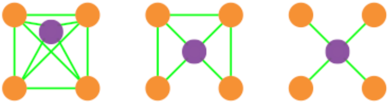
Clustering coefficient

How neighbors of a given node connected to each other

- *Local clustering coefficient* (per vertex):

$$C_i = \frac{\text{number of links in } \mathcal{N}_i}{k_i(k_i - 1)/2}$$

(a)



C_i=1 C_i=1/2 C_i=0

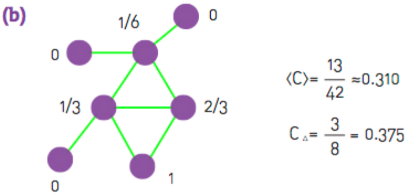
- Average clustering coefficient:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

Clustering coefficient

- *Global clustering coefficient*:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}}$$



igraph:transitivity(type="global")

Statistical properties

- Power-law degree distribution
- Small average path length
- High clustering coefficient (transitivity)
- Gigantic connected component

NATIONAL RESEARCH
UNIVERSITY

Mathematical models of networks

Network models

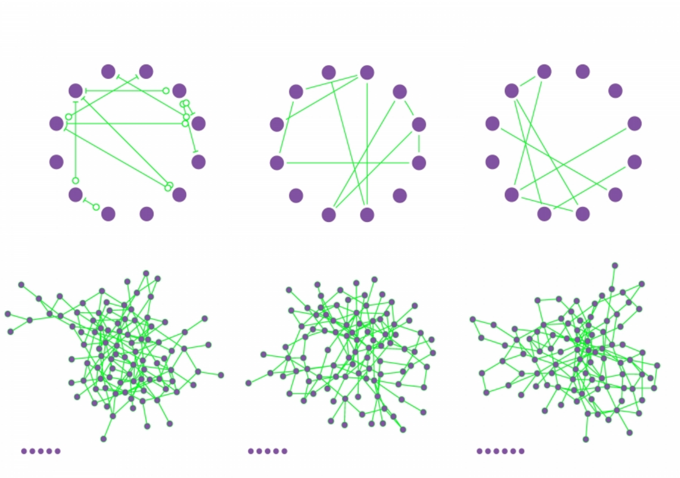
network features:

- Power-law (heavy-tailed) degree distribution
- Small average distance (graph diameter)
- Large clustering coefficient (transitivity)
- Giant connected component

Generative models:

- Random graph model (Erdos & Renyi, 1959)
- Preferential attachement model (Barabasi & Albert, 1999)
- Small world model (Watts & Strogatz, 1998)

Random graph model



top: $n = 12, p = 1/6$ bottom: $n = 100, p = 0.03$

Random graph model

Graph $G\{E, V\}$, nodes $n = |V|$, edges $m = |E|$
Erdos and Renyi, 1959.

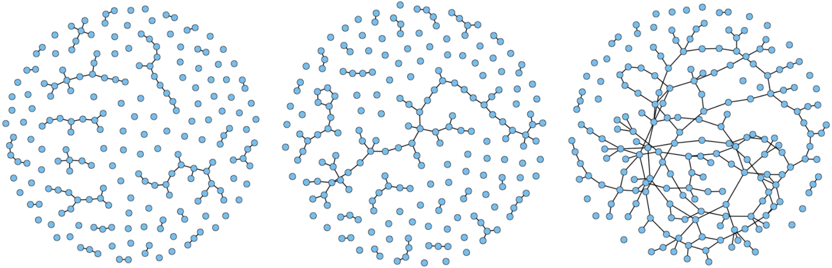
- $G_{n,p}$ - each pair out of $N = \frac{n(n-1)}{2}$ is connected with probability p,
  number of edges m - random number

$$\langle m \rangle = p\frac{n(n-1)}{2}$$

$$\langle k \rangle = \frac{1}{n}\sum_i k_i = \frac{2\langle m \rangle}{n} = p\,(n-1) \approx pn$$

