Network Science

Part II: Weighted Networks

Albert-László Barabási

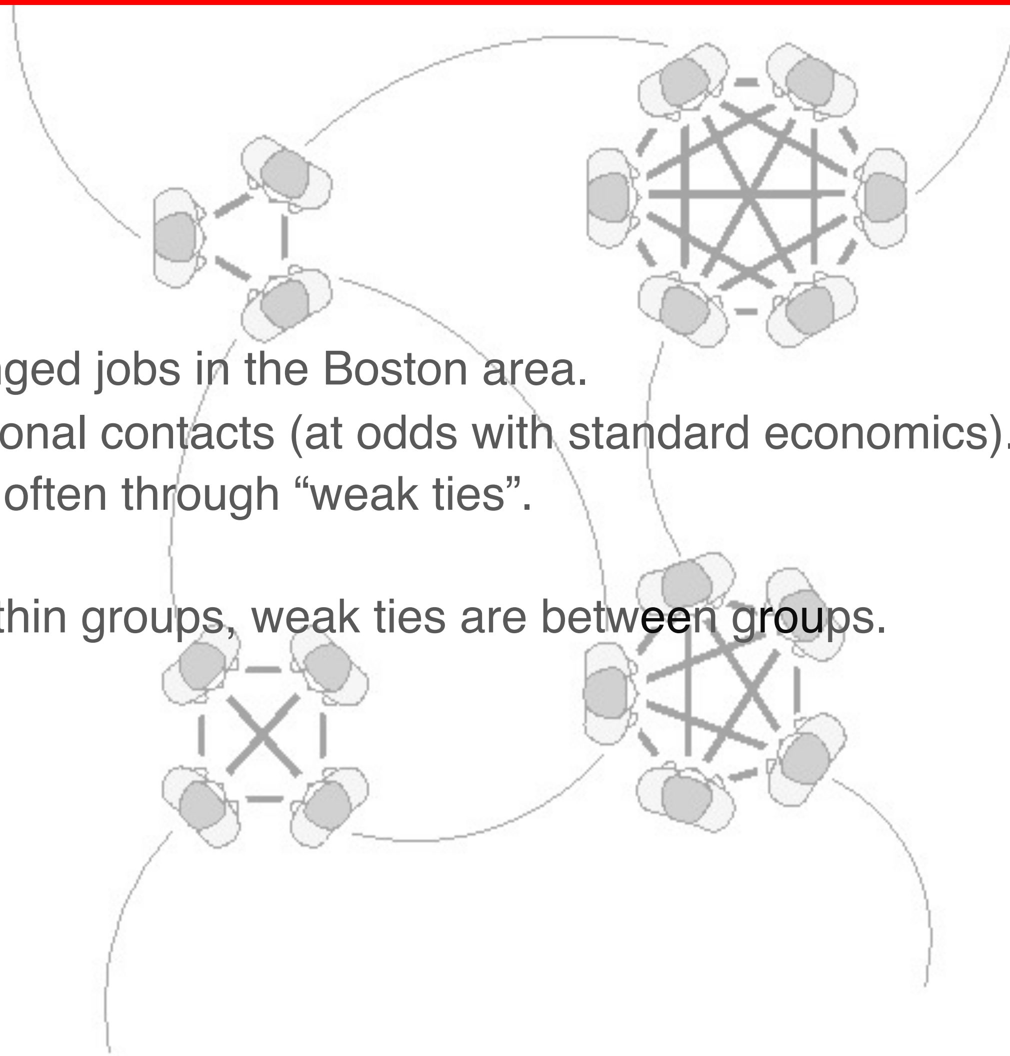
With

Emma K. Towlson, Sebastian Ruf, Michael Danziger, and Louis Shekhtman

www.BarabasiLab.com

WHY WEIGHTS MATTER: THE STRENGTH OF WEAK TIES

How do people find a new job?



- interviewed 100 people who had changed jobs in the Boston area.
- More than half found job through personal contacts (at odds with standard economics).
- Those who found a job, found it more often through "weak ties".

- HYPOTHESIS: The strong ties are within groups, weak ties are between groups.

Granovetter, 1973

OUTLINE

Examples of weighted networks

Extended metrics to weighted networks: Weighted degree and other definitions (C, knn).

Correlations between weights and degrees

Scaling of weights

EXAMPLE 1: AIRLINE TRAFFIC

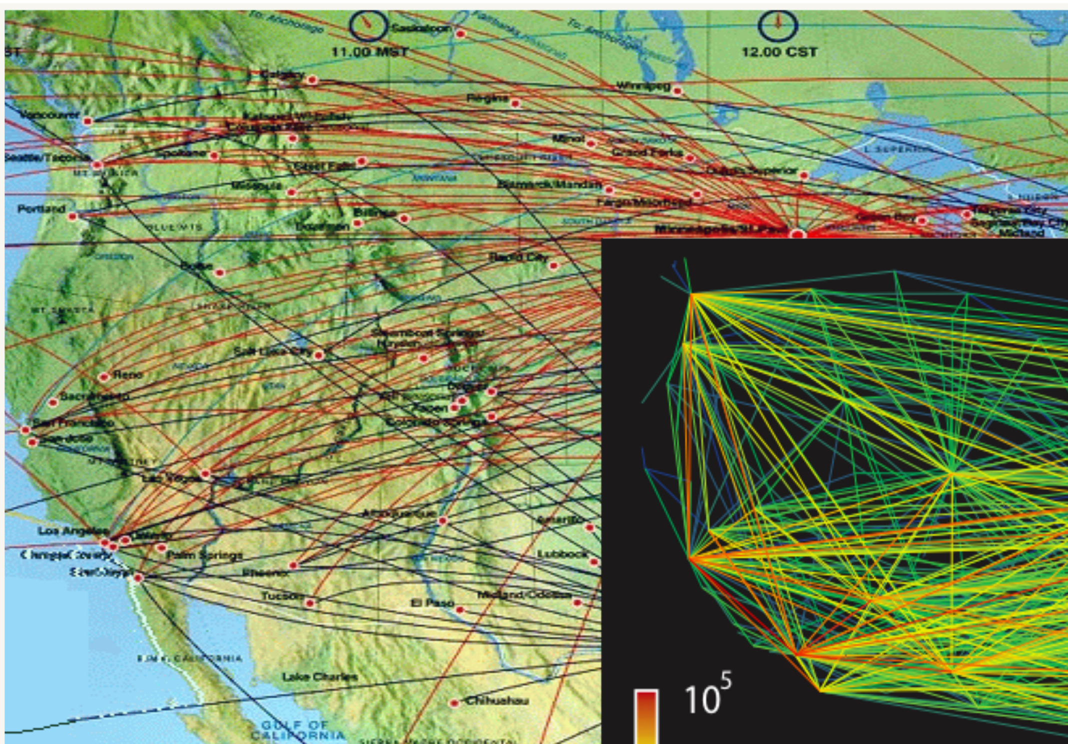
Nodes: airports

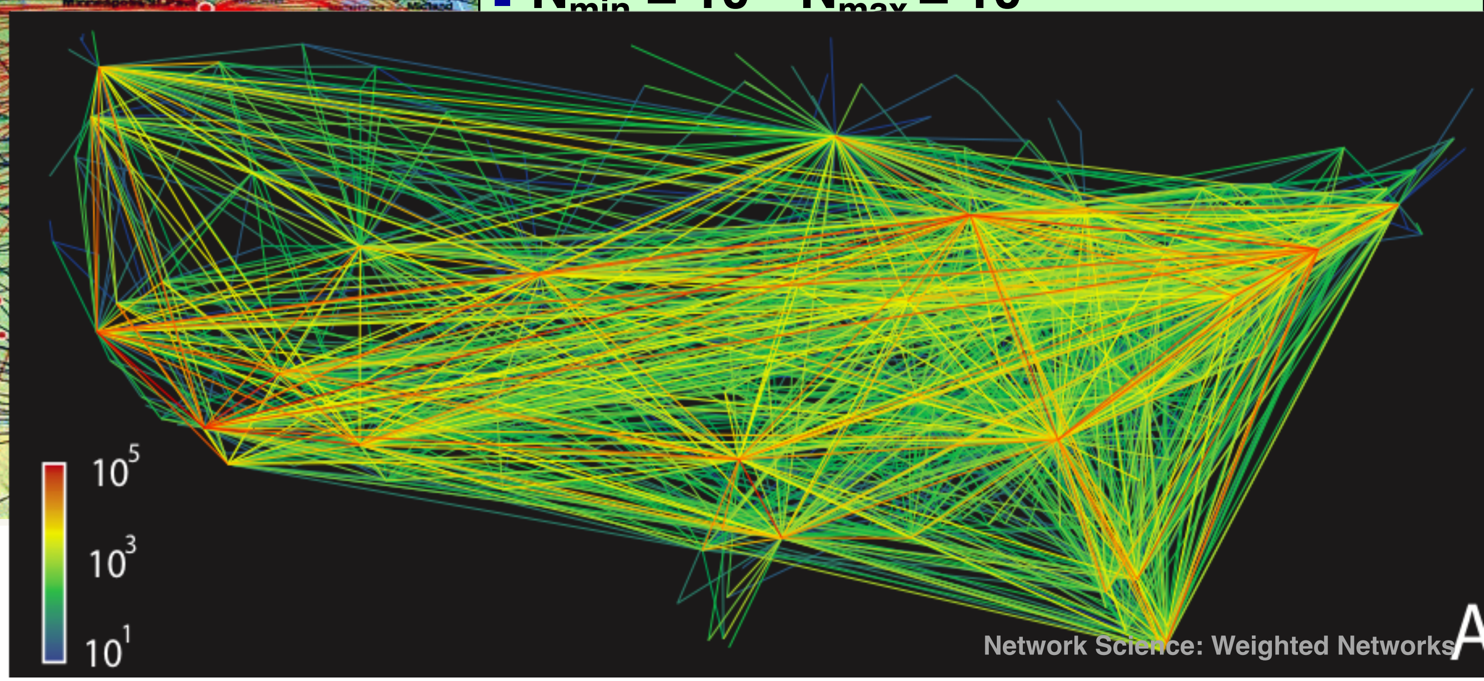
Links: direct flights

Weights: Number of seats

- ➢ **2002 IATA database**
- **V = 3880 airports**
- **E = 18810 direct flights**

- $\langle k \rangle = 10$ $k_{max} = 318$
- $\langle l \rangle = 4$ $l_{max} = 16$
- $\langle w \rangle = 10^5$ $w_{max} = 10^7$
- $N_{min} = 10^3$ $N_{max} = 10^7$





EXAMPLE 2: SCIENCE COLLABORATION NETWORK

- Nodes: scientists
- Links: joint publications
- Weights: number of joint pubs.

