**Shortest Paths-based Centrality**

\[ \sigma_{st}(i) \] is a number of \( s-t \) shortest paths containing \( i \)

\[ \sigma_{st} \] is a number of all \( s-t \) shortest-paths

**Observation:** In practice, communication or transport of goods in networks follow different kinds of paths that tend to be shortest.

**Question:** How much work can be done by a node?  

**Stress Centrality**

\[ c_S(i) = \sum_{s \neq i} \sum_{t \neq i} \sigma_{st}(i) \]  
for nodes

\[ c_S(ij) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(ij) \]  
for edges

Relation between stress centralities

\[ c_S(i) = \frac{1}{2} \sum_{i,j \in \Gamma(i)} c_S(ij) - \sum_{i \neq s \in V} \sigma_{si} - \sum_{i \neq t \in V} \sigma_{it} \]

**Betweenness Centrality**

\[ c_B(i) = \sum_{s \neq i} \sum_{t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \]

Interpretation: BC is a quantification of communication control which has a vertex over all pairs of nodes.
Communication Control Quantification: Example

From Brandes “Network Analysis”

\[ c_S(u_i) = 16 \text{ and } c_B(u_i) = \frac{1}{3}, \ i = 1, 2, 3 \text{ and } c_S(v) = 16 \text{ but } c_B(v) = 1 \]

**BC for edges**

\[ c_B(ij) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{st}(ij)}{\sigma_{st}} \]

**Problem**

BC is very sensitive to network dynamics (edge/node removal/addition)

**Solution?**

\( \varepsilon \)-BC: in BC replace all shortest s-t paths with all shortest paths that are not longer than \((1+\varepsilon)\sigma_{ij} \)
We have to be careful with shortest path-based centralities, example
\[ c_B(i) = \sum_{ij \in \Gamma+(i)} c_B(ij) - (n - 1) = \sum_{ji \in \Gamma-(i)} c_B(ji) - (n - 1) \]

Exercise: Prove that in directed graphs the relation between centralities holds

\[ G[T_{ij}] \]

is an illuminating measure

Another measure based on \( T_{ij} \) is the edge centrality index

\[ c_{ts}(ij) = |H|, \text{ where } H \text{ is a minimum vertex cover in } G[T_{ij}] \]

that can be used in characterization of networks with hierarchical organization.

Note: minimum size of a vertex cover = the size of a maximum matching in bipartite graphs
Counting All Paths

Simple, directed $G = (V, E)$, no loops

$$\forall i \in V \ c_K(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha_k (A^k)_{ji}$$

The sum converges with restricted $\alpha_k$.

**Theorem.** If $A$ is the adjacency matrix of $G$, $\alpha > 0$, and $\lambda_1$ the largest eigenvalue of $A$, then

$$\lambda_1 < \frac{1}{\alpha} \iff \sum_{k=1}^{\infty} \alpha^k A^k \text{ converges}$$

and $c_K = (I - \alpha A)^{-1} \cdot 1_n$.

Leo Katz “A new status index derived from sociometric analysis”
Vitality (robustness)

Let $G$ be the set of all simple, undirected and unweighted graphs $G = (V, E)$ and $f : G \rightarrow \mathbb{R}$ be any real-valued function on $G \in G$. A vitality index $\mathcal{V}(G, x)$ is the difference of the values of $f$ on $G$ and on $G$ without element $x$, i.e., $\mathcal{V}(G, x) = f(G) - f(G \setminus x)$.

Max-flow Betweenness Vitality

Q: How much flow must go over a vertex $i$ in order to obtain the maximum flow value? How does the objective function value change if we remove $i$ from the network?

$$c_{mf}(i) = \sum_{s, t \in V} \frac{f_{st}(i)}{f_{st}}, \text{ where } f_{st}(i) = f_{st} - \max \text{ s-t-flow in } G \setminus i$$
Closeness Vitality

Wiener index of a network

\[ I_W(G) = \sum_{i,j \in V} \delta_{ij} \]

or in terms of closeness centrality

\[ I_W(G) = n \cdot \sum_{i \in V} \frac{1}{c_i} \]

Closeness vitality is defined on both vertices and edges

\[ c_{CV}(x) = I_W(G) - I_W(G \setminus \{x\}) \]
Current Flow

- Electrical network is an undirected, simple, connected graph $G = (V, E)$
- Conductance function $c: E \rightarrow \mathbb{R}$
- Supply function $b: V \rightarrow \mathbb{R}$ (external electrical current enters and leaves network)
- Positive $b = \text{entering current}$
- Negative $b = \text{leaving current}$
  \[ \sum_{i \in V} b(i) = 0 \]
- Direction of the current: each edge $ij$ is oriented arbitrarily

Fuction $x: \overline{E} \rightarrow \mathbb{R}$ is called current if

\[ \sum_{ij \in E} x_{ij} - \sum_{ji \in E} x_{ji} = b(i) \quad \text{and} \quad \sum_{ij \in C} x_{ij} = 0 \]

for every cycle $C \subset E$.

