Network Science: Principles and Applications[1]

CS 695 - Fall 2016

Amarda Shehu, Fei Li

Department of Computer Science
George Mason University

[1] Some materials are from wikipedia.

Figure : from the Huffington Post (08/14/2003) and a post posted by Edward H. Kennedy (06/03/2014)

The Northeast blackout of 2003 was a widespread power outage that occurred throughout parts of the Northeastern and Midwestern US and Ontario in Canada on August 14, 2003.
Some power was restored 7 hours later. Many others did not get their power back until two days later. In more remote areas it took nearly a week to restore power.
10 million people in Ontario and **45 million people in US affected**.
This power failure shuts down **265 power plants**, with **508 generating units**, including **10 nuclear power stations**.

- San Diego (2011, 3 million people affected during one day)
- Switzerland-France-Italy (2003, 57 million people affected during one day)
- Brazil (1999, 75 million people affected for many hours)
- Northeast (1965, 30 million people affected for many hours)
- ...

Electric disturbances in the U.S. increases **265%** in the number of major outages occurring since 1984 with an annual cost as high as **$188 billion**.
The Northeast blackout of 2003 wasn't caused by a storm or a terrorist attack but by an **errant tree branch**.
A tree branch falling on a power line caused a fault which was propagated throughout the grid.

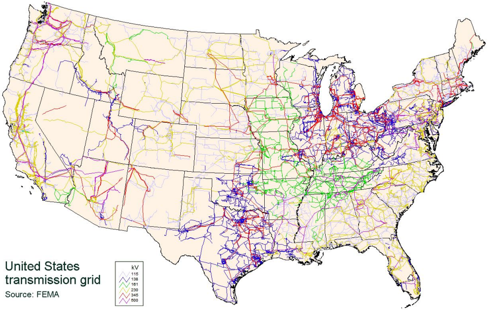


Figure : from GENI

The blackout's primary cause was a software bug in the alarm system at a control room of the FirstEnergy Corporation, located in Ohio.

1. A lack of alarm left operators unaware of the need to re-distribute power after overloaded transmission lines hit unpruned foliage, which triggered a **race condition**[2] in the control software.
2. What would have been a **manageable local blackout** cascaded into massive widespread distress on the electric grid.

[2] A race condition or race hazard is the behavior of an electronic, software or other system where the output is dependent on the sequence or timing of other uncontrollable events.

Vulnerability of fiber networks and power grids to geographically correlated failures

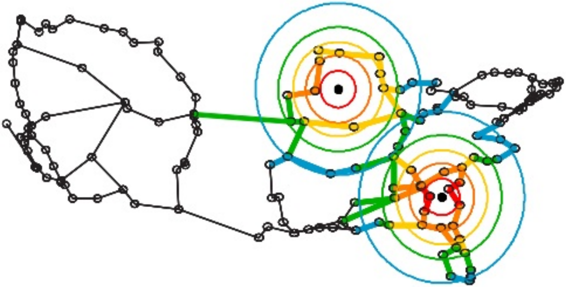
The following material[3] to be introduced was done by

- Saleh Soltan (EE, Columbia University)
- Mihalis Yannakakis (CS, Columbia University)
- Gil Zussman (EE, Columbia University)

[3] "Doubly balanced connected graph partitioning", ACM-SIAM SODA, 2017

1. Deterministic attacks

2. Probabilistic properties — realistic structures



Power grid vulnerability and cascading failures

Power grids are subject to cascading failures.

- Cascade propagation models differ from the classical epidemic/percolation-based models.

Power flow follows the laws of physics.

- AC (alternating current) model — Simulations are needed.

Control is difficult.

- It is difficult to "freeze" the network.

Modeling is difficult.

- Final report of the 2003 blackout — cause #1 was "inadequate system understanding"

Power grid islanding techniques to mitigate cascading failures in power grids[a].

