US’03, India’12, Turkey’15)

- An attack/failure will have a significant effect on many interdependent systems (communications, gas, water, etc.)
Failures/Attacks/Disasters

- Mismanagements
- EMP (Electromagnetic Pulse) attack
- Solar Flares - Federal Energy Regulatory Commission (FERC) has recently issued a rule for transmission grid operators to develop a plan to deal with the Geomagnetic disturbances


- Other natural disasters
- Physical attacks

Source: FERC, DOE, and DHS, Detailed Technical Report on EMP and Severe Solar Flare Threats to the U.S. Power Grid, 2010
“A sniper attack in April 2014 that knocked out an electrical substation near San Jose, Calif., has raised fears that the country's power grid is vulnerable to terrorism.” –The Wall Street Journal
Motivation

- Need to study vulnerabilities in the real grid topologies
- Real data may **not be available** to all researchers or **hard to obtain** due to security reasons
- **Not wise** to publish vulnerabilities of the real grid
- Current situation
  - Small-world, scale-free networks, etc. do not consider the geographical locations (The reactance value and type of a line is directly correlated to its length)
  - Spatial networks (e.g., random geometrics graphs) are not designed to generate networks with properties similar to power grid networks
  - Limited reference test cases (Polish Grid, IEEE benchmark systems, etc.) that do not contain the coordinates of the lines
- We present a procedure to generate synthetic networks with similar structural properties to power grids
Related Work

- The structural properties of the power grids around the World (North America, Europe, etc.) has been widely studied:
  - Watts and Strogatz (1998) → small-world property (average path length and clustering coefficient)
  - Barabasi and Albert (1999) → the power-law degree distribution
  - Studies of the power grids in European countries shows similar properties (2007)
  - Wang, Scaglione, et. al. (2010) → Algebraic connectivity
  - Recent work by Hines et. al. (2012) → show existent topological models do not satisfy all the structural properties of power grids

- Few synthetic models are available
  - Random Networks (Small-world, scale-free, etc.)
  - Wang, Scaglione, et. al. (2010)
  - Recent work by Schultz et. al. (2014)

- None has considered the spatial distribution of the nodes and the length distribution of the lines
What is important?

- **Number of Nodes**
- **Number of Edges**
- **Average path length**: The average shortest path lengths (number of edges) between all pairs of nodes
- **Clustering coefficient**: The fraction of connected pairs between all the neighbors of a node $i$, averaged over all nodes $i$
- **Degree distribution of the nodes**
- **Length Distribution of the lines**: Is directly correlated with physical properties of the lines (e.g., resistance, reactance)
Two major: The Western (WI) and Eastern Interconnections (EI)
Three minor: The Texas, Québec, and Alaska Interconnections

The data is obtained from the Platts Geographic Information System (GIS)
North American Electric Reliability Corporation (NERC) regional entities and the National Electricity Transmission Grid of Mexico (NETGM).
Structural Properties

- We consider:
  - Western Interconnection (WI)
  - SERC Reliability Corporation (SERC)
  - Florida Reliability Coordinating Council (FRCC)

<table>
<thead>
<tr>
<th>Network</th>
<th>WI</th>
<th>SERC</th>
<th>FRCC</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Nodes</td>
<td>14302</td>
<td>12946</td>
<td>1312</td>
</tr>
<tr>
<td># of Edges</td>
<td>18769</td>
<td>16658</td>
<td>1780</td>
</tr>
<tr>
<td>Average Path Length</td>
<td>17.33</td>
<td>19.71</td>
<td>11.68</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.049</td>
<td>0.049</td>
<td>0.075</td>
</tr>
<tr>
<td>Power-law exponent $\zeta$</td>
<td>-3.48</td>
<td>-3.93</td>
<td>-2.76</td>
</tr>
</tbody>
</table>
The Length Distribution of the Lines

- First time that has been studied
- Non-parametric fit

WI

SERC

FRCC
Outline of our Method

1. Generate nodes with the similar spatial distribution to a given network with $n$ nodes and $m$ edges

2. Connect nodes based on the distances and degrees
Western interconnection with \( n = 14302 \) substations (nodes) and \( m = 18769 \) lines (edges)
Cluster points based on their geographical proximity

Gaussian Mixture Models (GMM) are commonly used for clustering and density estimation

Mixtures describe two-stage sampling procedure
- Sample $j \sim \pi = (\pi_1, \ldots, \pi_c)$
- Sample $X_i \sim \mathcal{N}(\mu_j, \Sigma_j)$

Given a network, $\pi$ and $(\mu_j, \Sigma_j)$s can be estimated using algorithms like EM algorithm

Example of clustering nodes in WI into 10 clusters
- Each color represents nodes assigned to a cluster
Position of the Nodes

- Use the Bayesian Information Criterion (BIC) to select number of clusters → 55 clusters
- Use the obtained parameters from GMM to generate $n$ nodes with similar spatial distributions as the nodes in WI
Outline of our Method

1. Generate nodes with the similar spatial distribution to a given network with $n$ nodes and $m$ edges

2. Connect nodes based on the distances and degrees
Connection Between the Nodes

- Inspired by the way the power grids have evolved through time
- Find a spanning tree of nodes \((\text{Connectivity})\) \(\rightarrow n - 1\) edges
- Add more edges to the graph \((\text{Robustness})\) \(\rightarrow m - n + 1\) edges
  - Number of edges
  - Clustering coefficient
  - Degree distribution of the nodes
Connectivity

