Closeness Centrality

δ_{ij} = length of $i - j$ shortest path

$\forall i \in V \;\; C_i = 1/l_i$, where $l_i = \frac{1}{n}\sum_{j\in V}\delta_{ij}$

large for peripheral vertices

Problems:

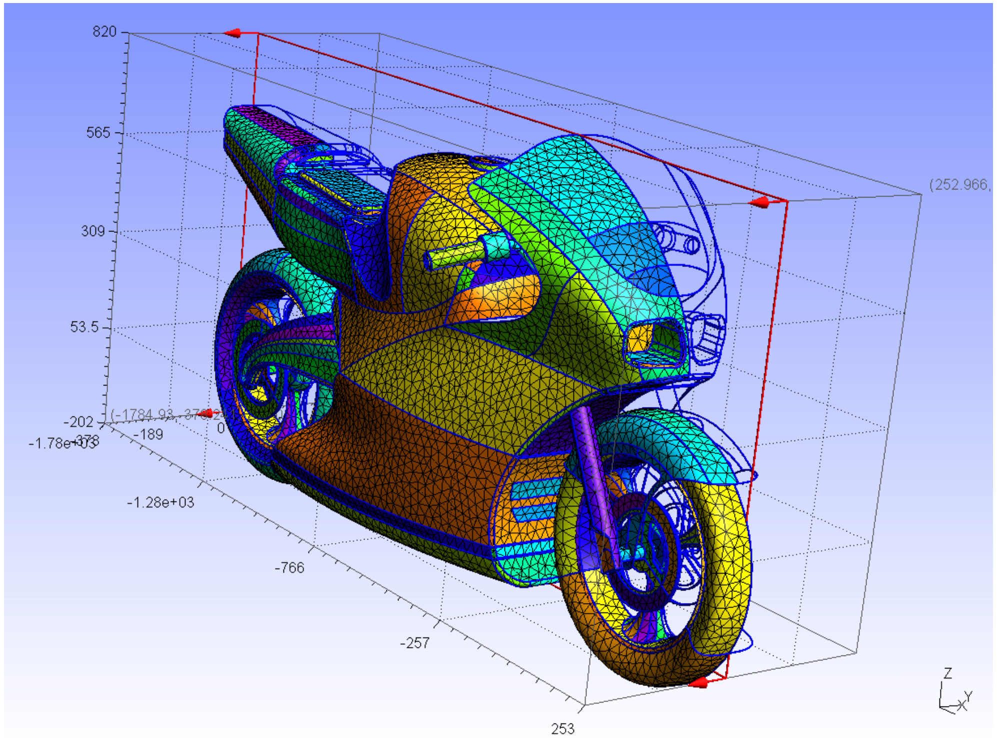
- *i-j* path is infinity for different connected components

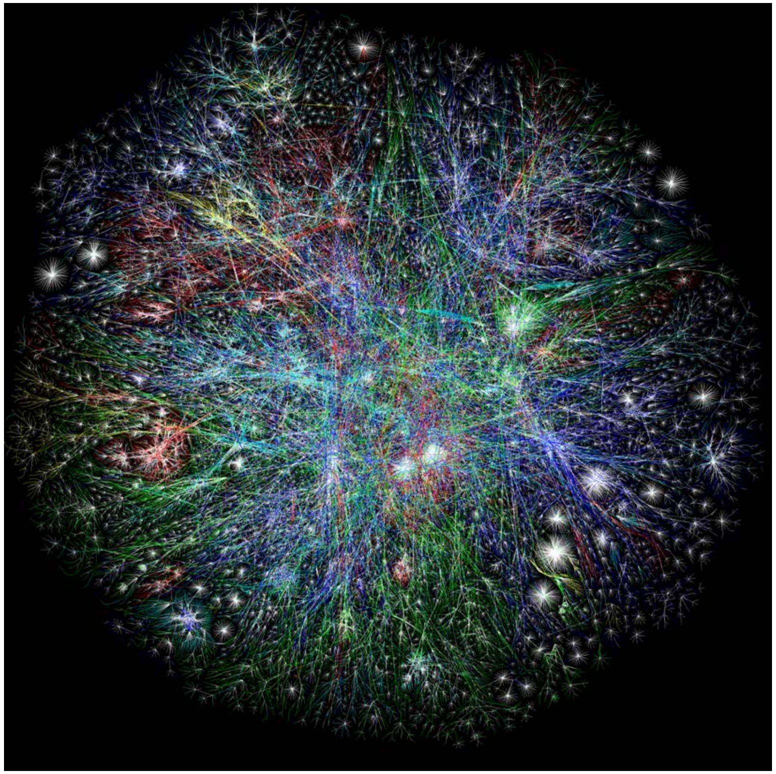
  Solution: Harmonic Mean Distance Centrality $\forall i \in V \;\; C_i' = \frac{1}{n-1}\cdot\sum_{j\neq i} 1/\delta_{ij}$