[a] "Splitting Strategies for Islanding Operation of Large-Scale Power Systems Using OBDD-Based Methods" by Kai Sun, Da-Zhong Zheng, and Qiang Lu, IEEE Transactions on Power Systems, 2003

The challenge is to **partition the network into smaller connected components**, called *islands*, such that each island can operate independently for a while.

1. The power **supply and demand** at that island are **almost equal**.
2. The infrastructure in that island has the physical capacity to safely transfer the power from the supply nodes to the demand nodes.

Suppose the nodes of a graph are partitioned into **red** and **blue** nodes. Find a partition of the graph into two large connected subgraphs that splits approximately evenly both the red and the blue nodes.

Doubly balanced connected graph partitioning (Saleh Soltan, Mihalis Yannakakis, and Gil Zussman (Columbia University) ACM-SIAM SODA'17)

Let $G = (V, E)$ be a connected graph with a weight (supply/demand) function $p : V \rightarrow \{-1, +1\}$ satisfying

$$p(V) = \sum_{j \in V} p(j) = 0$$

The objective is to partition G into (V_1, V_2) such that for some constants c_p and c_s,

1. $G[V_1]$ and $G[V_2]$ are connected,
2. $\max(|p(V_1)|, |p(V2)|) \leq c_p$,
3. $\max\left(\frac{|V_1|}{|V_2|}, \frac{|V_2|}{|V_1|}\right) \leq c_s$,

The problem calls for a partition into two connected subgraphs that simultaneously balances two objectives:

1. the supply/demand within each part,
2. the sizes of the parts.

Soltan, Yannakakis, and Zussman's results summarized

The problem is NP-hard in general. Consider *k-connected* graphs[4].

1. **Assume** G **is** 2-**connected**.
   A solution with $c_p = 1$ and $c_s = 3$ always exists and can be found in polynomial time.
2. **Assume** G **is** 3-**connected and** $V \equiv 0 \pmod 4$.
   There is always a 'perfect' solution (a partition with $p(V_1) = p(V_2) = 0$ and $|V_1| = |V_2|$, and it can be found in polynomial time.
3. **The techniques can be extended, with similar results**.
   There exist polynomial-time algorithms for cases in which the weights are arbitrary (not necessarily ± 1).

[4] The *connectivity* of a graph $G = (V, E)$ is the minimum size of a set $S \subset V$ such that $G \setminus S$ is not connected. A graph is *k-connected* if its connectivity is at least k.

Theorem (Perfect weight partition)

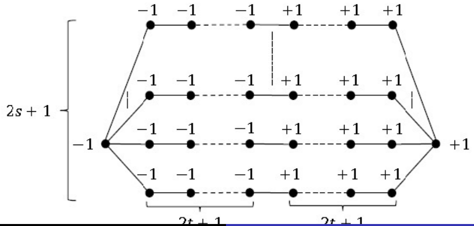
There is always a connected partition that is perfect with respect to the weight objective, $p(V_1) = p(V_2) = 0$.

Theorem (Perfect size partition)

There is always a connected partition that is perfect with respect to the size objective, $|V_1| = |V_2|$.

Theorem (Combined case — negative answer)

If $c_p = 0$*, then for any* $c_s < \frac{|V|}{2} - 1$ *(If* $c_s = 1$*, then for any* $c_p < \frac{|V|}{6}$*),* G *below does not have a solution.*



Theorem (Combined case — essentially perfect partition)

Let $G = (V, E)$ be a 3-connected graph. G has a partition that is **essentially perfect** *with respect to both objectives. (V is a multiplier of 4.)*

Proof.

Start with the special case in which G has a triangle. The main idea is:

1. Number (index) the nodes such that the size constraint ($|V_1| = |V_2| = \frac{|V|}{2}$) is satisfied.
2. Gradually meet the supply/demand constraint ($p(V_1) = p(V_2) = 0$) while the size constraint is kept being satisfied.