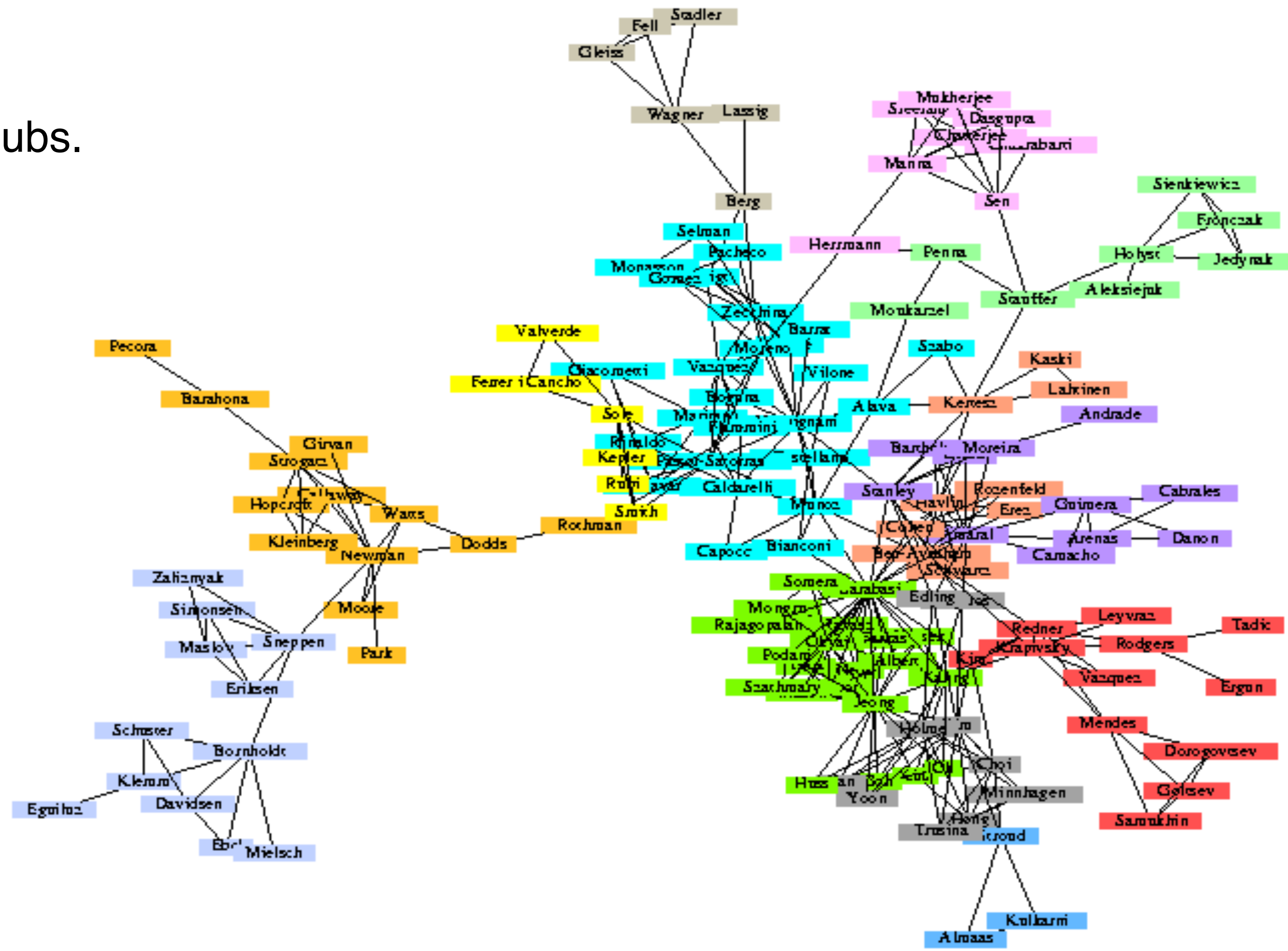
$$\mathbf{w_{ij}} = \Sigma_{\mathbf{k}} \frac{\delta_{\mathbf{i}}^{\mathbf{k}} \delta_{\mathbf{j}}^{\mathbf{k}}}{\mathbf{n_k} - 1}$$

i, j: authors

k: paper

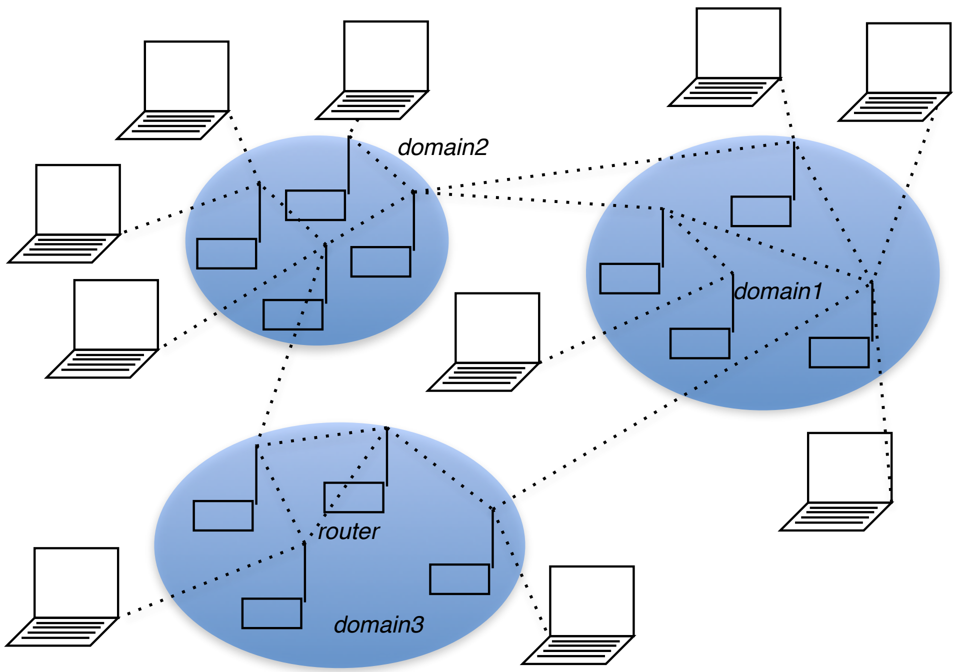
n_k: number of authors

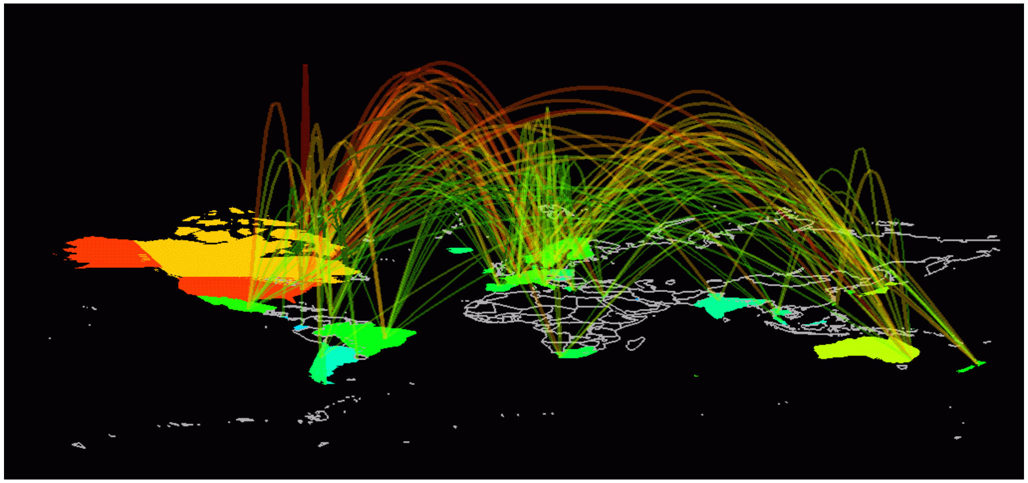
$\delta_i^k=1$ if author i contributed to paper k

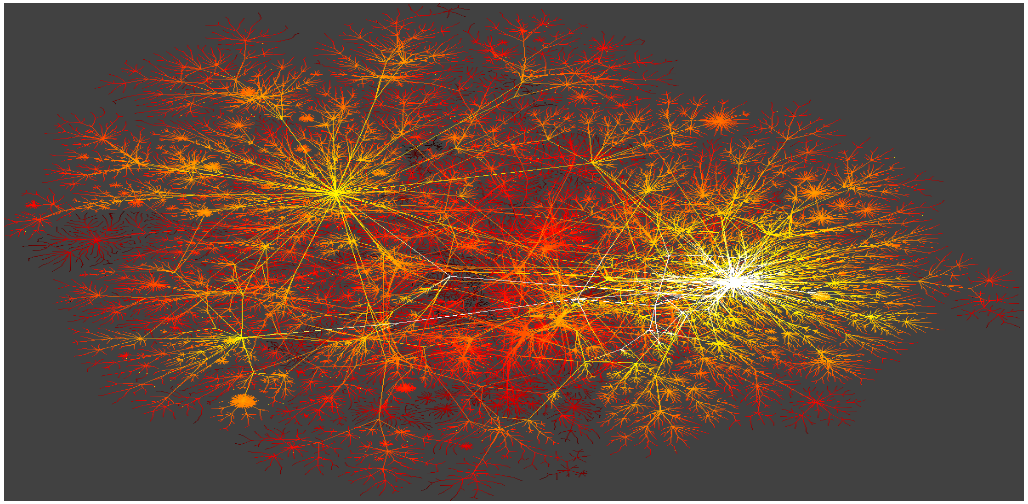


Newman and Girvan, cond-mat/0308217 (2003)

