Network Centrality as Statistical Inference in Large Networks

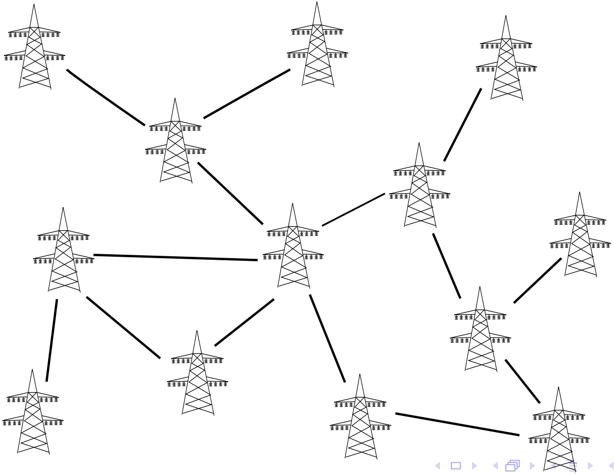
Chee Wei Tan

City University of Hong Kong

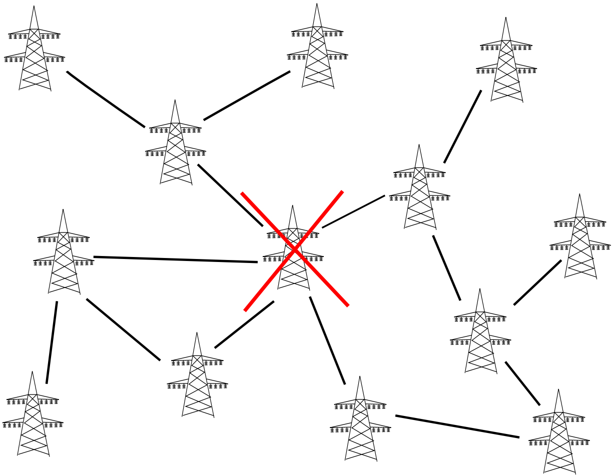
June 18, 2018

Problem Statement

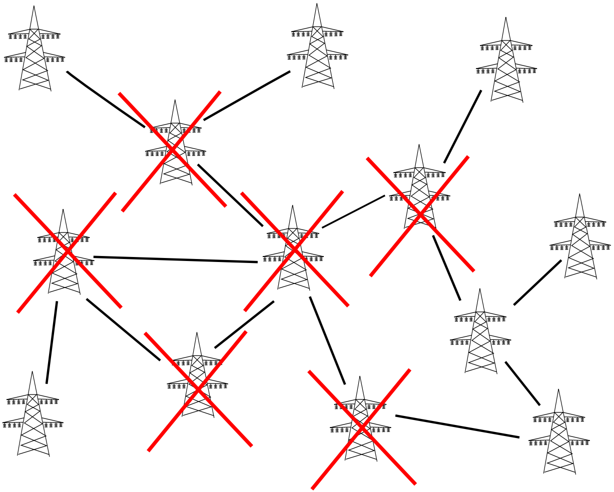
Averting Cascading Failures in Networked Infrastructures: Poset-constrained Graph Algorithms, IEEE Journal of Selected Topics in Signal Processing, 2018 [1]