- Minimum weight Spanning Tree (MST) → high average path length
- Connect node $i$ to its geographically closest node with index less than $i$
- Depends on the ordering of the nodes

There is an ordering that obtains the minimum spanning tree using this process.
Finding a spanning tree (tunable parameter $\kappa$):

- Randomly order nodes with probability proportion to $||X_{i} - \bar{X}||^{-\kappa}$
- Connect node $i$ to its geographically closest node with index less than $i$
**Tunable Parameter $\kappa$**

- $\kappa$ controls the total weight and the average path length of the obtained tree.
- Average path length in MST is $\approx 505$
Tunable Parameter $\kappa$

$\kappa = 0$

$\kappa = 2.5$

$\kappa = 9$

Get closer to the MST
Connection Between the Nodes

- Find a spanning tree of nodes (Connectivity) \( \rightarrow n - 1 \) edges
- Add more edges to the graph (Robustness) \( \rightarrow m - n + 1 \) edges
  - Number of edges
  - Clustering coefficient
  - Degree distribution of the nodes
Add More Edges to the Graph: Observations

1. Power-law degree distribution
   - Preferential attachment
   - Less degree 1 & 2 nodes

2. Longer edges have lower probabilities

3. Nodes in denser areas have higher degrees
   - Define $\rho_i$ as the average geographical distance of the node $i$ from its $k$ closest nodes
   - $\rho_i^{-1}$ is the density of nodes around node $i$
Add more Edges to the Graph

- Repeat \( m - n + 1 \) times:
  1. Pick a low degree node in a dense area
  2. Connect it to a high degree but close node

- If Small Network: Sample node \( i \) with probability \( \propto d_i^{-\eta} \rho_i^{-\alpha} \)
- If Large Network: From all nodes with degree less than 3, sample node \( i \) with probability \( \propto \rho_i^{-\alpha} \)

- Connect node \( i \) to node \( j \) sampled from all other nodes with probability \( \propto ||X_i - X_j||^{-\beta} d_j^\gamma \)
  - \( \beta \) tune the probability distribution of length of the lines
  - \( \gamma \) tune the exponent of the power-law distribution for degrees

- \( \alpha, \beta, \gamma, \eta \) are tunable parameters

\( d_i \): Degree of the node \( i \)
\( \rho_i^{-1} \): Density of the nodes around the node \( i \)

\( X_i \): Position of the node \( i \)
Add more Edges to the Graph
Evaluation
Kullback-Leibler (KL) divergence

- Measure the similarity between the length distribution of the lines in two networks

- The KL-divergence is a non-symmetric measure of the difference between two probability distributions $P$ and $Q$

$$D_{KL}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}$$

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx$$

- We do not have the actual distribution $\rightarrow$ use samples to estimate the KL-divergence as in the paper by Boltz et. al.

Evaluation: WI

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Gen.</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Nodes</td>
<td>14302</td>
<td>14302</td>
</tr>
<tr>
<td># of Edges</td>
<td>18769</td>
<td>18769</td>
</tr>
<tr>
<td>Average Path Length</td>
<td>17.33</td>
<td>17.28</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.049</td>
<td>0.050</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>-3.48</td>
<td>-3.84</td>
</tr>
<tr>
<td>$D_{KL}$</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Length Distribution $D_{KL} = 0.1$

Degree Distribution

$\zeta = -3.48$

$\zeta = -3.84$
Evaluation: SERC

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Gen.</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Nodes</td>
<td>12946</td>
<td>12946</td>
</tr>
<tr>
<td># of Edges</td>
<td>16658</td>
<td>16658</td>
</tr>
<tr>
<td>Average Path Length</td>
<td>19.71</td>
<td>20.46</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.049</td>
<td>0.049</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>-3.93</td>
<td>-4.20</td>
</tr>
<tr>
<td>$D_{KL}$</td>
<td>0</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Length Distribution $D_{KL} = 0.064$

Degree Distribution

$\zeta = -3.93$

$\zeta = -4.20$
Evaluation: FRCC

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Gen.</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Nodes</td>
<td>1312</td>
<td>1312</td>
</tr>
<tr>
<td># of Edges</td>
<td>1780</td>
<td>1780</td>
</tr>
<tr>
<td>Average Path Length</td>
<td>11.68</td>
<td>10.65</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.075</td>
<td>0.057</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>-2.76</td>
<td>-2.73</td>
</tr>
<tr>
<td>$D_{KL}$</td>
<td>0</td>
<td>0.08</td>
</tr>
</tbody>
</table>
A method to generate synthetic networks with structural properties similar to a given network
Applied it to different parts of the North America grid
It can be applied to any spatial network (e.g., transportation, gas pipes)
It can be used to generate networks similar to a given network with less number of nodes and edges (turn back the time!)

- Publish our code as a Library in $R$ for generating synthetic networks similar to any part of the North America grid as well as any given network
- Submit our work to the IEEE Transactions on Network Science and Engineering
Thank You!

saleh@ee.columbia.edu

http://wimnet.ee.columbia.edu