EXAMPLE 3: INTERNET

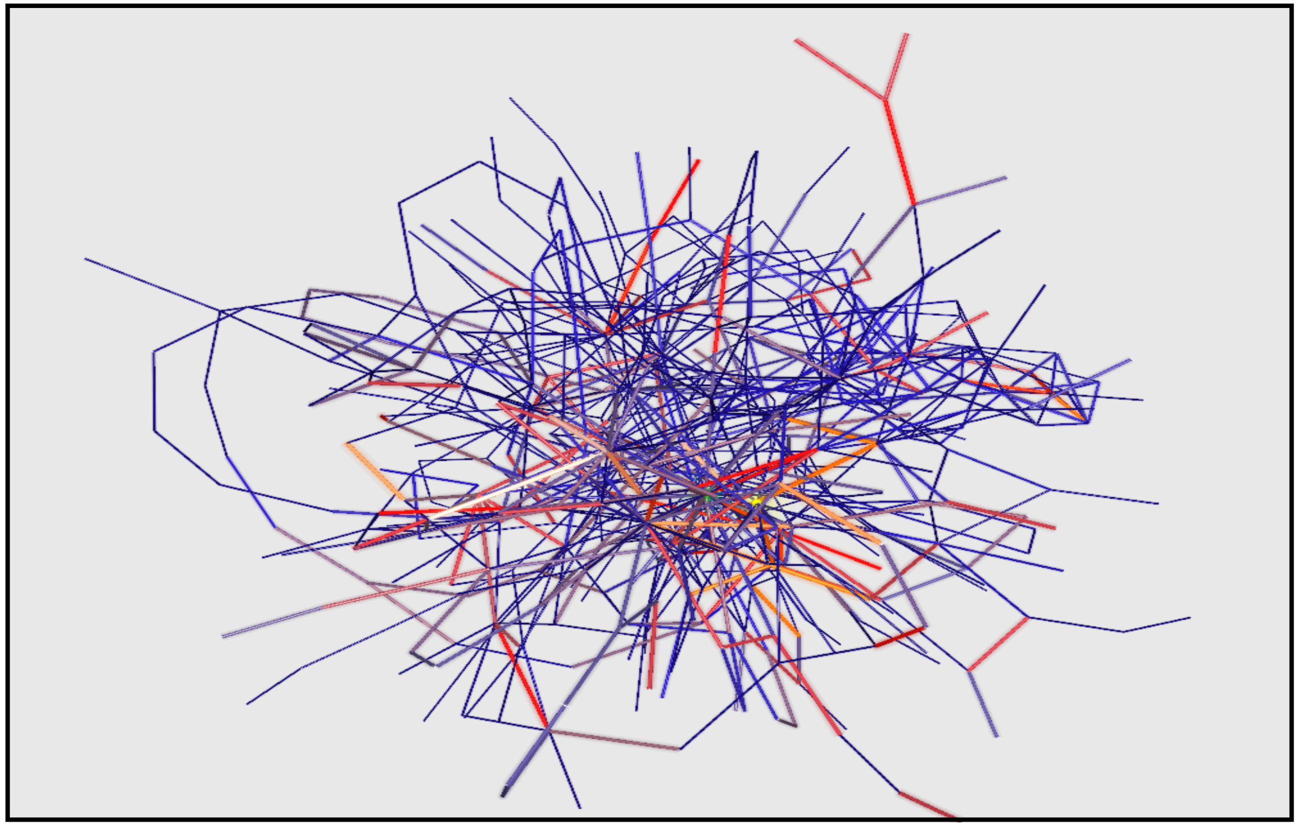






- Nodes: routers
- Links: physical lines
- Link Weights: bandwidth or traffic

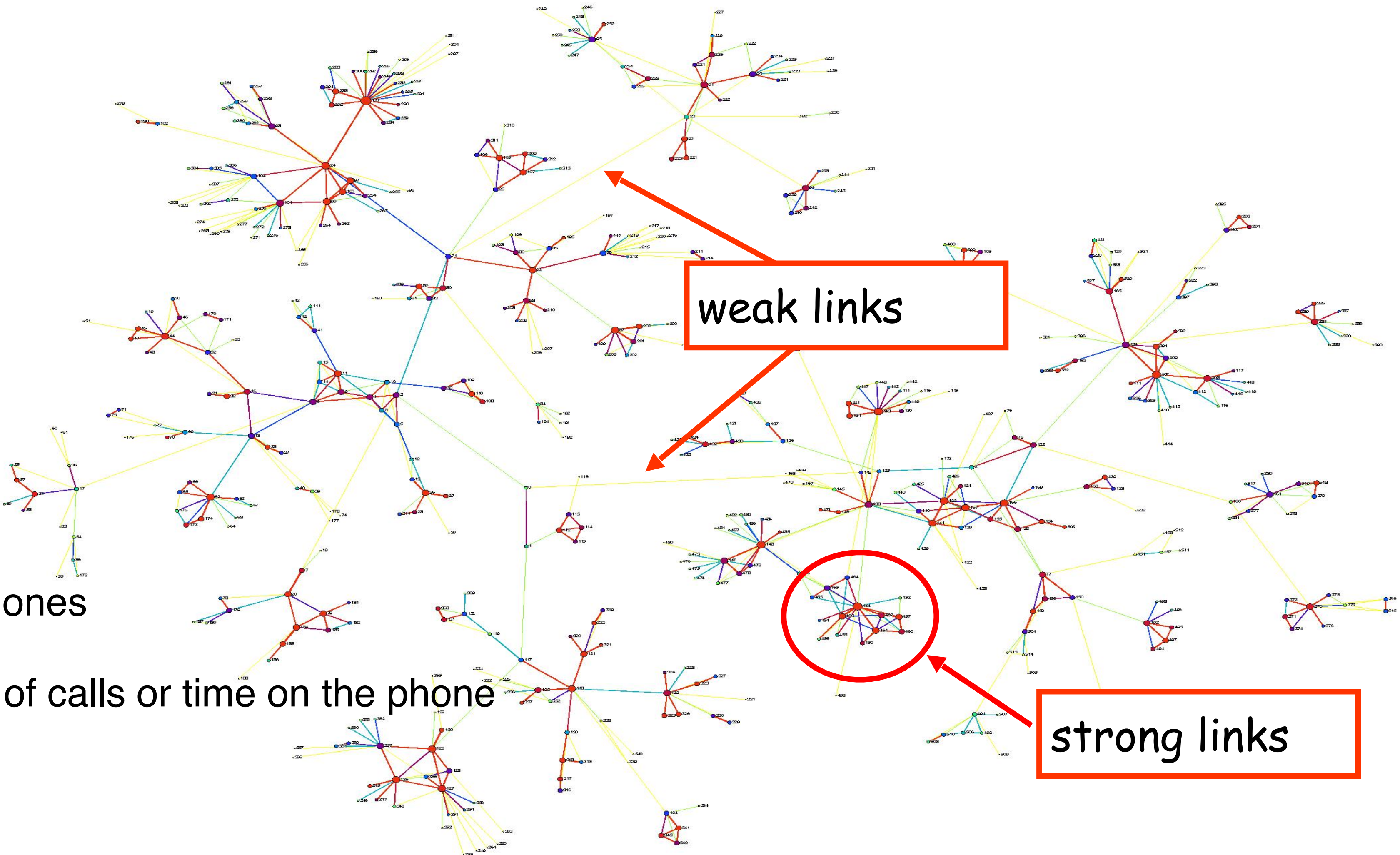
EXAMPLE 4: METABOLIC NETWORK



- Nodes: metabolites
- Links: reactions
- Link Weights: reaction flux

E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A.-L. B. Nature, 2004; Goh et al, PRL 2002.

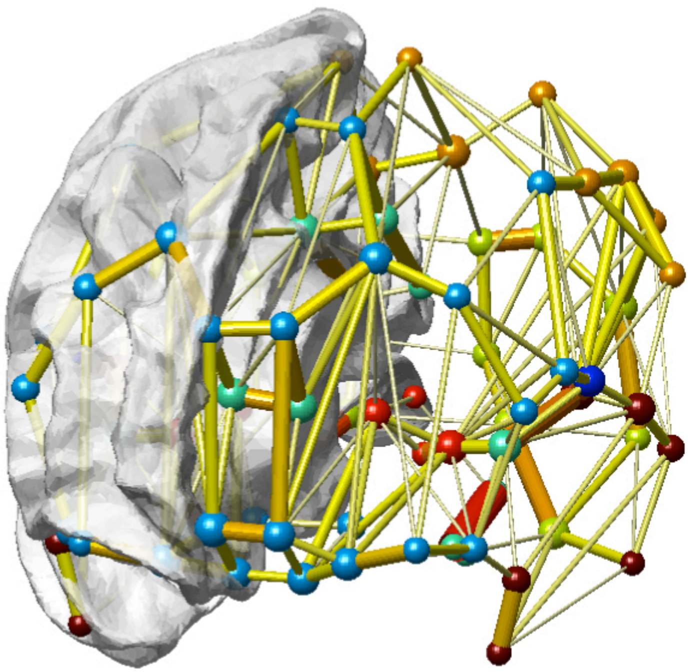
EXAMPLE 5: MOBILE CALL GRAPH

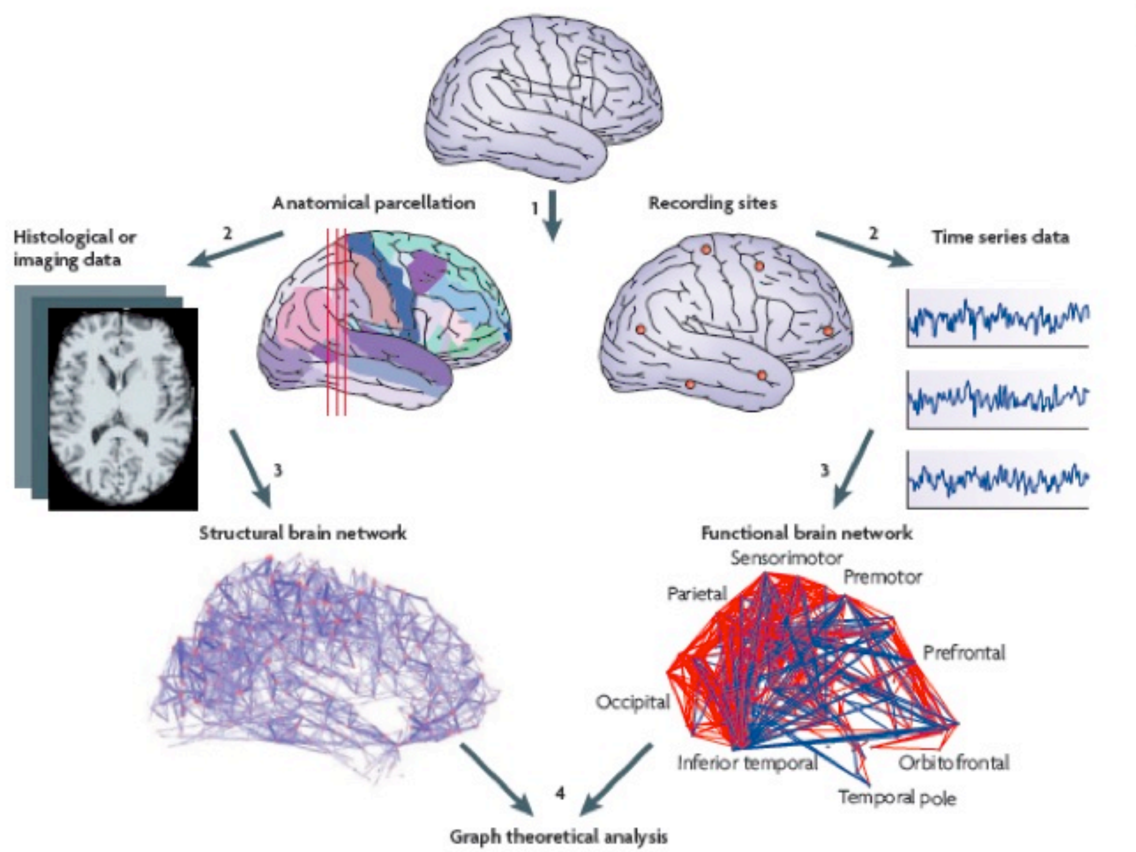


- Nodes: mobile phones
- Links: calls
- Weights: number of calls or time on the phone

Onella et al, PNAS 2007

EXAMPLE 6: fMRI NETWORKS





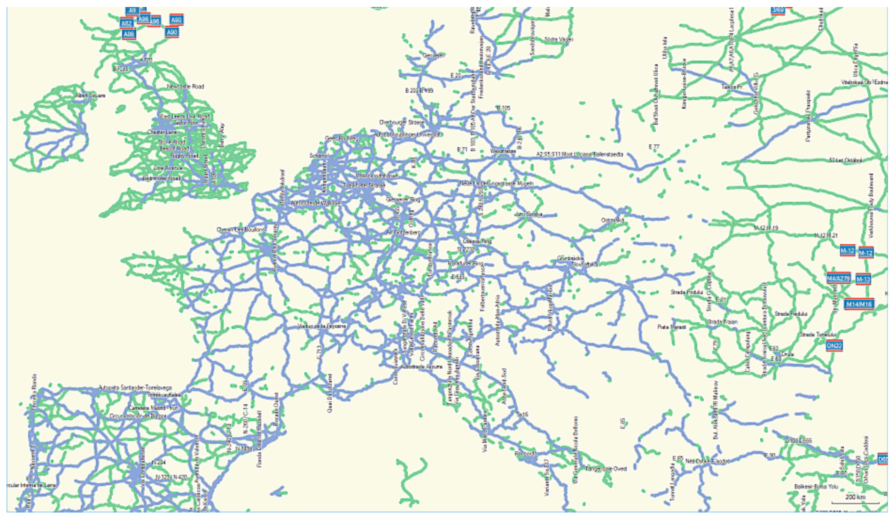
- Nodes: macroscopic brain regions
- Links: correlation of BOLD timeseries
- Weights: strength of correlation

Bullmore & Sporns, Nat Rev Neuroscience 2009

THINK CAREFULLY: WHAT DO YOUR WEIGHTS MEAN?