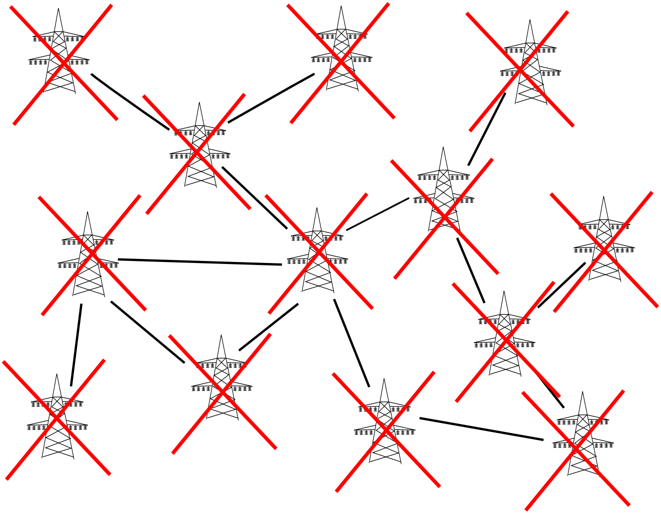
Problem Statement



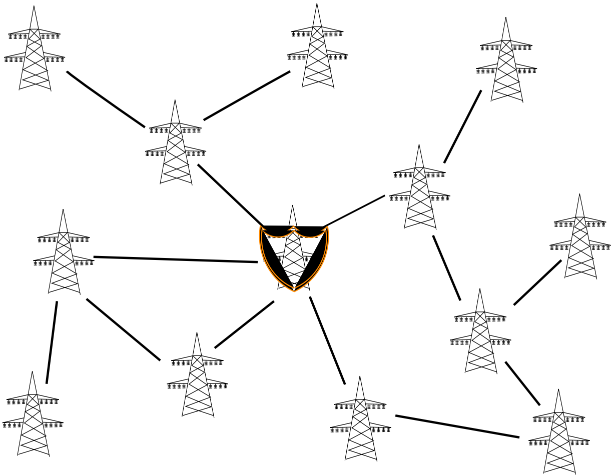
Problem Statement



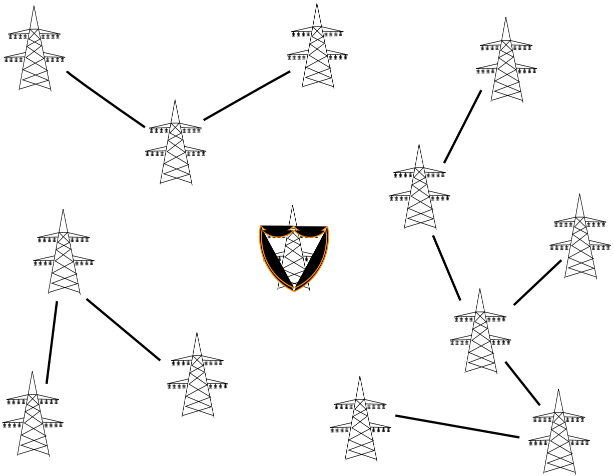
Problem Statement



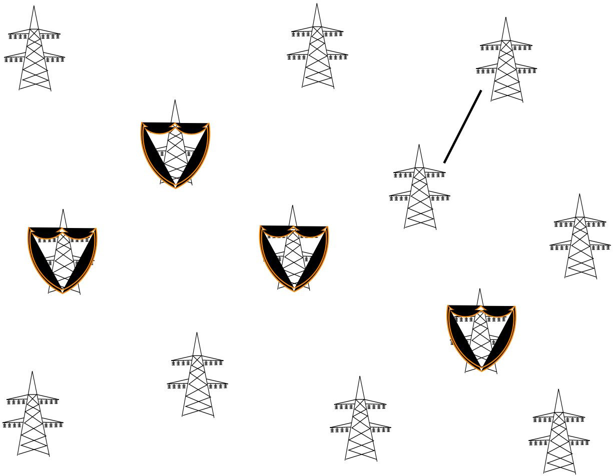
Problem Statement



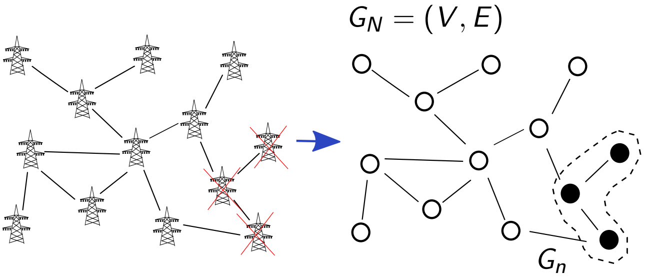
Problem Statement



Problem Statement



The Model and Assumptions



The Model and Assumptions

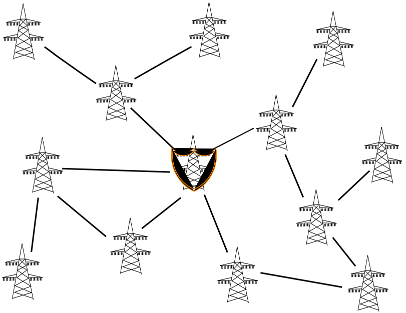
In the extended SI model, we have three types of nodes described as following:

- Susceptible node: Nodes that are susceptible to failure.
- Infected node: Nodes that are under the effect of failure.
- Protected node: Nodes that are protected and can not spread the failure further.
- Every vertex is equally like to be the source
- Assume that in each time period, one vertex is uniformly chosen from the neighbors of those infected vertices to be infected.

