Identification of Edge Disconnections in Networks Based on Graph Filter Outputs

Shlomit Shaked and Tirza Routtenberg

School of Electrical and Computer Engineering Ben-Gurion University of the Negev, Beer-Sheva, Israel

Motivation and Objective

- Most graph signal processing (GSP) methods are based on known network topology.
- However, small topology changes significantly degrade the performance of GSP tasks.
- In particular, edge disconnections, i.e. links between the graph vertices that have been dropped, is a common problem, especially in physical networks.
- For example, in power networks, the problem of identifying line outages due to environmental factors, damages, aging, and malicious attacks, is a significant problem.

Goal: By using graph signals, identify edge disconnections, where the original network topology is known.

Background: GSP Definitions

- An undirected, connected, weighted graph $\mathcal{G}=\{\mathcal{V},\mathcal{E},\mathbf{W}\}$, where $N\triangleq|\mathcal{V}|$ and $\mathbf{W}$ is the *adjacency matrix*. If $(i,j)\in\mathcal{E}$, the entry $\mathbf{W}_{i,j}>0$ represents the weight of the edge; otherwise, $\mathbf{W}_{i,j}=0$.
- The *Laplacian matrix*: $\mathbf{L}\triangleq\text{diag}(\mathbf{W1})-\mathbf{W}$. with the eigenvalue decomposition $\mathbf{L}=\mathbf{U}^{(\mathbf{L})}\Lambda^{(\mathbf{L})}(\mathbf{U}^{(\mathbf{L})})^\top$.
- A *graph signal* is a vector measured over the vertices, $\mathbf{a}:\mathcal{V}\rightarrow\mathbb{R}^N$.
- The *graph Fourier transform (GFT)* and *inverse GFT (IGFT)* w.r.t $\mathbf{L}$ are $\tilde{\mathbf{a}}^{(\mathbf{L})}=(\mathbf{U}^{(\mathbf{L})})^\top\mathbf{a}$ and $\mathbf{a}=\mathbf{U}^{(\mathbf{L})}\tilde{\mathbf{a}}^{(\mathbf{L})}$.
- The *smoothness* is measured by
$$Q_{\mathbf{L}}(\mathbf{a})\triangleq\frac{1}{2}\Sigma_{(i,j)\in\mathcal{E}}\mathbf{W}_{i,j}[\mathbf{a}_i-\mathbf{a}_j]^2=\mathbf{a}^\top\mathbf{L}\mathbf{a}.$$
- A *graph filter* is a linear operator $h(\mathbf{L})\triangleq\mathbf{U}^{(\mathbf{L})}\text{diag}([h(\lambda_1^{(\mathbf{L})}),\ldots,h(\lambda_N^{(\mathbf{L})})]^T)(\mathbf{U}^{(\mathbf{L})})^T$.
- We give a formal definition of smooth graph filters for an input-output system $\mathbf{a}_{\text{out}}=h(\mathbf{L})\mathbf{a}_{\text{in}}$:
$$\frac{\text{E}[Q_{\mathbf{L}}(\mathbf{a}_{\text{out}})]}{\text{E}[Q_{\mathbf{L}}(\mathbf{a}_{\text{in}})]}<1.$$

Model and Problem Formulation

The measurement model is a vector of M time samples of smooth graph filter output, $\mathbf{y}\triangleq[\mathbf{y}^T[1],\ldots,\mathbf{y}^T[M]]^T$, where
$$\mathbf{y}[m]=h(\mathbf{L})\mathbf{x}[m]+\mathbf{w}[m],\ m=1\ldots M.$$
Problem formulation: Identification of edge disconnections from a set of possible graphs:
$$\mathcal{H}_k:\quad \mathbf{L}=\mathbf{L}^{(k)},\quad k=0,1,\ldots,K$$
based on the graph signals, $\mathbf{y}$, where

- $\mathbf{L}^{(0)}$ is the Laplacian of the original, known topology (set of edges: $\mathcal{E}$)
- $\mathbf{L}^{(k)}$ is the Laplacian matrix after edge disconnections (set of edges: $\mathcal{E}\setminus\mathcal{C}^{(k)}$).