The way to appropriately handle and interpret the extra information from edge weights is sensitive to what the weights are describing.

Eg path lengths (and any metric exploiting this concept!)



Higher weights can represent higher distances (or other ‘costs’). Eg road networks, spatial networks…
-> Want high strengths to correspond to long path lengths.

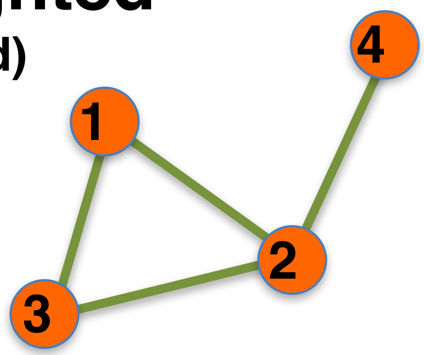
Higher weights can represent higher strength of connections/interactions. Eg fMRI, metabolic...
-> Want high strengths to correspond to short paths.

DEFINITIONS:

Weighted Degree

GRAPHOLOGY 2

Unweighted
(undirected)



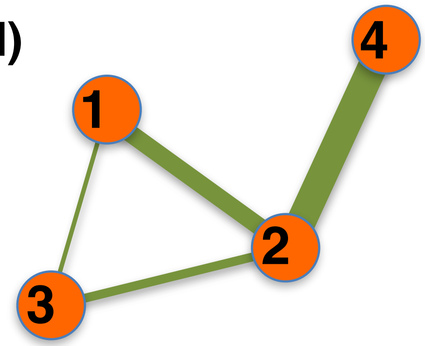
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad < k >= \frac{2L}{N}$$

Protein-protein interactions, www

Weighted
(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

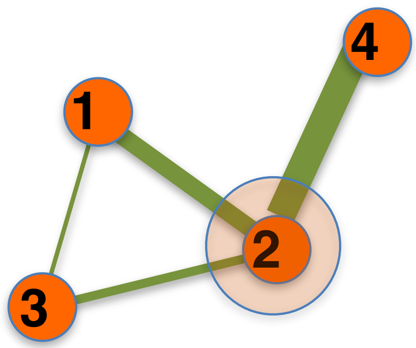
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad < k >= \frac{2L}{N}$$

Call Graph, metabolic networks

a_{ij} and w_{ij}

In the literature we often use a double notation: A_{ij} and w_{ij} (somewhat redundant)



Adjacency Matrix (A_{ij})

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Weight Matrix (W_{ij})

$$W_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

Node Strength (*weighted degree*) s:

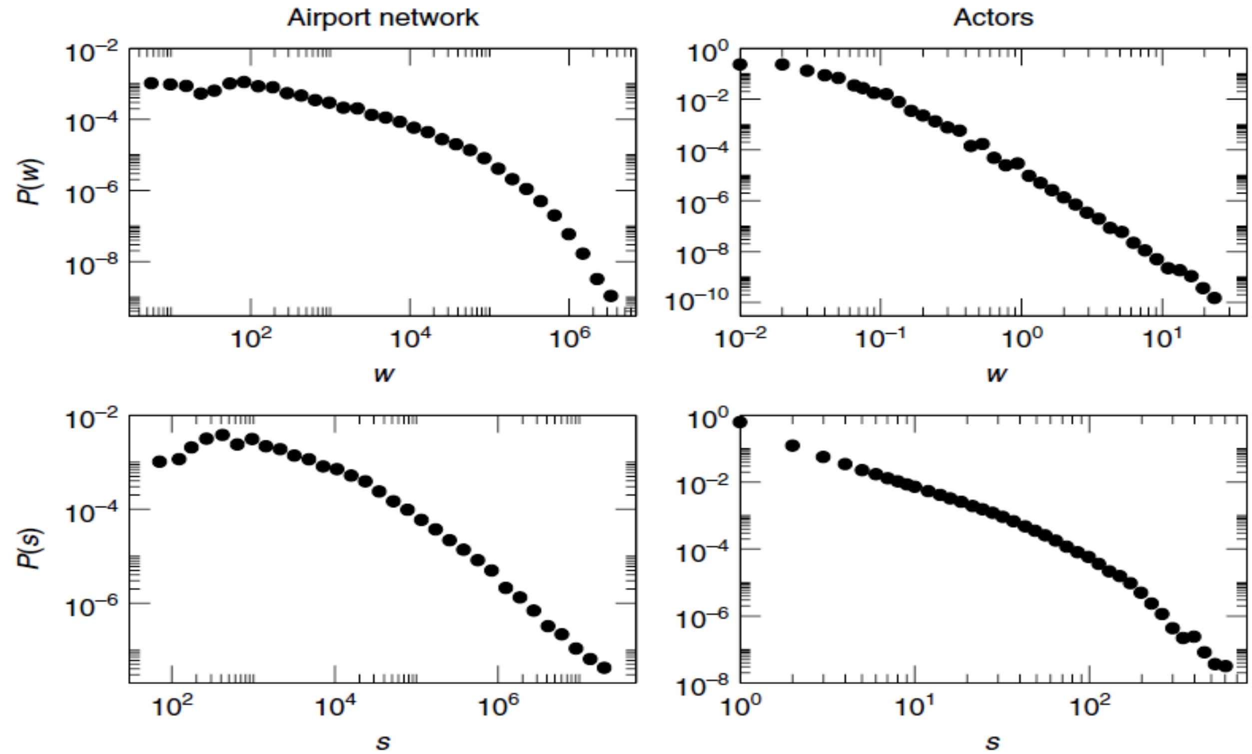
$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij} = \sum_{j=1}^{N} w_{ij}$$

$$s_2 = \sum_{j=1}^{N} w_{2j} = w_{21} + w_{23} + w_{24}$$

EMPIRICAL FINDING 1: P(s) AND P(w) ARE FAT TAILED

Strength distribution *P(s):*
probability that a randomly chosen node has strength *s*

Weight distribution ***P(w)***:
probability that a randomly chosen link has weight ***w***



In most real systems P(s) and P(w) are fat tailed.

EMPIRICAL 2: RELATIONSHIP BETWEEN STRENGTH AND DEGREE

If there are no correlations between k and s, we can approximate w_{ij} with $<w>$:

$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij} = <w> \sum_{j=1}^{N} a_{ij} = <w> k_i$$

Science collaboration network:

Weights appear to be assigned randomly

Barrat, Barthelemy, Pastor-Satorras, Vespignani, PNAS 2004

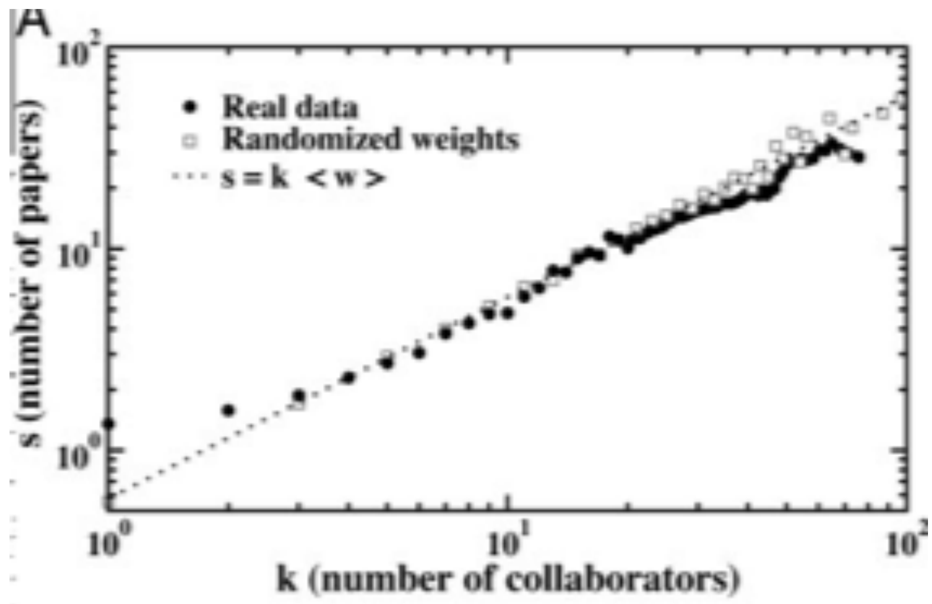
EMPIRICAL 2: RELATIONSHIP BETWEEN STRENGTH AND DEGREE

If there are no correlations between k and s, we can approximate w_{ij} with <w>:

$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij} = <w> \sum_{j=1}^{N} a_{ij} = <w> k_i$$

Science collaboration network:

Weights appear to be assigned randomly



Barrat, Barthelemy, Pastor-Satorras, Vespignani, PNAS 2004

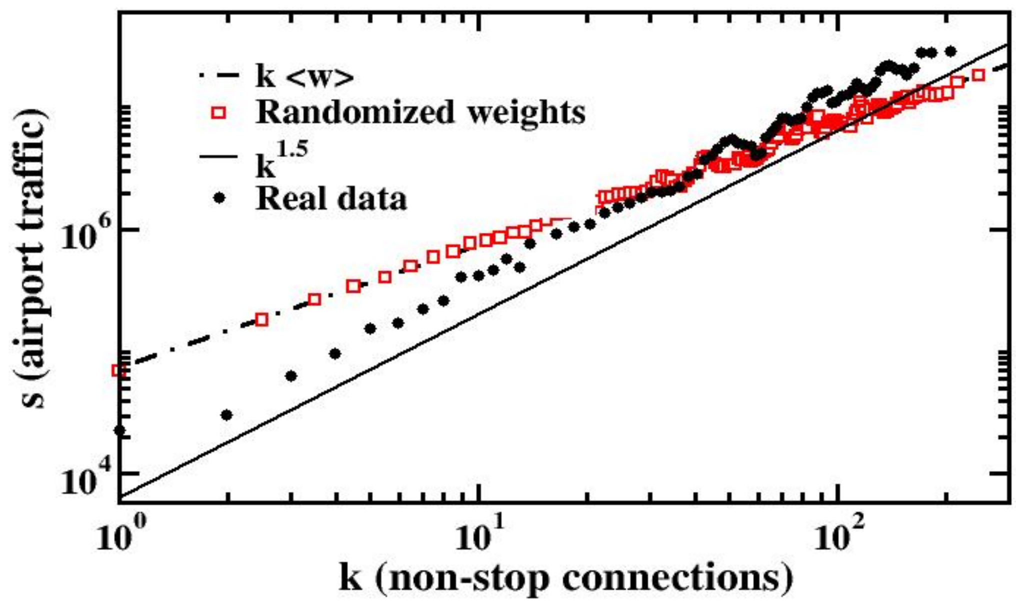
EMPIRICAL 2: NODE STRENGTH AND DEGREE

In most systems we observe:

$$s_i = A k_i^{\beta}$$

Airport network: β=1.5

Randomized weights:
s=<w>k: β=1



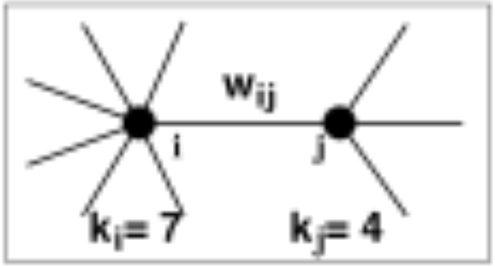
Correlations between topology and dynamics:

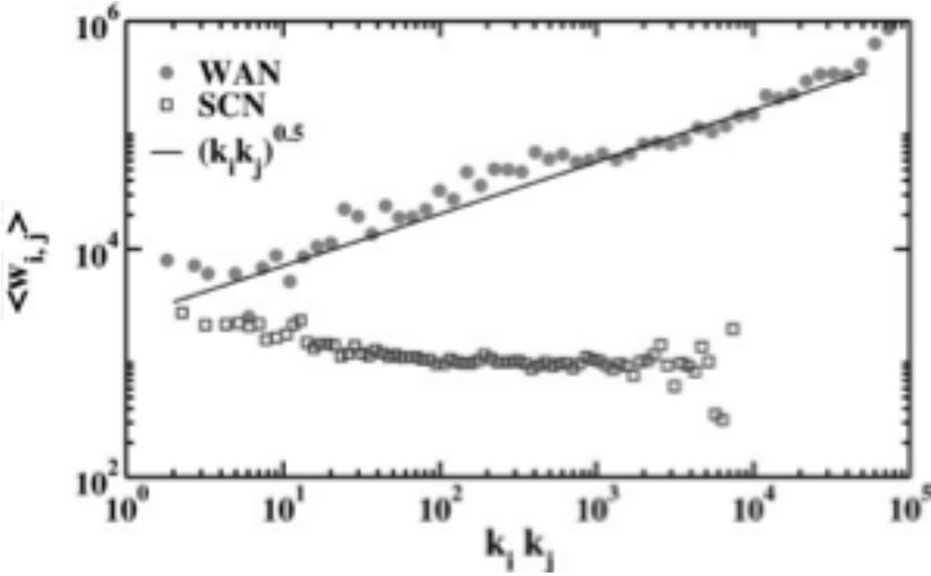
β>1: strength of vertices grows faster than their degree
àthe weight of edges belonging to highly connected vertices have a value higher than the one corresponding to a random assignment of weights.
à*the larger is an airport's degree (k), disproportionally more traffic it can handle (s).*

Barrat, Barthelemy, Pastor-Satorras, Vespignani, PNAS 2004

WEIGHT CORRELATIONS(2)

Weight correlation:





$$< w_{ij} >\sim \left(k_i k_j\right)^{\theta}$$

Scaling relationship:

$$s_i \sim k_i < w_{ij} >\sim k_i^{1+\theta} k_j^{\theta}$$

$$s_i = Ak_i^{\beta}$$

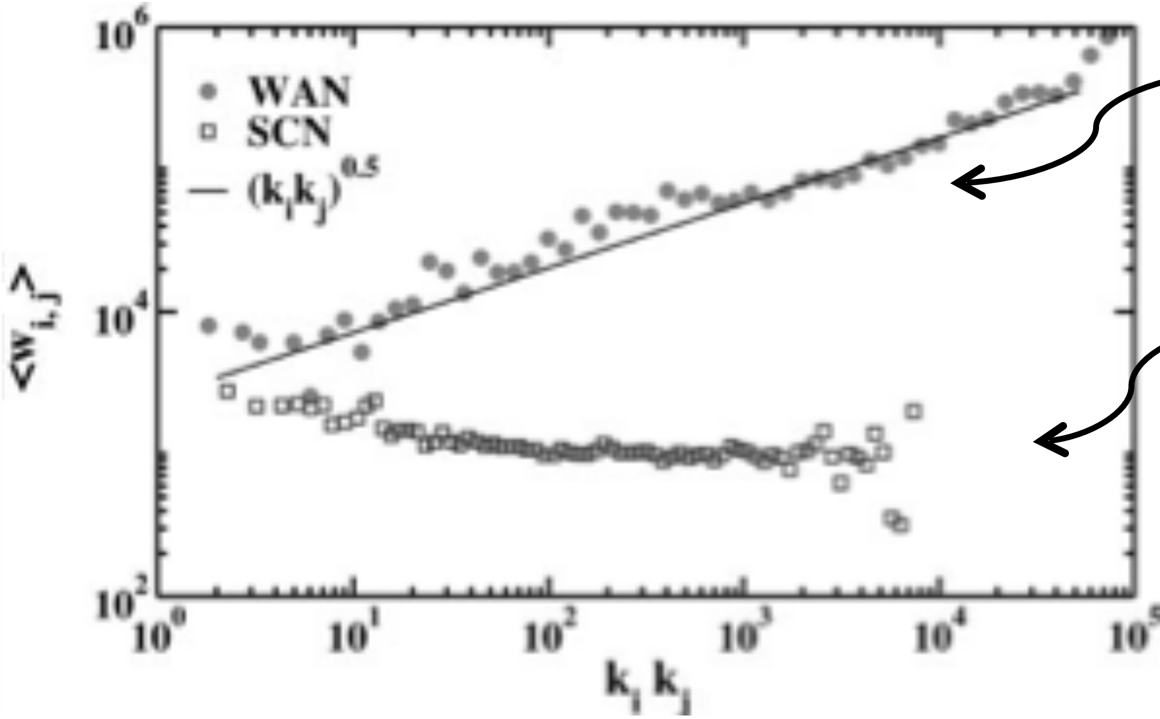
→ (assuming no degree correlations)

$$\beta = 1 + \theta$$

WEIGHT CORRELATIONS(2)

Weight correlation: $$< w_{ij} > \sim \left(k_i k_j\right)^{\theta}$$

Scaling relationship: $$\beta = 1 + \theta$$



Barrat, Barthelemy, Pastor-Satorras, Vespignani, PNAS 2004

Airline network:

β=1.5 → Θ=0.5 (YES!)

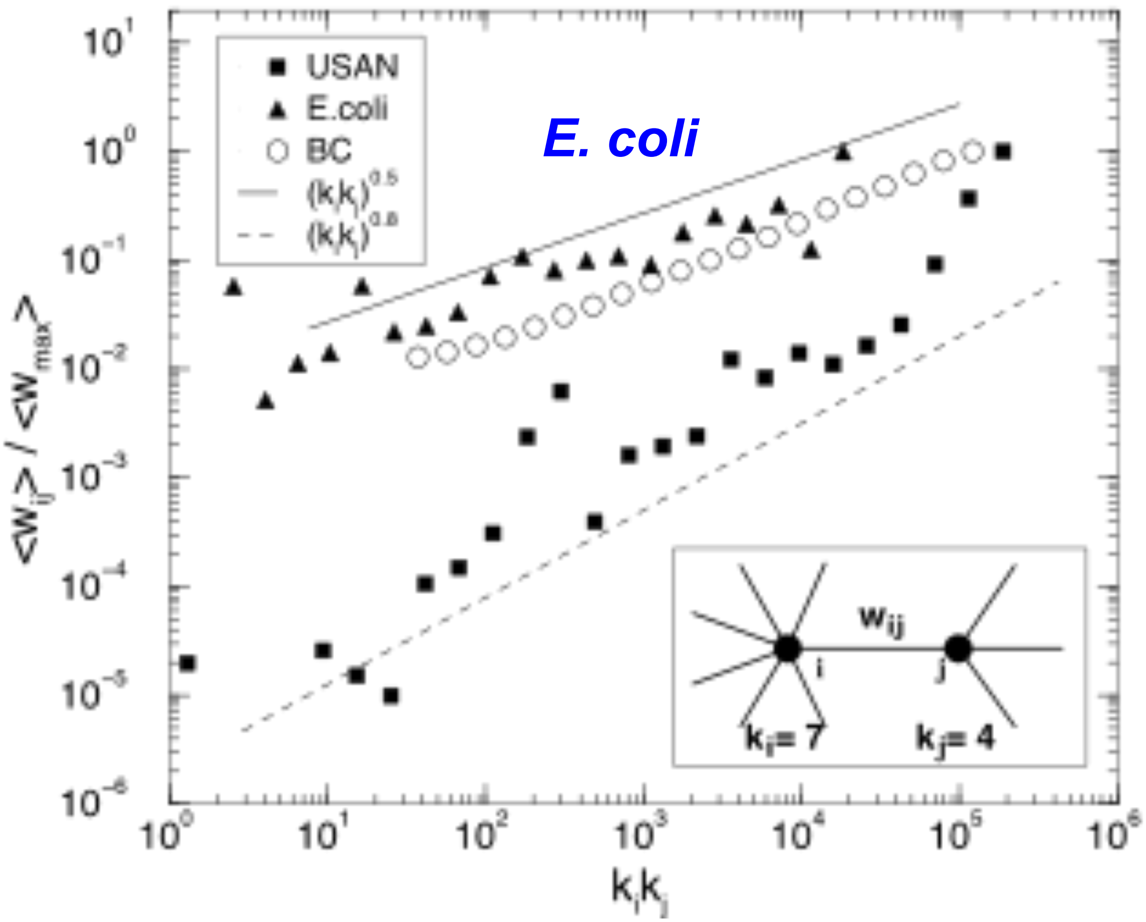
→ **The traffic between two airports depends on their individual degrees.**

Science collaboration:

β=1 → Θ=0 (YES!)
(confirming the lack of correlations between k and w)

→ **The weight between two authors does not depend on their degree.**

WEIGHT CORRELATIONS(2)



Weight correlation:

$$< w_{ij} > \sim \left(k_i k_j\right)^{\theta}$$

Scaling relationship:

$$\beta = 1 + \theta$$

Macdonald, Almaas, Barabasi, Europhys Lett 2005

SUMMARY

		Airline	*Collaboration*
Weight-degree:	$s_i = A k_i^{\beta}$	β=1.5	β=1
Weight correlation:	$< w_{ij} > \sim \left(k_i k_j\right)^{\theta}$	Θ=0.5	Θ=0