The Protection Node Placement Problem

$$
\begin{aligned}
\underset{v \in G_n}{\text{minimize}} \quad & \mathbf{E}(|G_n|) \\
\text{subject to} \quad & |V_P| = k,
\end{aligned} \tag{1}
$$

Example: $|V_P| = 1$



$$\mathbf{E}(|G_n|) = \frac{1}{13} \cdot [(3+3+3) + (3+3+3) + (5+5+5+5+5)]$$
$$= \frac{1}{13} \cdot [3^2 + 3^2 + 5^2]$$

The Protection Node Placement Problem

$$
\begin{array}{ll}
\underset{V_P \subseteq V(G_n)}{\text{minimize}} & (C_1^{\{V_P\}})^2 + (C_2^{\{V_P\}})^2 + \ldots + (C_m^{\{V_P\}})^2 \\
\text{subject to} & |V_P| = k,
\end{array}
\qquad (2)
$$

where $C_1^{\{V_P\}}, C_2^{\{V_P\}}, \ldots,$ and $C_m^{\{V_P\}}$ are the connected components after removing vertices in V_P from G_N.

Posets and Linear Extensions

Definition

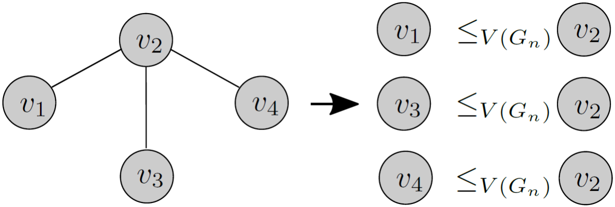
A non-strict partial order is a relation $\leq_S$ over a set S satisfying the following rules, for all $v_1, v_2, v_3 \in S$:

- $v_1 \leq_S v_1$ (reflexivity)
- if $v_1 \leq_S v_2$ and $v_2 \leq_S v_1$, then $v_1 = v_2$ (antisymmetry)
- if $v_1 \leq_S v_2$ and $v_2 \leq_S v_3$, then $v_1 \leq_S v_3$ (transitivity)

A **total order** has one more rule that every two elements in the set must be assigned a relation.

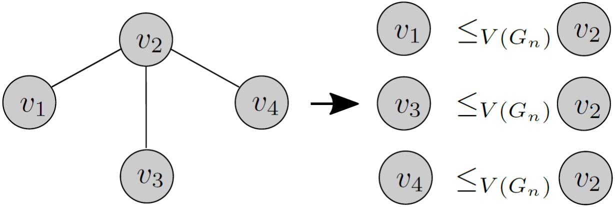
A **linear extension** $\leq_S^*$ of a partial order $\leq_S$ is a total order which preserve the relation in $\leq_S$, i.e., for all $v_1 \leq_S^* v_2$ whenever $v_1 \leq_S v_2$.

Posets and Rooted Trees



There is no relation between v_1, v_3 and v_4, hence this order is a **partial** order.

Linear Extensions and Cascading Failure



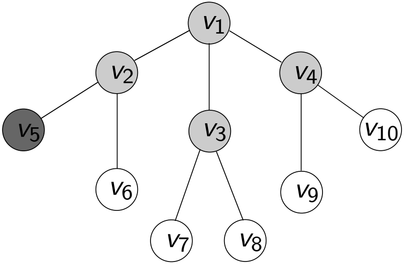
Consider a cascading failure on this graph with a specific order, for example $v_2 \rightarrow v_1 \rightarrow v_3 \rightarrow v_4$, then there is relation between any two vertices in this set, i.e., this specific order is a linear extensions on this posets (rooted tree). Intuitively, choosing the vertex with the maximum number of linear extensions to be protected is a good choice! [2]

Network Centrality to Determine Maximum Number of Linear Extensions of a Poset

Definition

Let G_n be a tree with n vertices, for any $u, v \in G_n$, let t_v^u be the subtree rooted at v by removing the edge (u, v) from G_n and slightly abusing the notation of the subtree size t_v^u as t_v^u.

