Multivariate analysis and complex networks

3: Complex Networks & Metrics

Antonio Scala

CNR-ISC, the Institute for Complex Systems

INFN School of Statistics
Ischia, May 2017

Closeness Centrality

Paths

- l_{ij} = minimum path among node i and node j
- $\langle l \rangle = \frac{1}{n(n-1)} \sum_{\{ij\} \in E} l_{ij}$ average path length
- $\delta = \max_{\{ij\} \in E} l_{ij}$diameter
- Closeness Centrality $close_i = \sum_j \frac{1}{l_{ij}}$

Path Calculation

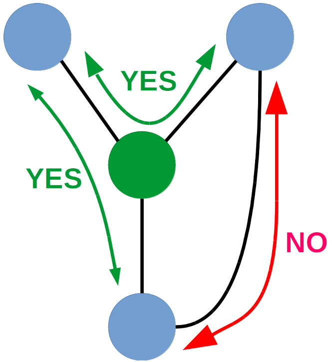
- Dijkstra algorithm
- Bellman-Ford
- brute force: calculate A^k, $k < |V|$

Betweeness Centrality

Betweeness

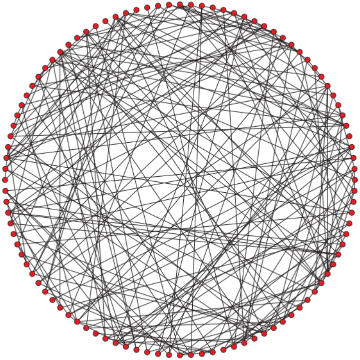
- b_i = number of shortest paths passing through node i
- b_e = number of shortest paths passing through edge e

inspired by studies on communication networks and transport networks



Erdos-Renyi Networks

- n nodes, p probability that two nodes are connected
- $|E| \sim p \cdot n \cdot (n-1)/2$
- $\langle k \rangle = 2|E|/n \sim pn$
- $p(k) \sim e^{\langle k \rangle} \frac{\langle k \rangle^{-k}}{k!}$
- $d \sim \ln n / \ln \langle k \rangle$
- $\langle c \rangle = p = \langle k \rangle / n$



Clustering Coefficient

A node i has a high CLUSTERING COEFFICIENT if its neighbors resemble a clique, i.e. the subgraph induced by its k_i neighbors has $k_i(k_i-1)$ directed edges or $k_i(k_i-1)/2$ undirected edges

- $c_i = \begin{cases} \frac{|\{e_{ik}\}_i|}{k_i(k_i-1)} & \text{if G is undirected} \\ \frac{2|\{e_{ik}\}_i|}{k_i(k_i-1)} & \text{if G is directed} \end{cases}$

with $j,k \in nn(i)$

Calculation:

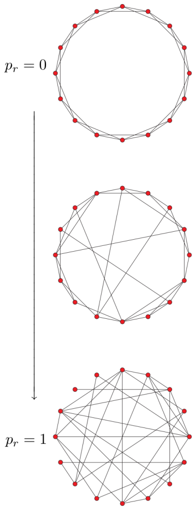
- $j,k \in nn(i) \iff A_{ij}A_{ik} \neq 0$
- $j,k \in nn(i) \wedge \exists e_{jk} \iff A_{ij}A_{ik}A_{jk} \neq 0$
- undirected graphs $|\{e_{ik}\}_i| = \frac{1}{2}\left(A^3\right)_{ii}$

Watts-Strogatz Networks

1. start with a LATTICE graph of dimension d
2. rewire a fraction p_r of links

- $p_r = 0 \rightarrow \delta \sim n^{1/d}$
- $p_r = 1 \rightarrow$Erdos-Renyi graph

$$p_r \sim 1/n \rightarrow \begin{array}{c} \delta \sim \ln n \\ c \sim c(p_r = 0) \end{array}$$



Degree Centrality

Degree

- $k_i^{in} = \sum_j A_{ji}$
- $k_i^{out} = \sum_j A_{ij}$
- $k_i = k_i^{in} + k_i^{out} - A_{ii}$

Degree Distribution

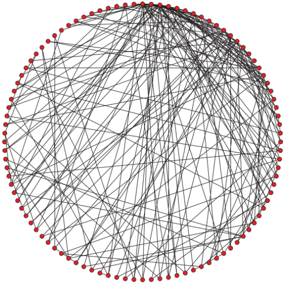
- $\langle k \rangle = \frac{1}{|V|} \sum_i k_i$
- $P(k)$ degree distribution
- k_{max} maximum degree

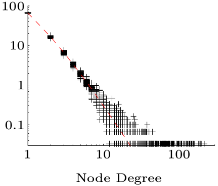
Barabasi-Albert Networks

"Rich get Richer"

1. start with a seed graph
2. add a new node v
3. add m new links from node v to the other nodes choosing a node i with a probability proportional to its degree d_i