Scaling relationship:

$$\beta = 1 + \theta$$

Macdonald, Almaas, Barabasi, Europhys Lett 2005

WEIGHTED CORRELATIONS

(Clustering, Assortativity)

WEIGHTED CLUSTERING COEFFICIENT

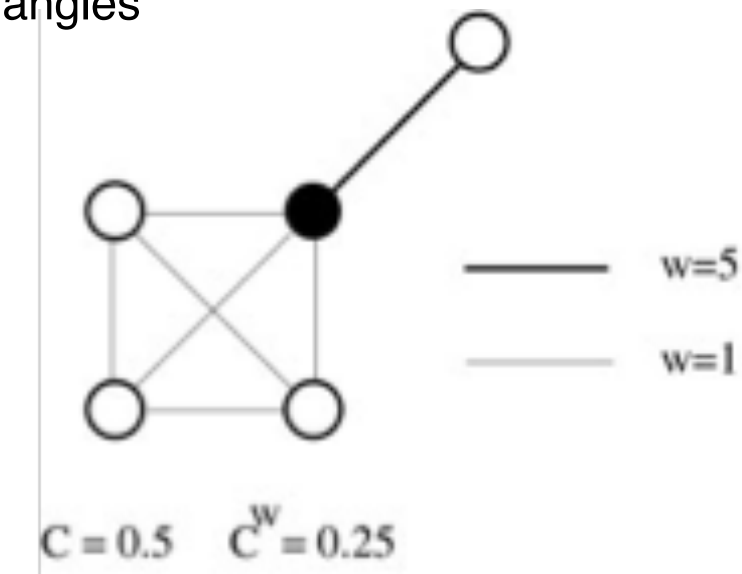
Un-weighted

$$c_i = \frac{1}{k_i(k_i-1)} \sum_{j,h} a_{ij} a_{ih} a_{jh}$$

$a_{ij}a_{ih}a_{jh}=1$ only for closed triangles

Weighted

$$c_i^w = \frac{1}{s_i(k_i-1)} \sum_{j,h} \frac{(w_{ij}+w_{ih})}{2} a_{ij} a_{ih} a_{jh}$$

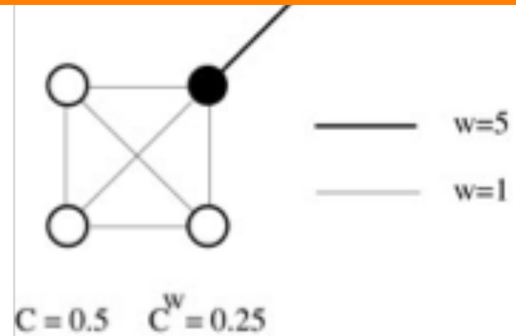


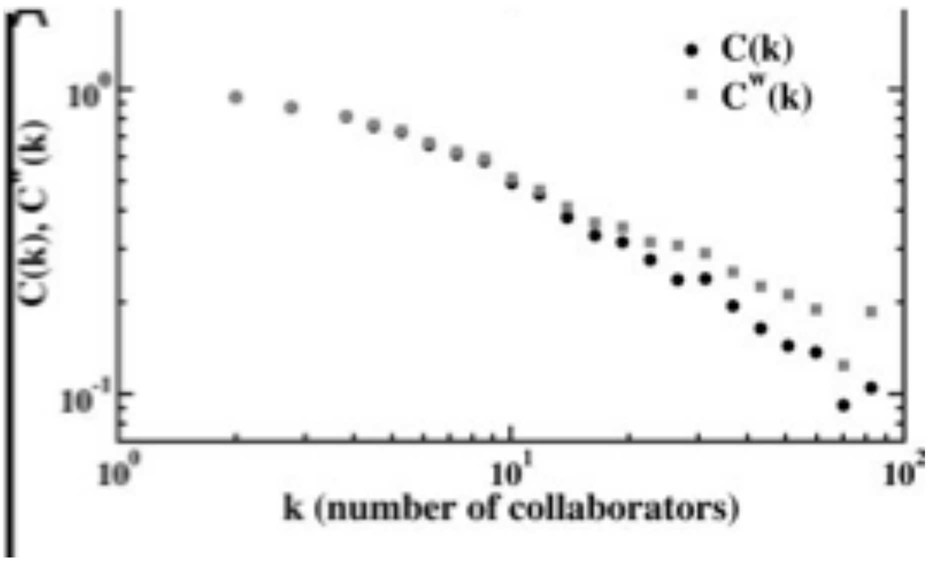
- If $w_{ij}=\langle w \rangle \rightarrow c_i=c_i^w$
- If $c_i^w/c_i>1$: Weights localized on cliques
- If $c_i^w/c_i<1$: Important links don't form cliques
- If w and k uncorrelated: $c_i^w=c_i$

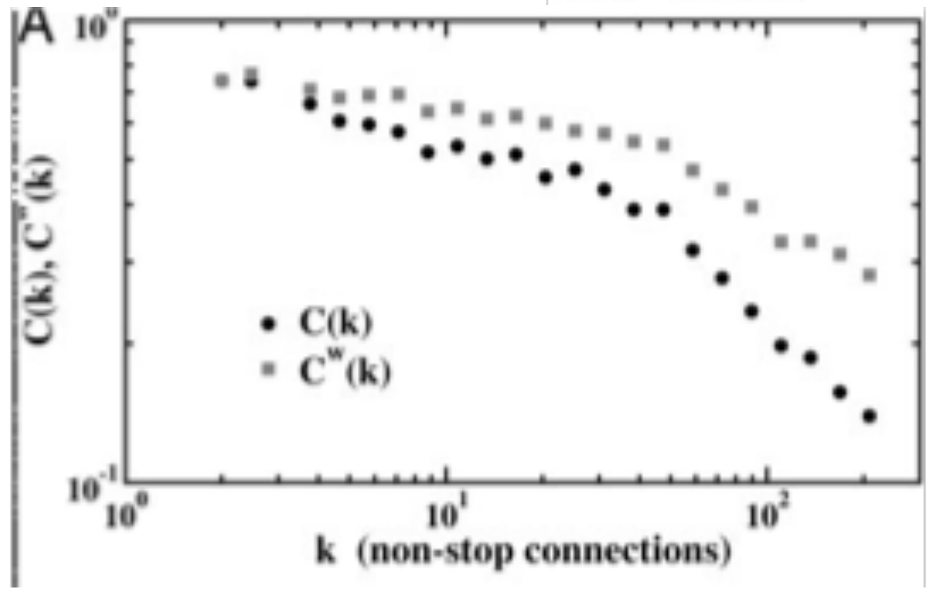
Barrat, Barthelemy, Pastor-Satorras, Vespignani, PNAS 2004

Weighted

$$c_i^w = \frac{1}{s_i(k_i-1)} \sum_{j,h} \frac{(w_{ij}+w_{ih})}{2} a_{ij} a_{ih} a_{jh}$$







Collaboration network:
C(k) and C^w(k) are comparable

Airline network:
C(k) < C^w(k): accumulation of traffic on high degree nodes

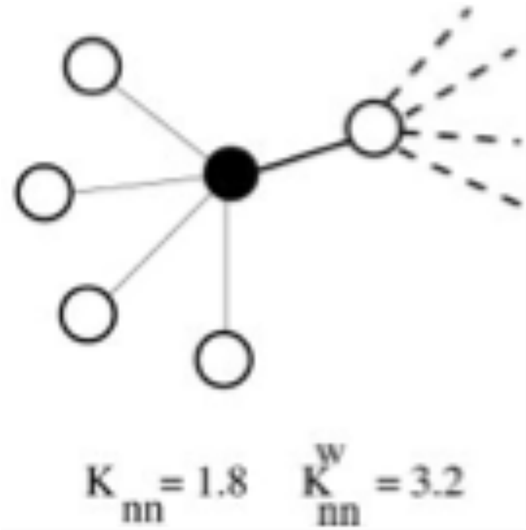
WEIGHTED ASSORTATIVITY

Un-weighted

$$k_{nn,i} = \frac{1}{k_i} \sum_{k'} a_{ij} k_j$$

Weighted

$$k^{w}_{nn,i} = \frac{1}{s_i} \sum_{k'} w_{ij} k_j$$



- If $w_{ij}=\langle w\rangle \rightarrow k_{nn,i}=k_{nn,i}^{w}$
- If $k_{nn}^{w}(i)/k_{nn}(i) > 1$: Edges with larger weights point to nodes with larger k

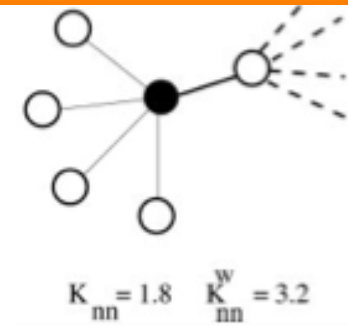
•Measures the affinity to connect with high- or low-degree neighbors according to the magnitude of the interactions.

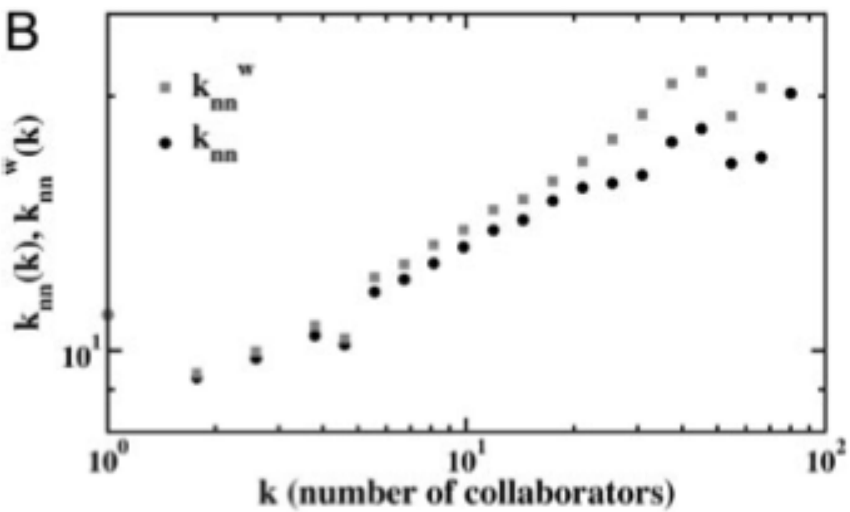
Barrat, Barthelemy, Pastor-Satorras, Vespignani, PNAS 2004