For example, $t_{v_1}^{v_2} = 7$ and $t_{v_2}^{v_1} = 3$.

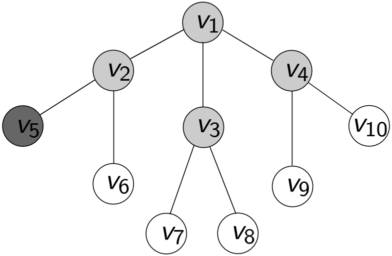


Definition

Define the branch weight of a vertex v in G_n by

$$\text{weight}(v) = \max_{c \in \text{child}(v)} t_c^v.$$

The vertex of G_n with the *minimum weight* is called the *centroid* of G_n [3]. For example, v_1 has the minimum weight, hence v_1 is the centroid.



Theorem

Let G_N be a general tree graph. Then, the rooted tree with the maximum number of linear extensions is rooted at v^* if and only if v^* is a centroid of G_N(proved in [4]).

Message Passing Algorithm to compute the Centroid of a Graph

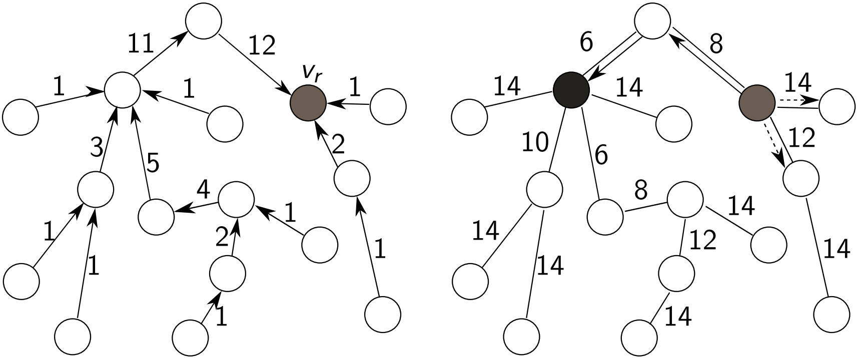
Let $M^{i \to j}$ denote the message from vertex i to vertex j. Let $\mathrm{Diff}(i,j)$ be defined by $\mathrm{Diff}(i,j) = |M^{i \to j} - M^{j \to i}|$.

Theorem

Given a tree G_n with n vertices.
$v_c \in G_n$ is the centroid if and only if $\forall v$ adjacent to v_c and $v_i, v_j \in V(G_n)$, $\min_{(v,v_c) \in E(G_n)} \{\mathrm{Diff}(v_c, v)\} \le \{\mathrm{Diff}(v_i, v_j)\}$. Moreover, for any $u \in G_n$, on the path from v_c to u say $(v_1, v_2, ..., v_D)$, where $v_1 = v_c$ and $v_D = u$. The sequence of $\mathrm{Diff}(v_i, v_{i+1})$ for $i = 1, 2...D$ is increasing.

Message Passing Algorithm to compute the Centroid of a Graph

$n = 12 + 1 + 2 + 1 = 16$



Assume G_N is a tree:

- When $|V_p| = 1$, we choose the centroid to be the solution.
- When $|V_p| > 1$, we use the centroid decomposition to select the protection set.

This may not be the optimal solution, but the performance can be bounded above.