$$\rho = \frac{\langle m \rangle}{n(n-1)/2} = p$$

Random graph model

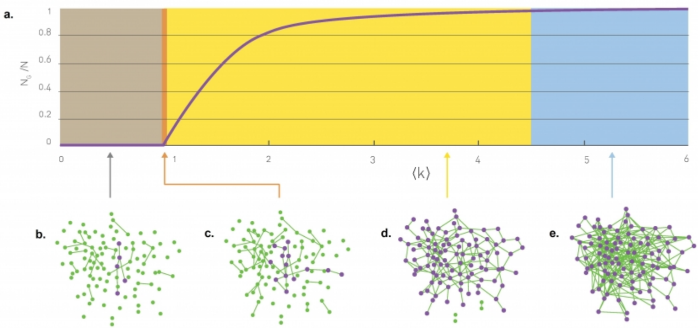


$p < p_c$ $\quad$ $p = p_c$ $\quad$ $p > p_c$

Structural changes happens when increasing p

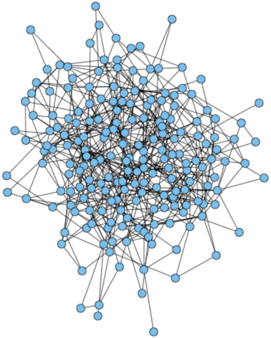
Phase transition

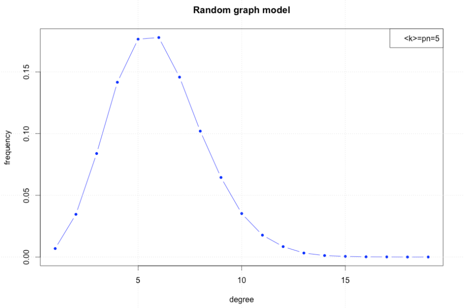
The size of largest connected component



Critical value: $\langle k \rangle = p_c n = 1$- on average one neighbor for a node

Random graph



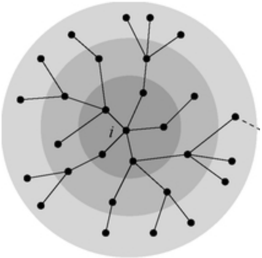


Node degree distribution (Poisson distribution):

$$P(k_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \lambda = pn = \langle k \rangle$$

Graph diameter

- $G(n,p)$ is locally tree-like (GCC) (no loops; low clustering coefficient)



- on average, the number of nodes d steps away from a node $\langle k \rangle^d$
- in GCC, around p_c , $\langle k \rangle^d \sim n$,

$$d \sim \frac{\ln n}{\ln \langle k \rangle}$$

Clustering coefficient

- Clustering coefficient

$$C(k) = \frac{\text{\#of links between NN}}{\text{\#max number of links NN}} = \frac{pk(k-1)/2}{k(k-1)/2} = p$$

$$C = p = \frac{\langle k \rangle}{n}$$

- when $n \to \infty, \; C \to 0$

Random graph model

- Node degree distribution function - Poisson:

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \;\; \lambda = pn = \langle k \rangle$$

- Average path length:

$$\langle L \rangle \sim \log(N) / \log\langle k \rangle$$

- Clustering coefficient:

$$C = \frac{\langle k \rangle}{n}$$

Preferential attachment model

Barabasi and Albert, 1999

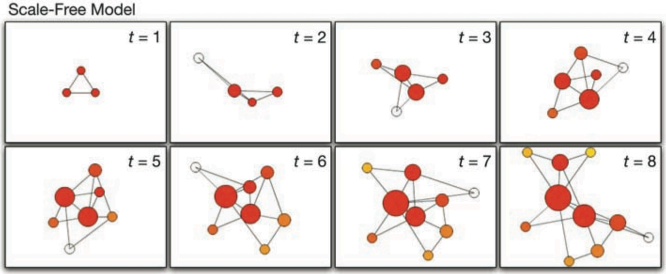
Dynamical growth model

- $t = 0$, n_0 nodes
- growth: on every step add a node with m_0 edges ($m_0 \leq n_0$), $k_i(i) = m_0$
- Preferential attachment: probability of linking to existing node is proportional to the node degree k_i