- In practice *C* spans relatively small range of values
- Highly sensitive (one edge removal and all distances are increased)
- Geodesic distances are integers, **often small because distance increases logarithmically with the size of the network**
- Example: Internet Movie Database

  highest centrality = 0.4143, lowest centrality = 0.1154, ratio 3.6 for 500K actors

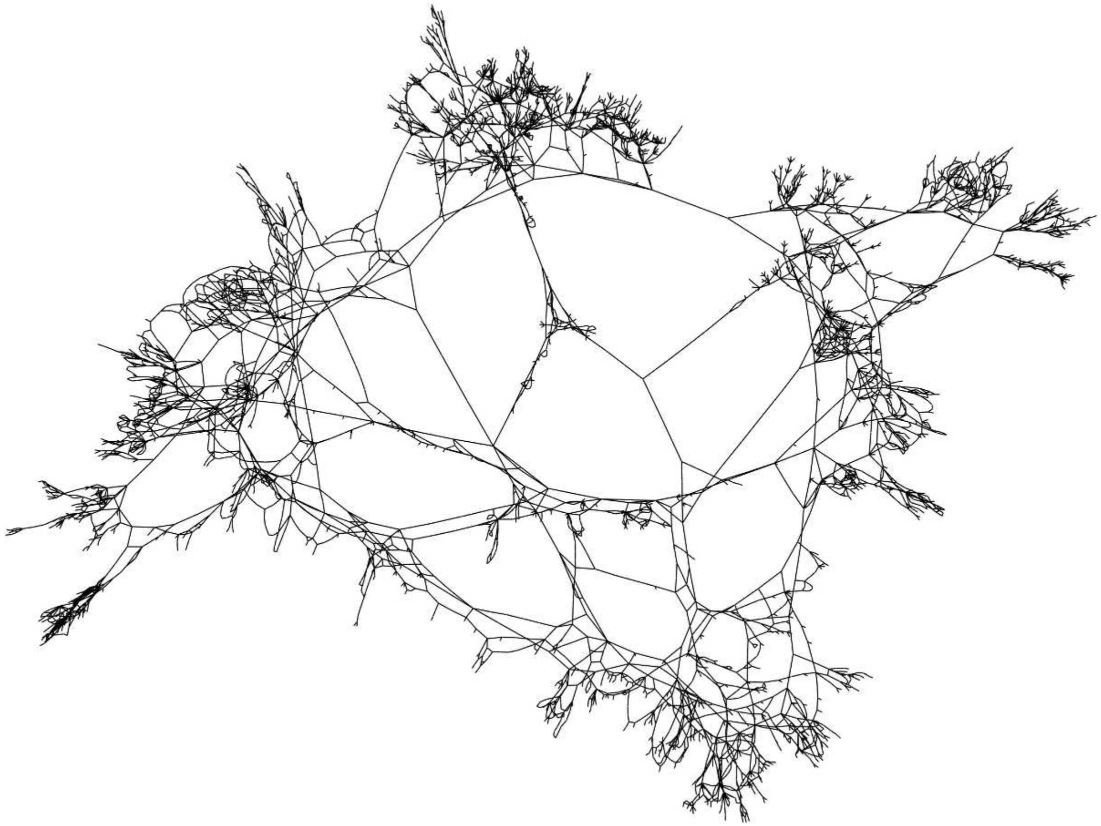
Closeness Centrality





http://geuz.org/gmsh/

Closeness and Degree-based Centralities



Example in Gephi: degrees vs eigenvector centrality
Some correlation is expected for certain types of networks

Enumeration of Shortest Paths-based Centrality

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

σ_{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.