□

Definition (*st*-numbering of a graph)

Given any edge (s, t) in a 2-connected graph G, an *st-numbering for G* is a numbering for the nodes in G defined as follows[a]: the nodes of G are numbered from 1 to n so that s receives number 1, node t receives number n, and every node except s and t is adjacent both to a lower-numbered and to a higher-numbered node.
It is shown in[b] that such a numbering can be found in $O(|V| + |E|)$.

[a] A. Lempel, S. Even, and I. Cederbaum, "An algorithm for planarity testing of graphs", International symposium of Theory of graphs, 1967.

[b] S. Even and R. E. Tarjan, "Computing an *st*-numbering", Theoretical Computer Science, 1976.

Theorem ([a])

[a] L. Lovasz, "A homology theory for spanning tress of a graph", Acta Mathematica Hungarica, 1977. E. Gyori, "On division of graphs to connected subgraphs", Combinatorics, 1976

Let $G = (V, E)$ be a k-connected graph. Let $n = |V|$, $v_1, v_2, \ldots, v_k \in V$ and let $n_1, n_2, \ldots, n_k$ be positive integers satisfying $n_1 + n_2 + \cdots + n_k = n$. Then, there exists a partition of V into $(V_1, V_2, \ldots, V_k)$ satisfying $v_i \in V_i$, $|V_i| = n_i$, and $G[V_i]$ is connected for $i = 1, 2, \ldots, k$.

1. Although the existence of such a partition has long been proved, there is no polynomial-time algorithm to find such a partition for $k > 3$.
2. For $k = 2$, it is easy to find such a partition using *st*-numbering.
3. For $k = 3$, [5] provided an $O(n^2)$ algorithm using the nonseparating ear decomposition of 3-connected graph.

[5] K. Wada and K. Kawaguchi, "Efficient algorithms for tripartitioning triconnected graphs and 3-edge-connected graphs", Graph-Theoretic Concepts in Computer Science, 1994.

Definition (Convex embedding of the k-connected graphs[a])

[a] N. Linial, L. Lovasz, and A. Wigderson, "Rubber bands, convex embeddings and graph connectivity", Combinatorica, 1988.

Let $Q = \{q_1, q_2, \ldots, q_m\}$ be a finite set of points in R^d. The *convex hull* $conv(Q)$ of Q is the set of all points

$$\sum_{i=1}^{m} (\lambda_i q_i)\,,\ \text{subject to}\ \sum_{i=1} \lambda_i = 1.$$

The *rank* of Q is defined by $rank(Q) = 1 + dim(conv(Q))$.
Let G be a graph and $X \subset V$. A *convex X-embedding of G* is any mapping

$$f : V \rightarrow R^{|X|-1} \Rightarrow \forall v \in V \setminus X, f(v) \in conv(f(N(v))).$$

Q is in *general position* if $rank(S) = d + 1$ for every $(d+1)$-subset $S \subseteq Q$. The convex embedding is in *general position* if the set $f(V)$ of the points is in general position.

Theorem (Rubber band drawing of a graph[a])

[a] N. Linial, L. Lovasz, and A. Wigderson, "Rubber bands, convex embeddings and graph connectivity", Combinatorica, 1988.

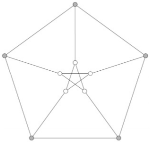
Let G be a graph on n vertices and $1 < k < n$. Then the following two conditions are equivalent:

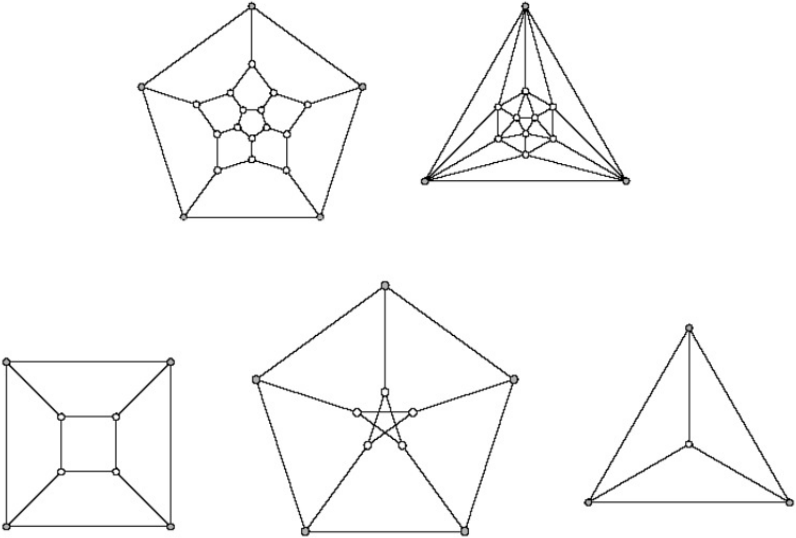
1. *G is k-connected*
2. *For every $X \subset V$ with $|X| = k$, G has a convex X-embedding in general position.*

Algorithm 1 Rubber band interpretation

1: All edges are rubber bands.
2: Nail down some nodes X in the plane, let the graph go.

The algorithm converges to a unique state — rubber band representation extending X. The graph is at equilibrium.





Algorithm 2 Mass-spring model

```
Assume nodes on the boundary know their location.
Fix the nodes on the outer boundary.
while a node can move more than distance δ do
    every node moves to the center of gravity of its neighbors,
```

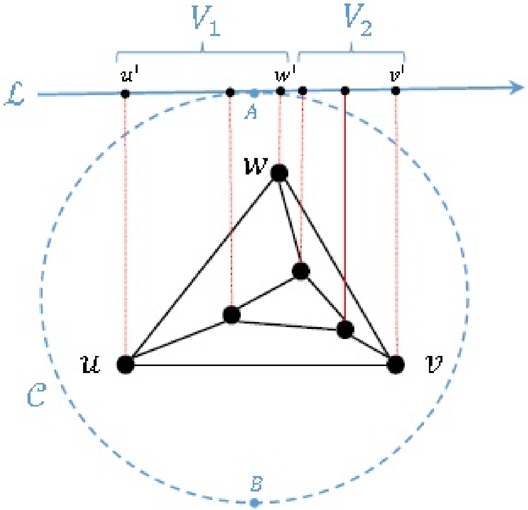
$$x_i \leftarrow \frac{\sum_{j \in N(i)} x_j}{d_i}$$

```
end while
```

End of preliminaries — the last step

1. If G is a k-connected graph, then for every $X \subset V$ with $|X| = k$, G has has a convex X-embedding in general position and this embedding can be found in polynomial-time[6].
2. Assume $u, v, w \in V$ form a triangle in G. Set $X = \{v, u, w\}$. Consider a circle C around the triangle $f(u)$, $f(v)$, $f(w)$. Also consider a directed line L tangent to the circle C at point A. If we project the nodes of G onto the line L, since the embedding is convex and also $\{u, v\}, \{u, w\}, \{w, v\} \in E$, the order of the nodes' projection gives an *st*-numbering between the first and the last node.
3. Since the embedding is in general position, there are exactly two points on every line that connects two points $f(i)$ and $f(j)$, so V_1 changes at most by one node leaving V_1 and one node entering V_1 at each step as we move L. Hence, $p(V_1)$ changes by either ± 2 or 0 value at each change.

[6] N. Linial, L. Lovasz, and A. Wigderson, *Rubber bands, convex embeddings and graph connectivity*, Combinatorica, 1988


V_1
V_2
u'
w'
v'
$\mathcal{L}$
A
w
u
v
$\mathcal{C}$
B

Open problems

1. For 2-connected graphs it is always possible to achieve both objectives at the same time, and for 3-connected graphs there is a partition that is essentially perfectly balanced in both objectives. Furthermore, these partitions can be computed in polynomial time.
   Can we reduce the running time further?
2. The novel techniques used in this paper can be applied to partitioning heterogeneous networks in various contexts.
   Can we apply randomization here? Can we introduce the 'cut' here?
3. **How can we extend the theory and algorithms to find doubly balanced connected partitions to more than two parts**?