$$\Pi(k_i) = \frac{k_i}{\sum_i k_i} = \frac{k_i}{2m_0 t}$$

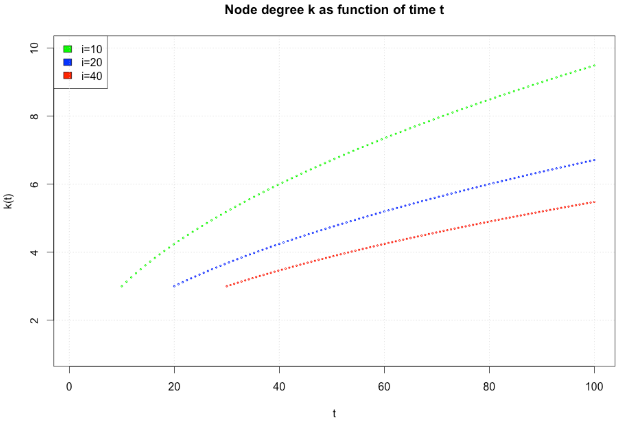
after t steps: $n_0 + t$ nodes, $m_0 t$ edges

Preferential attachment model



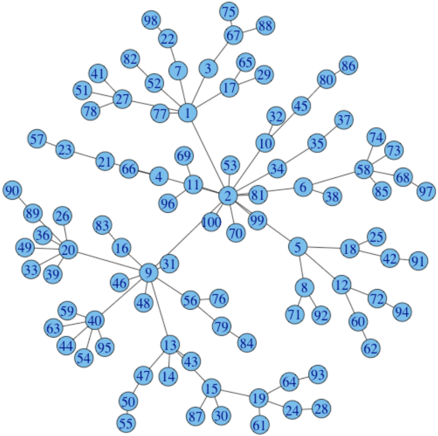
Barabasi, 1999

Preferential attachement

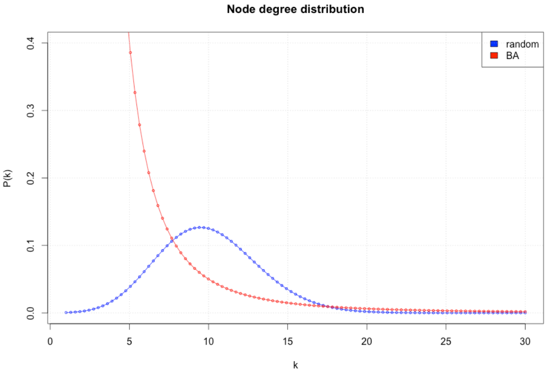


$$k_i(t) = m_0 \left(\frac{t}{i}\right)^{1/2}$$

Preferential attachement



Preferential attachement

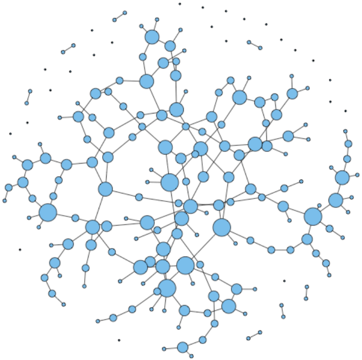


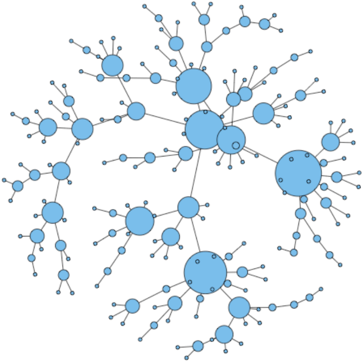
Node degree distribution:

$$P(k_i = k) = \frac{2m_0^2}{k^3}$$

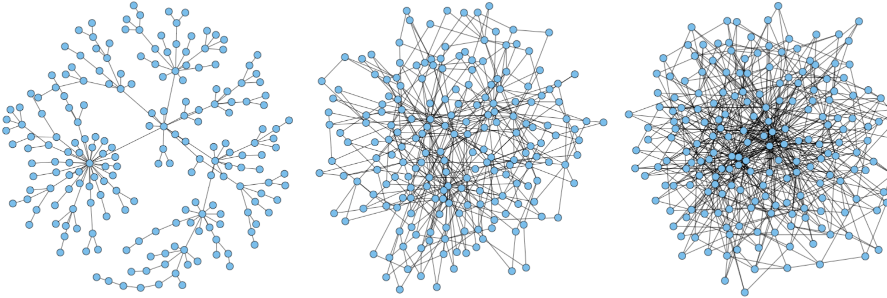
Preferential attachement

Preferential attachment vs random graph





Preferential attachment model



$m_0 = 1$ $\quad$ $m_0 = 2$ $\quad$ $m_0 = 3$

Preferential attachment model

- Node degree distribution - power law):

$$P(k) = \frac{2m_0^2}{k^3}$$

- Average path length :