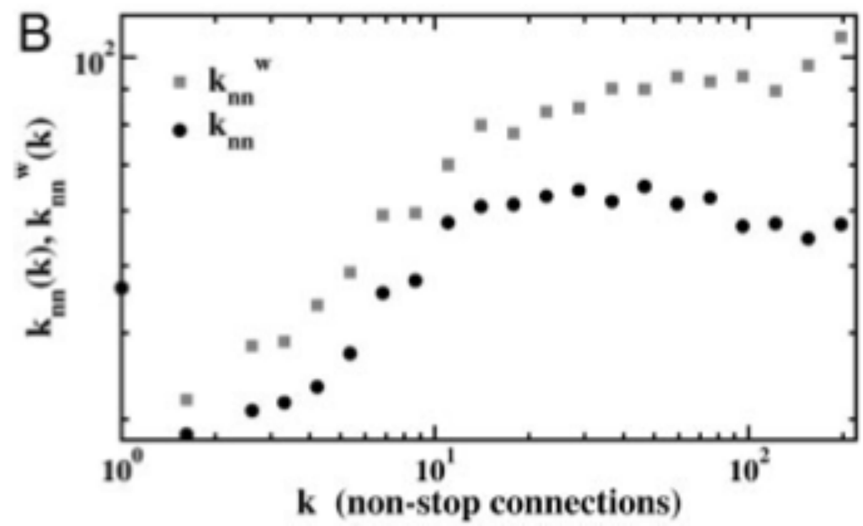
EXAMPLES: WEIGHTED ASSORTATIVITY

Weighted

$$k_{nn,i}^{w} = \frac{1}{s_i}\sum_{k'} w_{ij} k_j$$





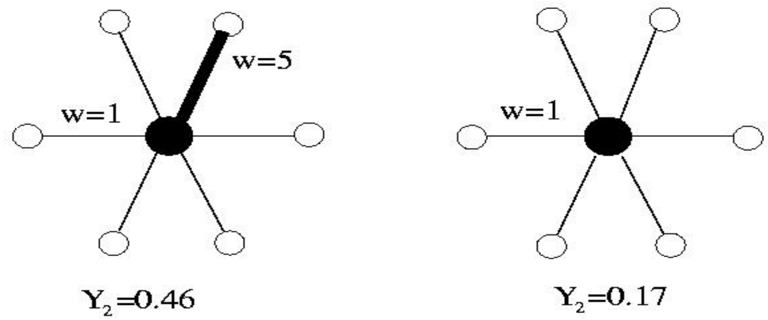


Assortative behavior in agreement with evidence that social networks usually have a strong assortative character

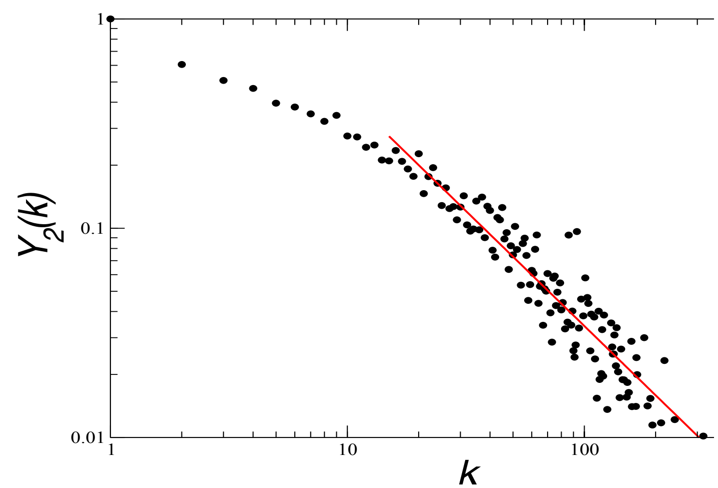
Assortative only for small degrees. For k>10, $k_{nn}(k)$ approaches a constant uncorrelated structure in which vertices with very different degrees have similar k_{nn}

DISPARITY

$$Y_2(i) = \sum_{j \in \mathcal{V}(i)} \left[\frac{w_{ij}}{s_i}\right]^2$$



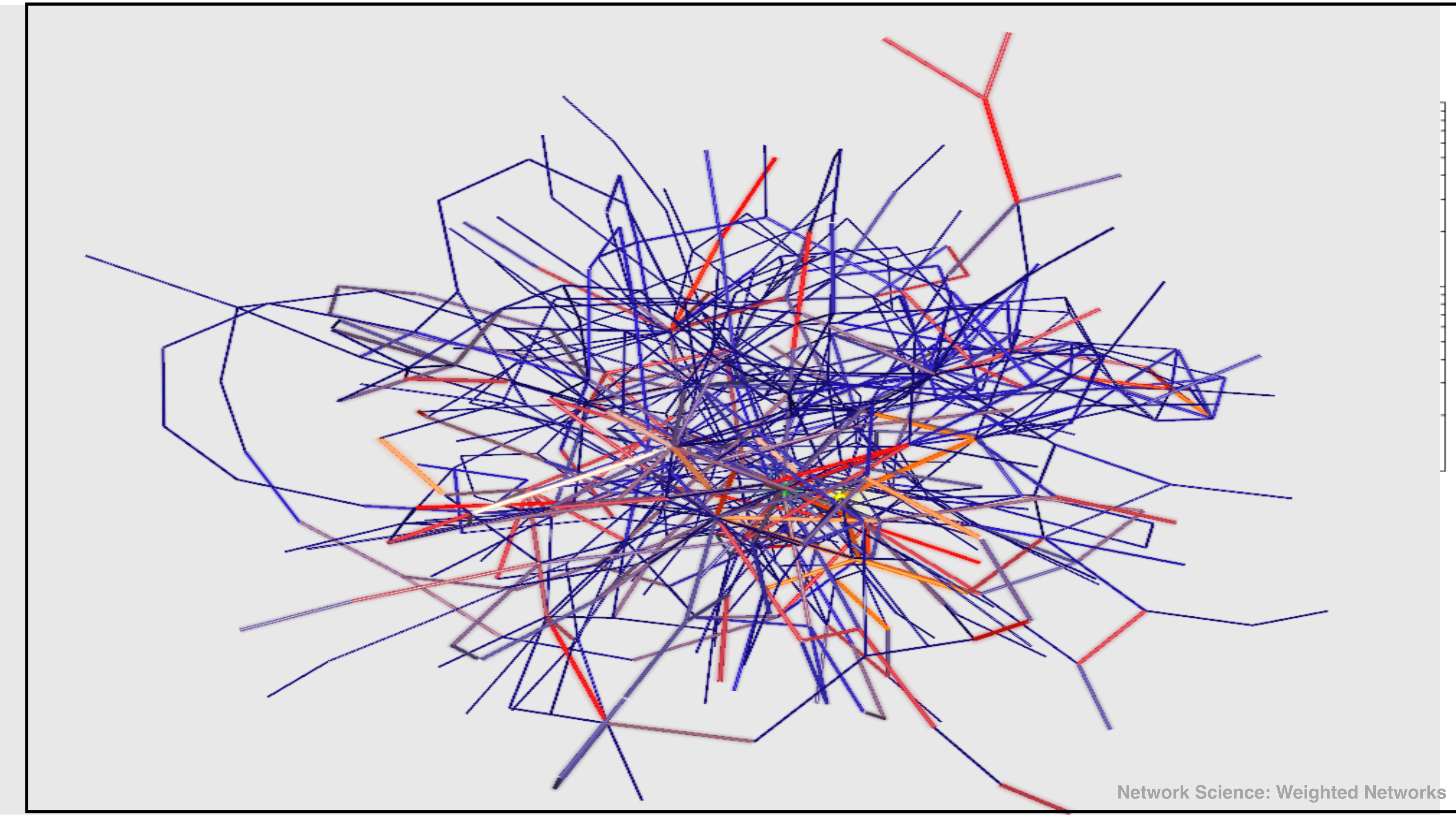
Air Traffic:

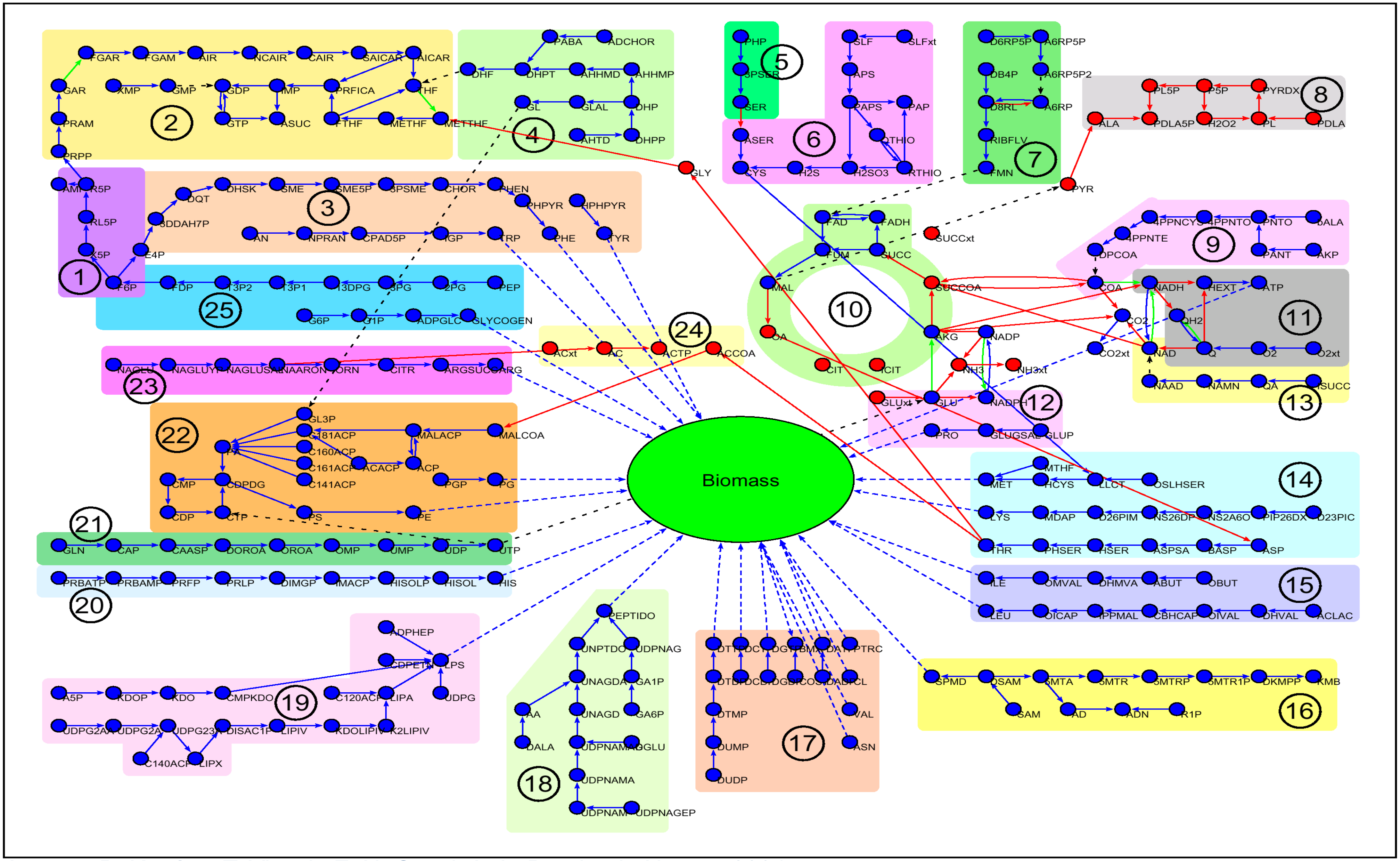


- If $Y_2(i)$» $1/k_i$ 1: A few dominant link weights
- If $Y_2(i)$~ $1/k_i$: No dominant link weights