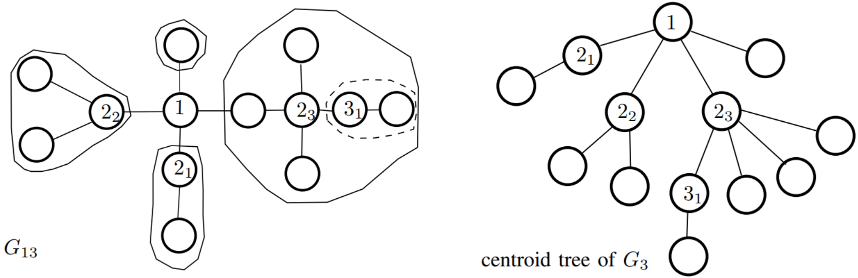
Theorem

Let $f(\{V_p\})$ denote the objective function in (2) and let V_p^* denote the optimal solution of (2). The centroid decomposition approach guarantees that

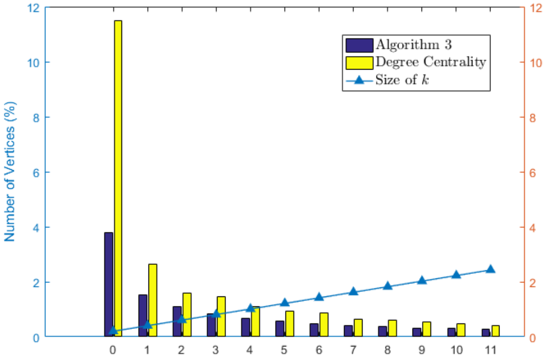
$$1 \leq \frac{f(\{V_p\})}{f(\{V_p^*\})} \leq c\frac{N}{k+1},$$

where k is the size of the protection set V_p and c is a small constant.

Centroid Decomposition

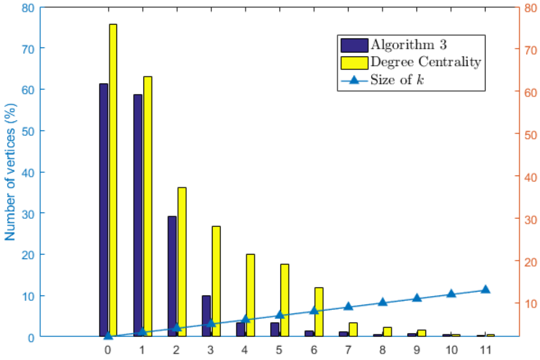


Experimental Results: $N = 4941$



A simulation result when G_N is a random tree. The y-axis represents the number of vertices in percentage and the x-axis represents each trial with different size of k.

Experimental Results: $N = 4941$



A simulation result when G_N is a real world network: Western United State Power Grid Network. The y-axis represents the number of vertices and the x-axis represents each trial with different size of k.

Network Centrality as Statistical Inference

In the reverse engineering perspective, we ask:

- Given a network centrality, what are the statistical inference optimization problems that it implicitly solves?
- Distance centrality and branch weight centrality solve the rumor source detection problem for degree-regular tree graphs.
- Betweenness centrality solves the protection node placement problem for a single node special case.
- Network centrality provides guiding principle on algorithm design and can compute exact or approximate solutions.

In the reverse engineering perspective, we ask:

- Given a stochastic optimization formulation over a network, how to transform it or to decompose it to one whose subproblems are graph-theoretic and can utilize network centrality, then solve or approximate the whole problem?
- Rumor source detection as a maximum-likelihood estimation problem solved by rumor centrality.
- Expected cascade size minimization problem solved by vaccine centrality.
- New algorithms can be designed based on message-passing (belief propagation) graph analysis
- Deep connections between network centrality on induced abstract data types with probability on trees and graphs.

Thank You!

http://www.cs.cityu.edu.hk/~cheewtan
Email: cheewtan@cityu.edu.hk

- P. D. Yu, C. W. Tan, and H. L. Fu, "Averting cascading failures in networked infrastructures: Poset-constrained graph algorithms," , *IEEE Journal of Selected Topics in Signal Processing*, p. forthcoming, 2018.
- D. Shah and T. Zaman, "Rumors in a network: Whos's the culprit?" *IEEE Trans. Information Theory*, vol. 57, no. 8, pp. 5163–5181, 2011.
- B. Zelinka, "Medians and peripherians of trees," *Arch. Math.*, vol. 4, no. 2, pp. 87–95, 1968.
- C. W. Tan, P. D. Yu, C. K. Lai, W. Zhang, and H. L. Fu, "Optimal detection of influential spreaders in online social networks," *Proc. of Conference on Information Systems and Sciences*, pp. 145–150, 2016.