A function $p: V \rightarrow \mathbb{R}$ is a potential if $p(i) - p(j) = x_{ij}/c_{ij}$ for all $ij \in \overline{E}$. As an electrical network $N = (G, c)$ has a unique current $x$ for any supply $b$, it also has a potential $p$ that is unique up to an additive factor.
Given edge weights $c(i)$, we define electrical network Laplacian $L$.

We can find $p$ and $b$ by solving $Lp=b$.

**Current-Flow Betweenness Centrality**

Unit $s-t$-supply $b_{st}$ is a supply of one unit that enters the network at $s$ and leaves at $t$, that is, $b_{st}(s) = 1$, $b_{st}(t) = -1$, and $b_{st}(i) = 0$ for all $i \in V \setminus \{s, t\}$.

Throughput of $i \in V$ with respect to a unit $s-t$-supply $b_{st}$ is defined as

$$\tau_{st}(i) = \frac{1}{2} \left( - |b_{st}(i)| + \sum_{ij \ni i} |x(-i,j)| \right)$$

$$c_{CB}(i) = \frac{1}{(n-1)(n-2)} \sum_{s,t \in V} \tau_{st}(i)$$
How to compare different centrality concepts?

Normalization in one network

$p$-norm of the centrality vector for concept $X$

$$||c_{X}||_{p} = \begin{cases} 
\left(\sum_{i=1}^{n} |c_{X_i}|^p\right)^{1/p} & 1 \leq p \leq \infty \\
\max_{i}\{ |c_{X_i}| \} & p = \infty 
\end{cases} \implies \frac{c_{X}}{||c_{X}||_{p}} \implies c_{X_i} \leq 1$$

separation of positive and negative values of $c_{X}$

$$c'_{X} = \begin{cases} 
c_{X_i}/\left(\sum_{j:c_{X_j}>0} |c_{X_j}|^p\right)^{1/p} & c_{X_i} > 0 \\
0 & c_{X_i} = 0 \\
c_{X_i}/\left(\sum_{j:c_{X_j}<0} |c_{X_j}|^p\right)^{1/p} & c_{X_i} < 0 
\end{cases}$$

**Exercise:** Is $c'_{X}$ a norm? Prove or disprove.

Freeman “Centrality in social networks: Conceptual clarification”
Normalization for different networks

Point-centrality

\[ c''_{X_i} = c_{X_i} / \left( \max_{G \in \mathcal{G}_n} \max_{i \in V(G)} c_{X_i} \right) \]

set of all graphs with \( n \) vertices

Examples

- Degree centrality = normalization by factor \((n-1)\)

- Shortest paths betweenness centrality

\[ c_B(i) = \sum_{s \neq i} \sum_{t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \]

What is the upper bound (or normalization factor)?

**Star graph**, \( c_B(i) = (n-1)(n-2)/2 \)

- Closeness centrality

\[ \forall i \in V \quad C_i = 1 / l_i, \text{ where } l_i = \frac{1}{n} \sum_{j \in V} \delta_{ij}, \delta_{ij} = \text{length of } i - j \text{ shortest path} \]

upper bound (or normalization factor) is \( 1/(n-1) \)
Summary and How Does It Work in Practice

Categories of centrality measures

*Reachability.* A vertex is supposed to be central if it reaches many other vertices. Centrality measures of this category are the degree centrality, the centrality based on eccentricity and closeness, etc. All of these centralities rely on the distance concept between pairs of nodes.

*Amount of flow.* Based on the amount of flow $f_{st}(i)$ from a vertex $s$ to a vertex $t$ that goes through a vertex or an edge $i$. Can be based on current flow and random walks (will see how it works in Spectral Methods). Also measures that are based on the enumeration of shortest paths, stress centrality; betweenness centralities measure the expected fraction of times a unit flow goes through the element if every vertex $s$ sends one unit flow consecutively to every other vertex $t$.

*Vitality.* Based on the vitality, i.e., the centrality value of an element $x$ is defined as the difference of a real-valued function $f$ on $G$ with and without the element. Recall, a general vitality measure is identified $f(G) - f(G\\{x\})$. Such as the max-flow betweenness vitality.

*Feedback.* Centrality measures that are based on implicit definitions of a centrality given by the abstract formula $c(i) = f(c(v1), \ldots, c(vn))$, where the centrality value of $i$ depends on the centrality values of all vertices. Includes Katz and some of eigenvector-based centralities.

from U. Brandes “Network Analysis”
**Introduction to Network Analysis**

### Application

- **Network**

#### Aspect to be evaluated

- Reachability
- Flow
- Vitality
- Feedback
- ...

#### Personalization: rootset of nodes, weights, not all paths, ...

#### Term operator wrt other elements: max, sum, average, ...

### Normalization and Comparison

#### Centrality Index

**Approximation (hopefully linear), Computational Tools (mat/eig solvers)**