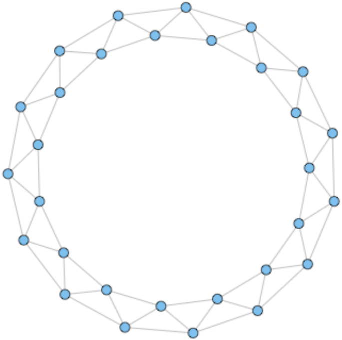
$$\langle L \rangle \sim \log(N)/\log(\log(N))$$

- Clustering coefficient (numerical result):

$$C \sim N^{-0.75}$$

Small world

Motivation: keep high clustering, get small diameter



Clustering coefficient $C = 1/2$
Graph diameter $d = 8$

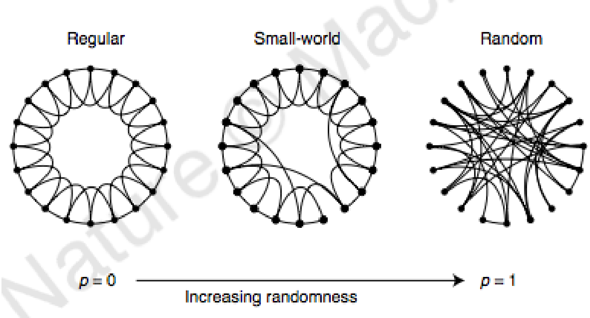
Small world

Watts and Strogatz, 1998

Single parameter model, interpolation between regular lattice and random graph

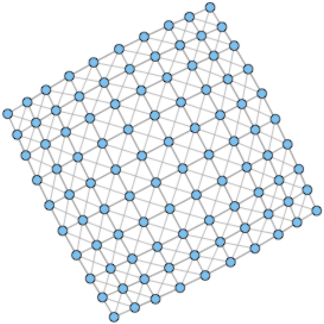
- start with regular lattice with n nodes, k edges per vertex (node degree), $k << n$
- randomly connect with other nodes with probability p, forms $pnk/2$ "long distance" connections from total of $nk/2$ edges
- $p = 0$ regular lattice, $p = 1$ random graph

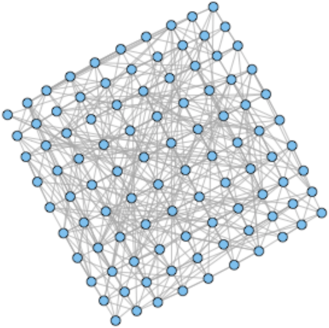
Small world



Watts, 1998

Small world model





20% rewiring:

ave. path length = 3.58 $\rightarrow$ ave. path length = 2.32

clust. coeff = 0.49 $\rightarrow$ clust. coeff = 0.19

igraph:watts.strogatz.game()

Small world model

- Node degree distribution function - Poisson like (numerical result)
- Average path length (analytical result) :

$$\langle L \rangle \sim \log(N)$$

- Clustering coefficient

$$C = const$$

Model comparison

	Random	BA model	WS model	Empirical networks
P(k)	$\frac{\lambda^k e^{-\lambda}}{k!}$	k^{-3}	poisson like	power law
C	$\langle k \rangle / N$	$N^{-0.75}$	const	large
$\langle L \rangle$	$\frac{\log(N)}{\log(\langle k \rangle)}$	$\frac{\log(N)}{\log \log(N)}$	$\log(N)$	small

NATIONAL RESEARCH
UNIVERSITY

Node Centrality and Ranking on Networks

Presentation from:http://www.leonidzhukov.net/hse/2019/sna/

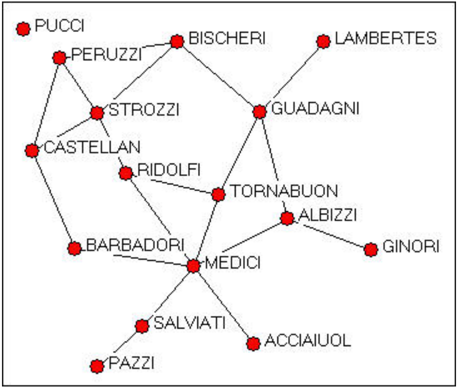
Centrality

Which vertices are important?