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

Disadvantage?

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

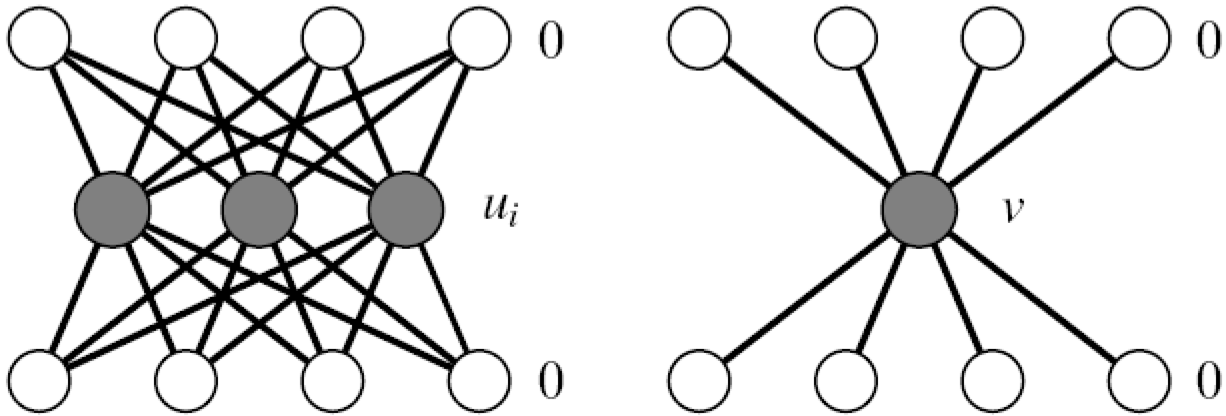
Relation between stress centralities $c_S(i) = \frac{1}{2} \sum_{ij \in \Gamma(i)} c_S(ij) - \sum_{i \neq s \in V} \sigma_{si} - \sum_{i \neq t \in V} \sigma_{it}$

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



$$c_S(u_i) = 16 \text{ and } c_B(u_i) = \tfrac{1}{3},\ i = 1,2,3 \text{ and } c_S(v) = 16 \text{ but } c_B(v) = 1$$

from Brandes "Network Analysis"