Maximum Likelihood Decision Rule

The maximum likelihood decision rule is given by
$$\underset{0\leq k\leq K}{\text{argmax}}\frac{\log f(\mathbf{y};\mathbf{L}^{(k)})}{\log f(\mathbf{y};\mathbf{L}^{(0)})}=\underset{0\leq k\leq K}{\text{argmax}}\, l(\mathbf{y}|\mathbf{L}^{(k)})-\rho(\mathbf{L}^{(k)}).$$

- $l(\mathbf{y}|\mathbf{L}^{(k)})$ presents the data term.
  - In the graph spectral domain, the sufficient statistics are NK scalars of the graph-frequency energy levels.
  - For the Gaussian Markov random fields (GMRF), $l(\mathbf{y}|\mathbf{L}^{(k)})$ includes only data measured over the vertices associated with the edges in the edge disconnection set.
- $\rho(\mathbf{L}^{(k)})$ presents the penalty term.
  - For nested edge disconnections subsets, a larger penalty for the hypothesis with a larger number of disconnections.
  - For the GMRF, $\rho(\mathbf{L}^{(k)})$ is a function of the second-order statistics of the graph signal over those vertices.

Greedy Approach

- The greedy approach starts with an empty set at the first iteration.
- At the lth iteration, we test all the available edges in the graph and choose the edge that maximizes the marginal likelihood for a single edge.
- If the likelihood ratio of the chosen edge is higher than zero, we add it to the edge disconnections set. Otherwise, the algorithm stops.
- If the maximum of edge disconnections, $r_{\max}$, is known, then it can be used as an additional stopping condition.

The Neighboring Strategy

The neighboring strategy is inspired by the local property of the GMRF model.

- For a candidate edge, (i,j), calculating the likelihood ratio of the measurements in the β-neighborhood of i and j, $\mathcal{N}(i,\beta)\cup\mathcal{N}(j,\beta)$, where $\mathcal{N}(i,\beta)$ is the set of vertices connected to vertex i by a path of at most β edges.
- In every iteration, building new suspicious edges set for the searching in the following iteration.
- The tunable parameter β provides a trade-off between the identification accuracy and the computation cost.

Identifying edge disconnections performance

We consider smooth graph filter. The initial graph was generated by using the Watts-Strogatz small-world graph model, with $N=50$ vertices, mean degree of $d=2$, and $|\mathcal{E}|=100$. The elements of $\mathbf{W}$ are independent, uniformly distributed weights in $[0.1,5]$. Topology change is obtained by removing an arbitrary set of r edges from $\mathcal{E}$.