image from M.Grandjean, 2014

Centrality Measures

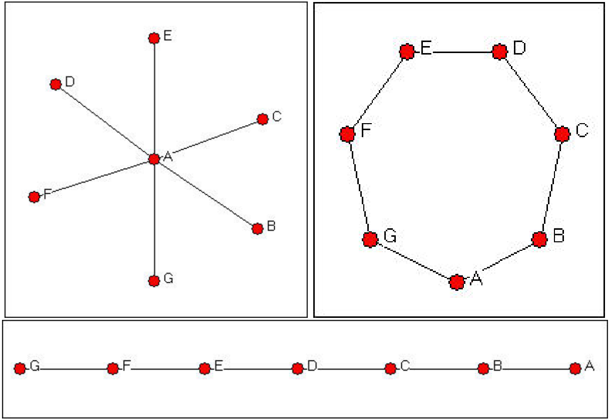
Determine the most ”important” or ”prominent” actors in the network based on actor location.



Marriage alliances among leading Florentine families 15th century.

Padgett, 1993

Three graphs



Star graph Circle graph Line Graph

Degree centrality

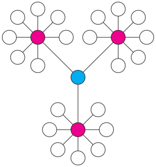
Degree centrality: number of nearest neighbors

$$C_D(i) = k(i) = \sum_j A_{ij} = \sum_j A_{ji}$$

Normalized degree centrality

$$C_D^*(i) = \frac{1}{n-1} C_D(i) = \frac{k(i)}{n-1}$$

High centrality degree -direct contact with many other actors



Closeness centrality

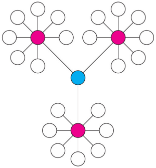
Closeness centrality: how close an actor to all the other actors in network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

Normalized closeness centrality

$$C_C^*(i) = (n-1)C_C(i) = \frac{n-1}{\sum_j d(i,j)}$$

High closeness centrality - short communication path to others, minimal number of steps to reach others



Betweenness centrality

Betweenness centrality: number of shortest paths going through the actor $\sigma_{st}(i)$

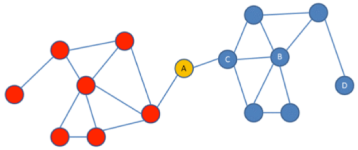
$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Normalized betweenness centrality

$$C_B^*(i) = \frac{2}{(n-1)(n-2)} C_B(i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Hight betweenness centrality - vertex lies on many shortest paths

Probability that a communication from s to t will go through i



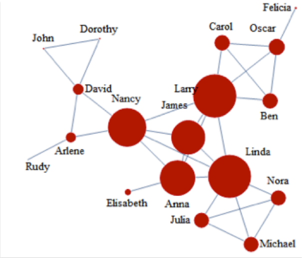
Eigenvector centrality

Importance of a node depends on the importance of its neighbors (recursive definition)

$$v_i \leftarrow \sum_j A_{ij} v_j$$

$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$

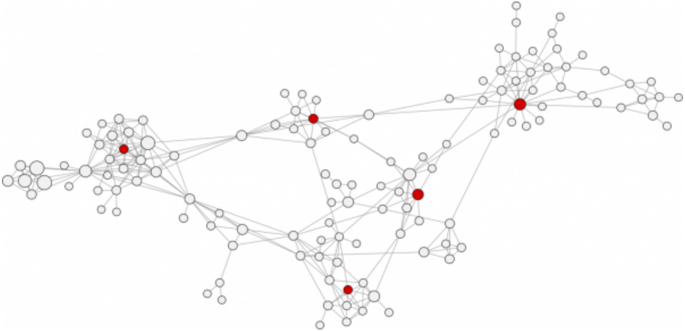
$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$



Select an eigenvector associated with largest eigenvalue $\lambda = \lambda_1$, $\mathbf{v} = \mathbf{v}_1$