BC for edges

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

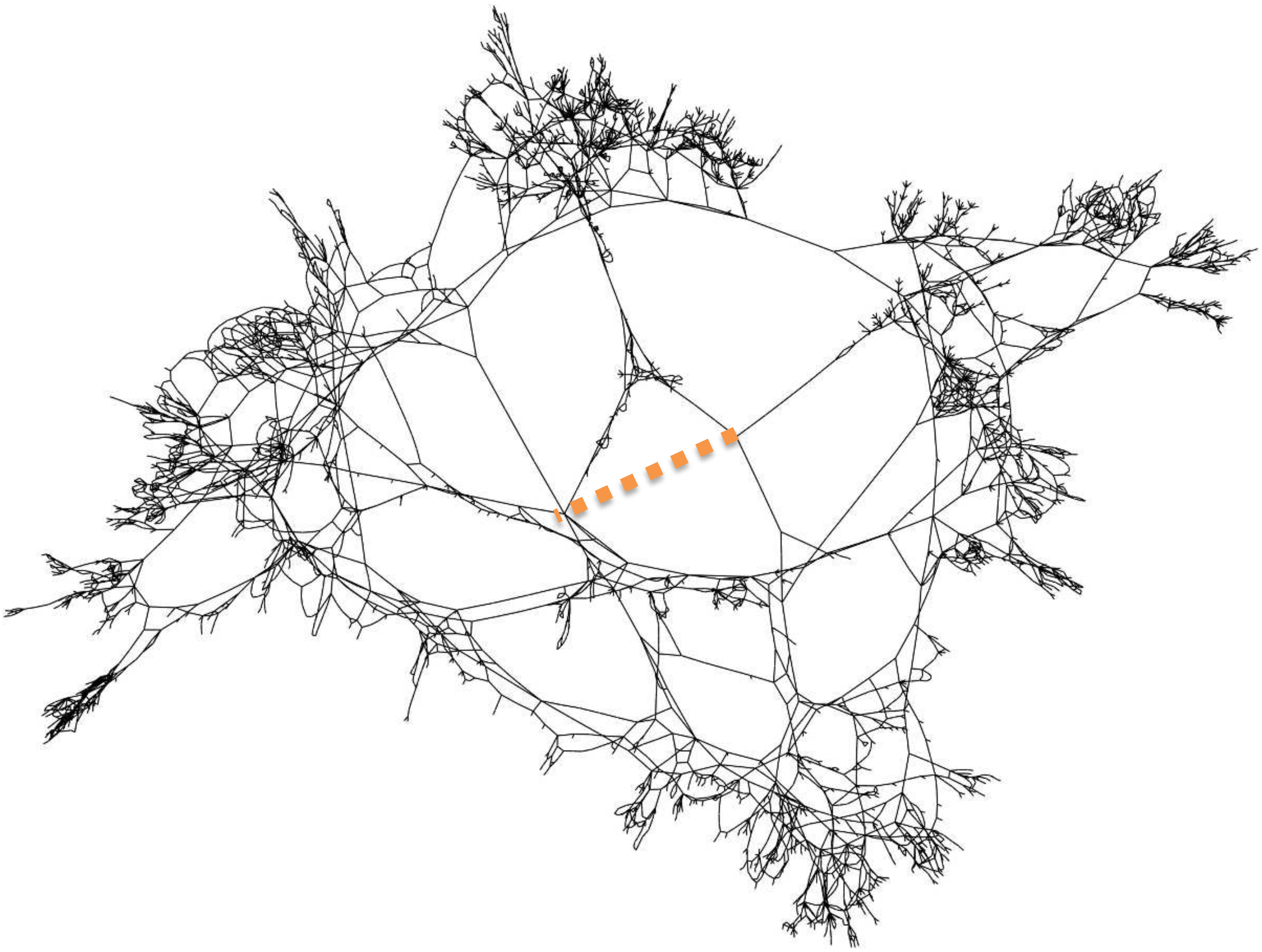
Problem

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

Solution?

ε-BC: in BC replace all shortest s-t paths with all shortest paths that are not longer than $(1+\varepsilon)\delta_{ij}$

We have to be careful with shortest path-based centralities, example



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

$$c_B(i) = \sum_{ij \in \Gamma^+(i)} c_B(ij) - (n-1) = \sum_{ji \in \Gamma^-(i)} c_B(ji) - (n-1)$$

Submit by 1/30/2014

Traversal Sets

$\forall ij \in E$ we define the edge's traversal set

$$T_{ij} = \{(s,t) \in V \times V |\text{ some shortest path } s-t \text{ contains } ij\}$$

and traversal set induced graph

$$G[T_{ij}] = (V', E'), \text{ where } V' = \{k \in V | (k,t) \in T_{ij} \text{ or } (s,k) \in T_{ij}\}, \text{ and}$$
$$E' = \{(s,t) \in T_{ij}\}.$$