- $\langle k \rangle \to 2m$
- $p(k) \sim k^{-3}$
- $d \sim \ln n / \ln(\ln n)$



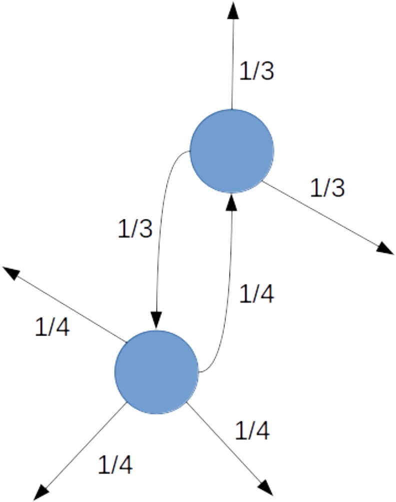


Random Walk Centrality

Random Walk

- transition matrix $M_{ij} = A_{ij}/d_j$
- discrete $p^{t+1} = Mp^t$
- continuous $\partial_t p_i \propto (Mp)_i - p_i$
- stationarity $p^\infty = Mp^\infty$

$$p_i^\infty \propto d_i$$



Adjacency Matrix Centralities

Eigen Centrality

$$Ax = \lambda_{max} x$$

Katz Centrality

$$h_i = \sum_j \sum_{k=0}^{\infty} \alpha^k \left(A^k\right)_{ij}$$

$$h = \alpha A(x + 1)$$

Katz, stress & Green

idea: the stress s_i propagates to its nearest neigbours

- $s^{t+1} = s^t + \alpha A s^t$
- $s^\infty = (1 - \alpha A)^{-1} s^0 = H^\alpha s^0$
- H^α converges if $\|\alpha A\| < 1$
- $h_i = \sum_j H^\alpha_{ji}$ stress induced by node i

inspired by studies on financial networks and input-output model

Connected Components

A COMPONENT of G is a disconnected subgraph of G. It can be also referred as a CLUSTER (for CM physicists)

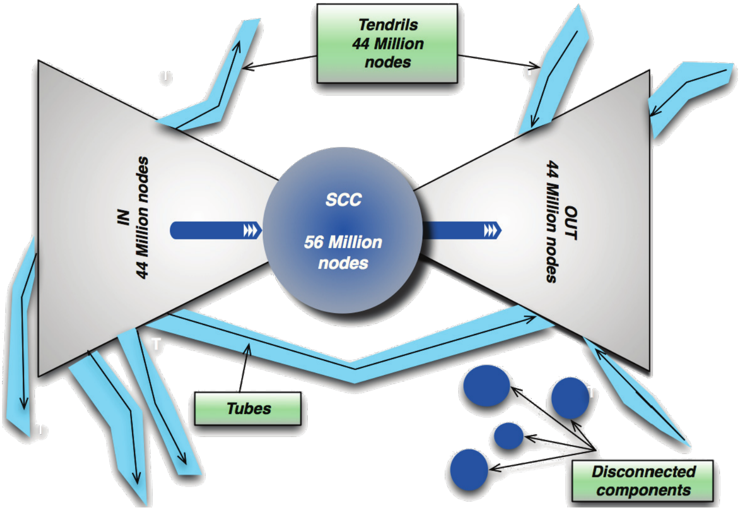
Component

- N_C number of connected components
- $P\infty$ fraction of nodes in the largest component
- $P(|C|)$ fraction of components of size $|C|$

Calculation

- Breadth/Depth first search
- Hoshen-Koppelman algorithm
- Matrix factorization
- Kosaraju algorithm, Tarjan algorithm (directed networks)

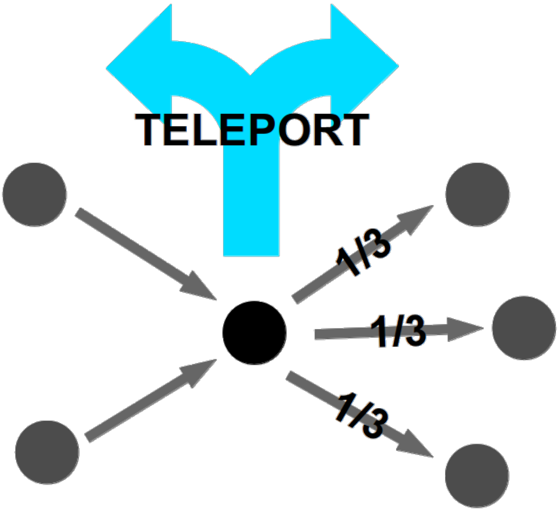
Bow-Tie structure of digraphs



Random Surfer

- web surfer follows at random the outgoing links
- web surfer gets bored at random (with probability $1-\rho$) and switches completely search
- ρ is related to the (measurable) average number of links followed by an user before switching search

RW on directed graphs



PageRank centrality

$$r_i = \frac{1-\rho}{N} + \rho \sum_j \frac{A_{ji}}{d_j^{out}} r_j$$

The process is a superposition of a random walk and a renewal process