Centrality examples

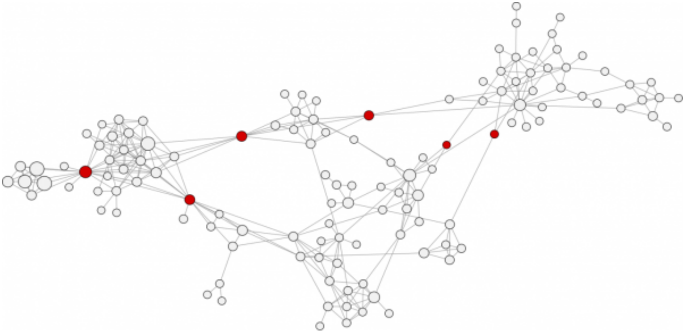
Closeness centrality



from www.activenetworks.net

Centrality examples

Betweenness centrality



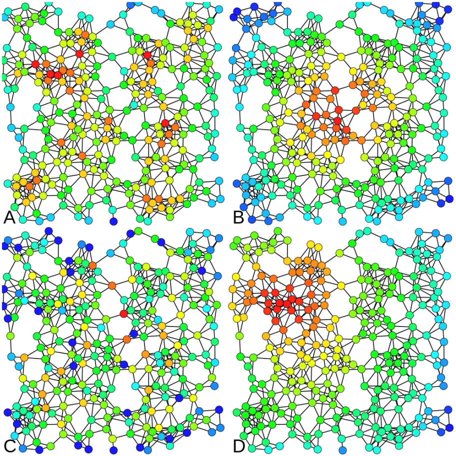
from www.activenetworks.net

Centrality examples

Eigenvector centrality

from www.activenetworks.net

Centrality examples



from Claudio Rocchini

- A) Degree centrality
- B) Closeness centrality
- C) Betweenness centrality
- D) Eigenvector centrality

Centralization

Centralization (network measure) - how central the most central node in the network in relation to all other nodes.

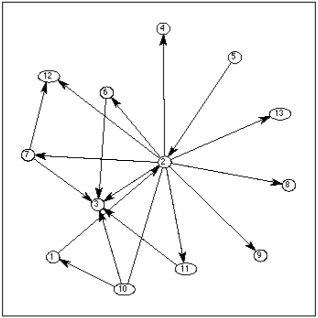
$$C_x = \frac{\sum_i^N [C_x(p_*) - C_x(p_i)]}{\max \sum_i^N [C_x(p_*) - C_x(p_i)]}$$

C_x - one of the centrality measures
p_* - node with the largest centrality value
max - is taken over all graphs with the same number of nodes (for degree, closeness and betweenness the most centralized structure is the star graph)

Linton Freeman, 1979

Directional relations

Directed graph: distinguish between choices made (outgoing edges) and choices received (incoming edges)



sending - receiving
export - import
cite papers - being cited

Centrality measures

All based on outgoing edges

- Degree centrality (normalized):

$$C_D^*(i) = \frac{k^{out}(i)}{n-1}$$

- Closeness centrality (normalized):

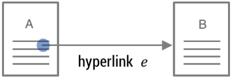
$$C_C^*(i) = \frac{n-1}{\sum_j d(i,j)}$$

- **Betweenness centrality (normalized):

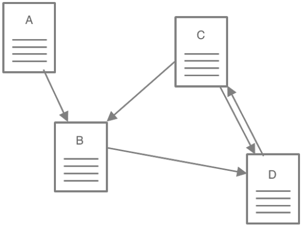
$$C_B^*(i) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Web as a graph

- Hyperlinks - implicit endorsements

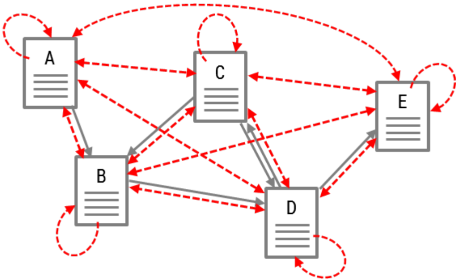


- Web graph - graph of endorsements (sometimes reciprocal)



PageRank

"PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The **probability** that the random surfer visits a page is its **PageRank**."