Homework: Prove that $G[T_{ij}]$ is bipartite.
Submit by 1/30/2014

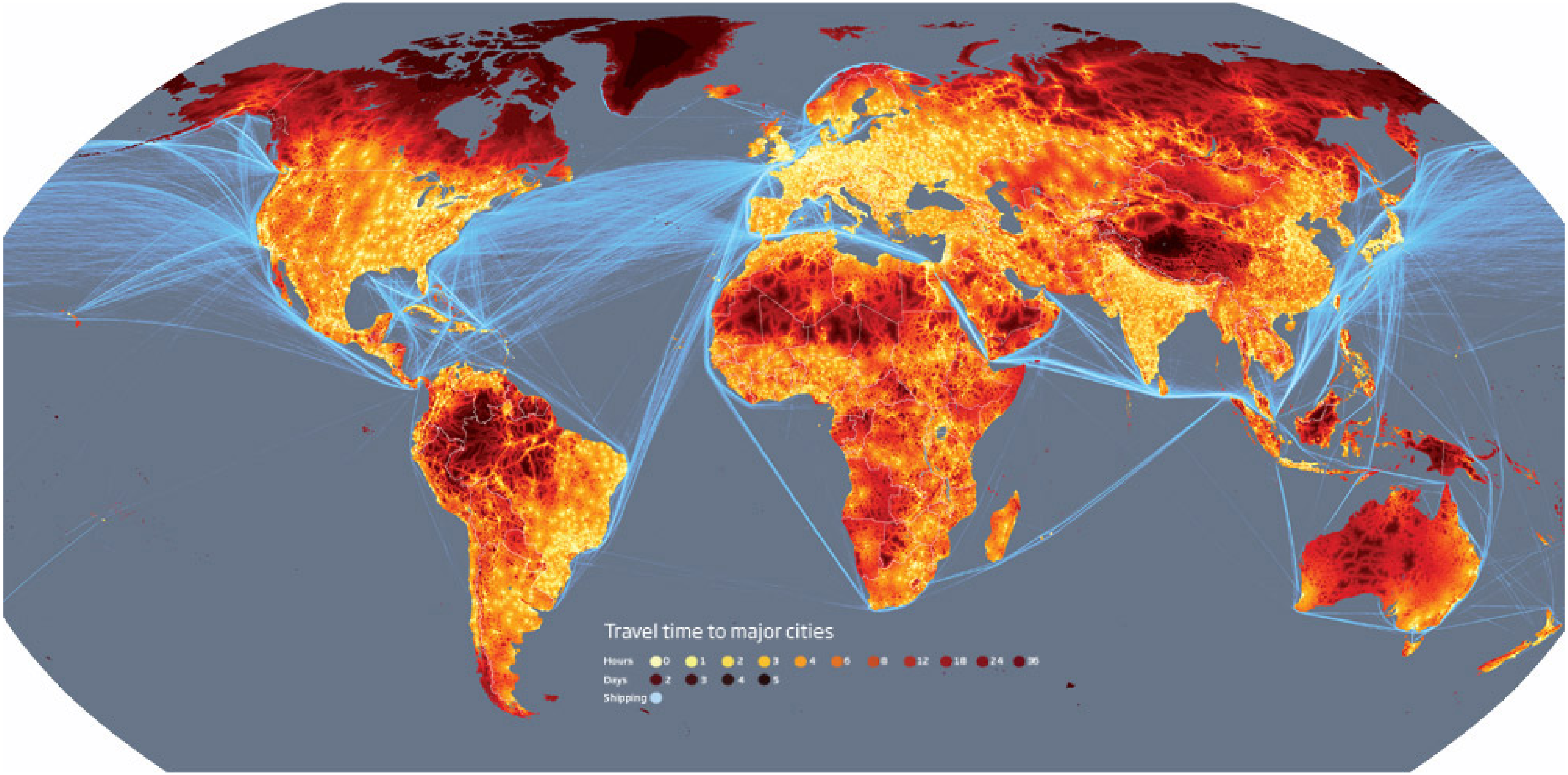
$|T_{ij}|$ is an illuminating measure
Another measure based on T_{ij} is the edge centrality index

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

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

Note: there is a theorem that says the following: in bipartite graphs the minimum size of a vertex cover = the size of a maximum matching

Transportation Roads Density



www.newscientist.com