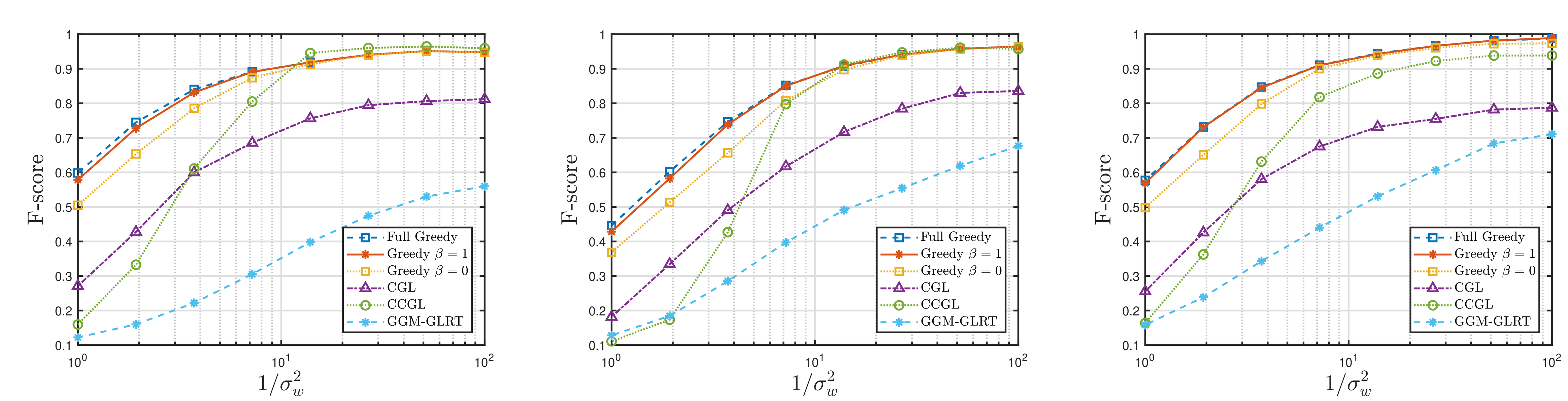

Figure 1:The F-score measure for the GMRF (left), the regularized Laplacian (middle), and the heat diffusion (right) filters versus SNR for the greedy algorithm, $\beta=0$, 1-neighbors greedy algorithm, CGL method, CCGL method, and the GGM-GLRT method with $M=1,000$ and $r=5$.

Comparison with existing methods: 1. Combinatorial graph Laplacian (CGL) method [2]; 2. Constrained CGL (CCGL) method: CGL + information on the initial Laplacian matrix, $\mathbf{L}^{(0)}$; and 3. Gaussian graphical model (GGM) - GLRT: edge exclusion test [3].

Computational Complexity

- The number of hypotheses in the general case is $K=\sum_{r=1}^{r_{\max}}\binom{|\mathcal{E}|}{r}$.
- The likelihood ratio calculations require an inversion of $N\times N$ matrix.
- The proposed methods search over linear number of possibilities, which is significantly smaller than K for large networks.
- Given a sparse graph with a small degree, most of the edges have small sets of β-local neighborhood. Thus, the matrix inversions are performed on smaller matrices and the size of the searching edge set is smaller.

Identifying outages in power system

- The vertices and the edges denote the buses (generators or loads), and the transmission lines between the buses, respectively. The branch susceptances determine the weights of the edges.
- We assume Phasor Measurement Units that acquire noisy measurements of the voltage phases at all buses. It has recently been shown that these voltages can be considered smooth graph signals.

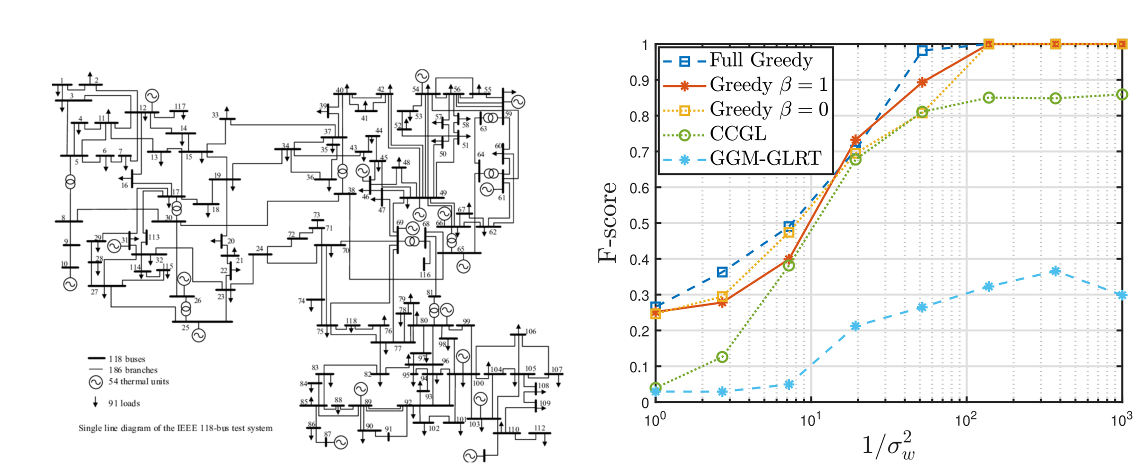

Figure 2:Identifying outages in power system: IEEE-118 bus test case (left) and the F-score measure (right) versus SNR by assuming the GMRF filter for random combinations of outages at the transmission lines.

References

[1] S. Shaked and T. Routtenberg. "Identification of edge disconnections in networks based on graph filter outputs", 2021.
[2] H. Egilmez, E. Pavez, and A. Ortega, "Graph Learning from Data under Structural and Laplacian Constraints", 2017.
[3] K. Tugnait, "Edge exclusion tests for graphical model selection: Complex Gaussian vectors and time series", 2019.

Acknowledgement

This work is supported by the Israeli Ministry of National Infrastructure, Energy and Water Resources and by the BGU Cyber Security Research Center.