Sergey Brin and Larry Page, 1998

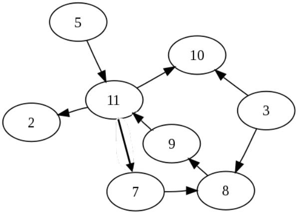
Random walk

- Random walk on graph

$$p_i^{t+1} = \sum_{j \in N(i)} \frac{p_j^t}{d_j^{out}} = \sum_j \frac{A_{ji}}{d_j^{out}} p_j$$

$$\mathbf{p}^{t+1} = \mathbf{P}^T \mathbf{p}^t$$

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}, \ \ \mathbf{D}_{ii} = diag\{d_i^{out}\}$$



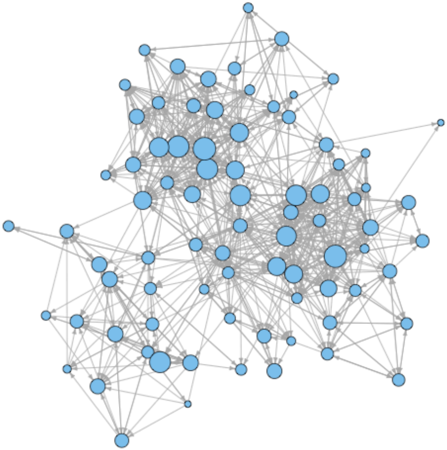
- with teleportation

$$\mathbf{p}^{t+1} = \alpha \mathbf{P}^T \mathbf{p}^t + (1-\alpha)\mathbf{v}$$

Perron-Frobenius Theorem guarantees existence and uniqueness of the solution $\lim_{t\to\infty} \mathbf{p} = \pi$

$$\pi = \alpha \mathbf{P}^T \pi + (1-\alpha)\mathbf{v}$$

PageRank



PageRank beyond the Web

1. GeneRank
2. ProteinRank
3. FoodRank
4. SportsRank
5. HostRank
6. TrustRank
7. BadRank
8. ObjectRank
9. ItemRank
10. ArticleRank
11. BookRank
12. FutureRank
13. TimedPageRank
14. SocialPageRank
15. DiffusionRank
16. ImpressionRank
17. TweetRank
18. TwitterRank
19. ReversePageRank
20. PageTrust
21. PopRank
22. CiteRank
23. FactRank
24. InvestorRank
25. ImageRank
26. VisualRank
27. QueryRank
28. BookmarkRank
29. StoryRank
30. PerturbationRank
31. ChemicalRank
32. RoadRank
33. PaperRank
34. Etc…

Hubs and Authorities (HITS)

Citation networks. Reviews vs original research (authoritative) papers

- authorities, contain useful information, a_i
- hubs, contains links to authorities, h_i

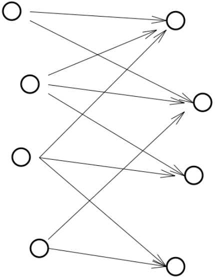
Mutual recursion

- Good authorities referred by good hubs

$$a_i \leftarrow \sum_j A_{ji} h_j$$

- Good hubs point to good authorities

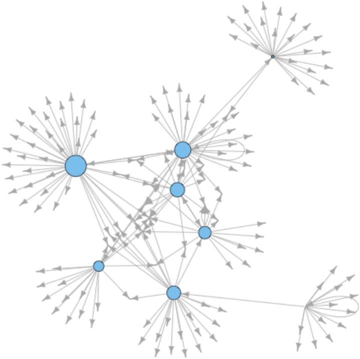
$$h_i \leftarrow \sum A_{ij} a_j$$

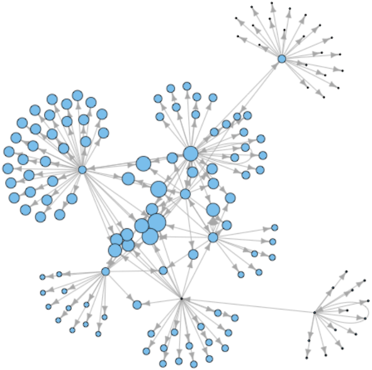


Hubs and Authorities

Hubs

Authorities





Florentines families