$Y_2(k)$» $1/k$ → No dominant connection

APPLICATION: DISPARITY IN METABOLIC FLUXES



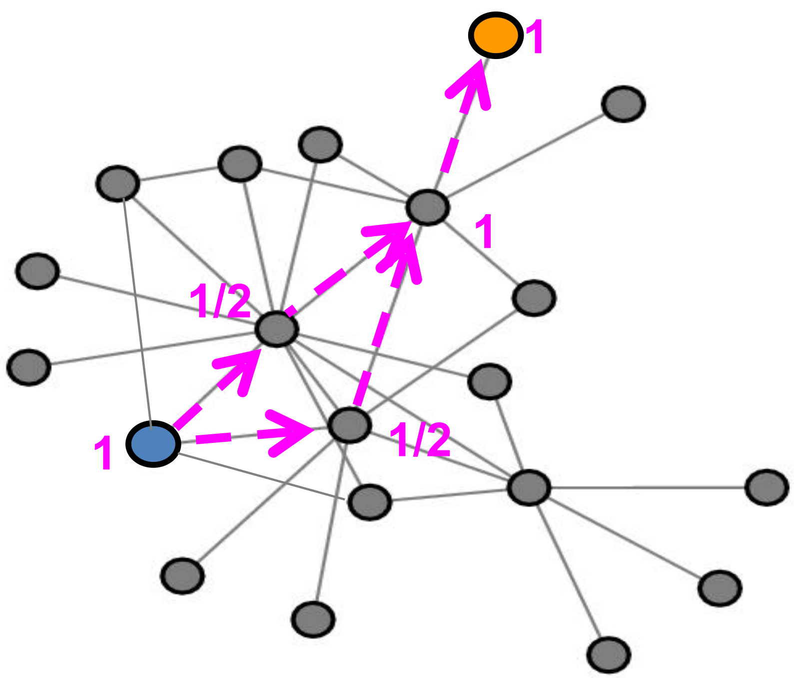


E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A.-L. Barabasi. Nature, 2004

BETWEENNESS CENTRALITY

BETWEENESS CENTRALITY OR LOAD IN NETWORKS

Load: # of packets that a node should handle during the shortest path-based transport from all nodes to all others.

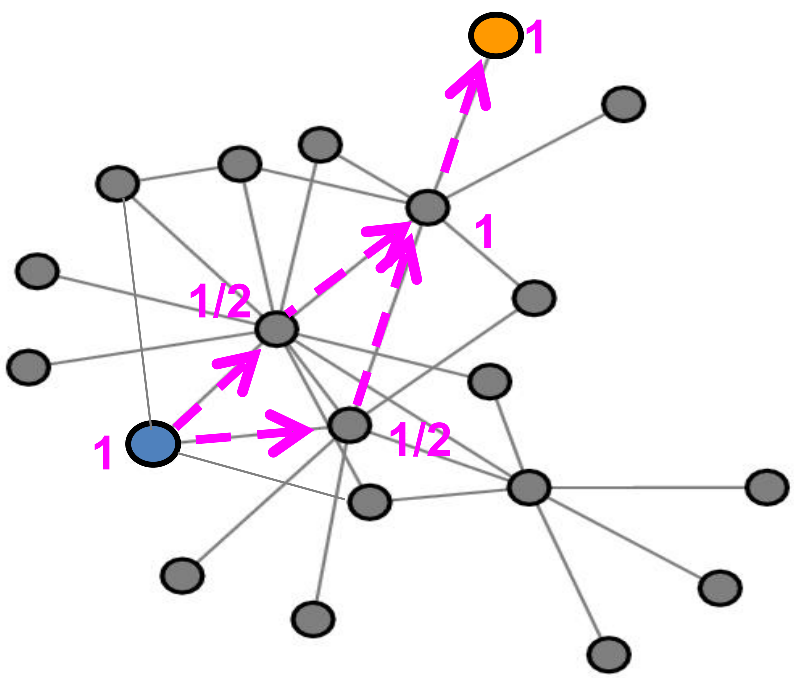


History: [Freeman '77]

- Measure of "centrality" of a person in a social network.
- How "influential" is a person, assuming that influence is determined by the mutual communication among the population via shortest paths?

Goh et al, PRL 2001

BETWEENNESS CENTRALITY OR LOAD IN NETWORKS



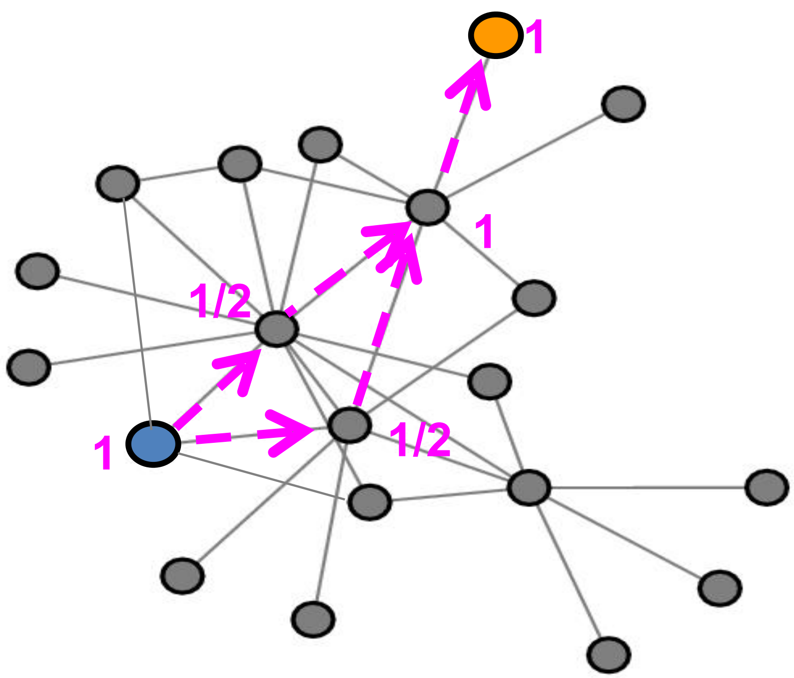
Betweenness centrality (BC) of a node v:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

σ_{st}: total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of shortest paths from s to t going through v.

BC scales as the number of pairs of nodes ($s \neq t \neq v$) so we rescale it by (N-1)(N-2)/2
(N: number of nodes in the giant component)

BETWEENNESS CENTRALITY OR LOAD IN NETWORKS



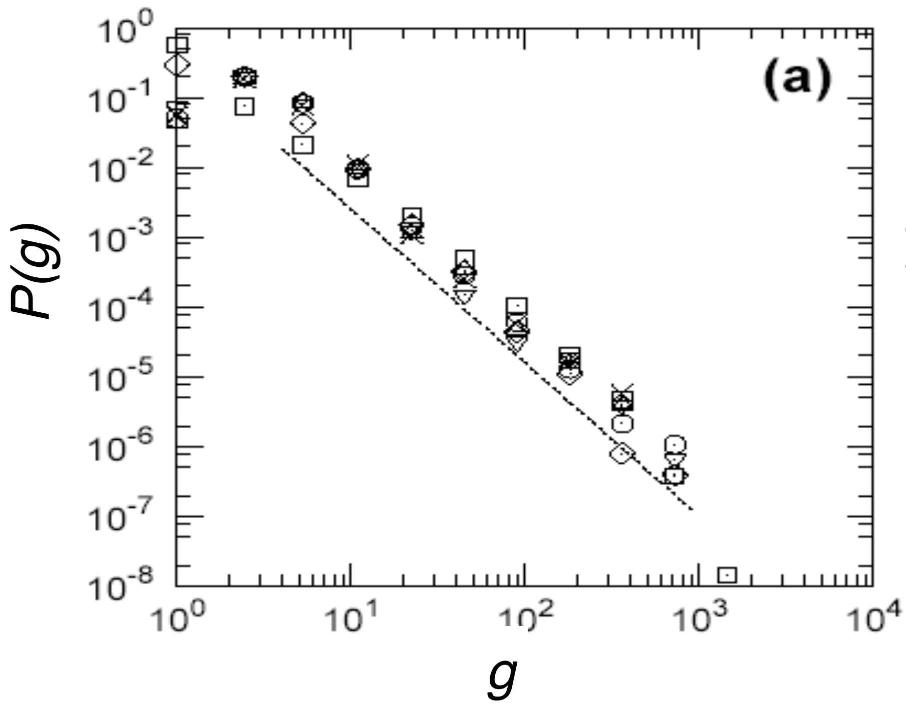
Betweenness centrality depends on shortest paths. These need consideration in weighted networks:

Does the weight represent a distance? (Higher = longer path)
A strength of connection? (Higher = shorter path).

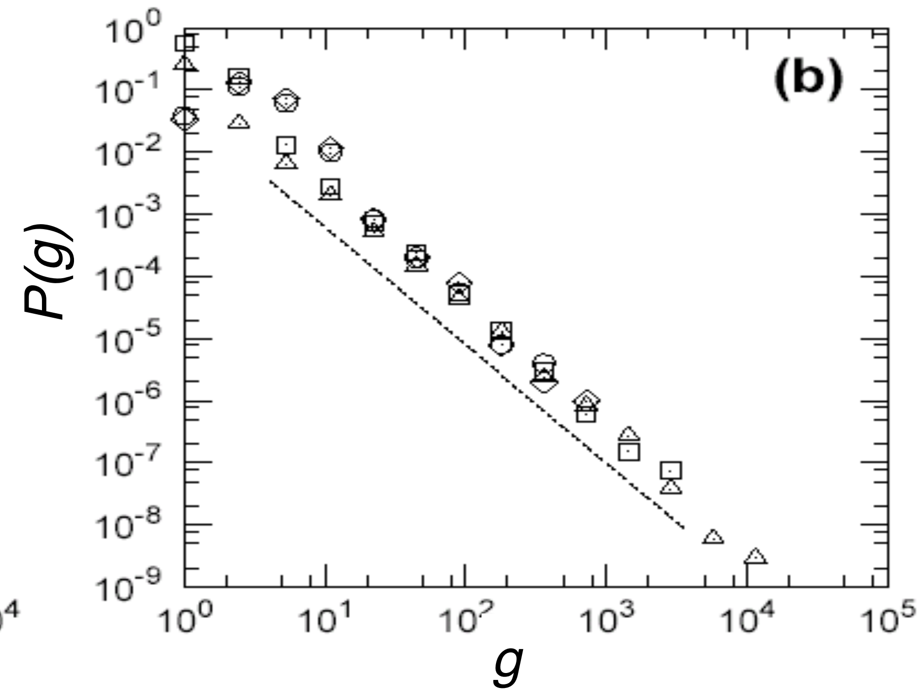
NetworkX handles most weighted measures, but be mindful of what you are feeding it and avoid 'Garbage In, Garbage Out'!

LOAD DISTRIBUTION OF REAL NETWORKS

The load distributions of many real-world SF networks also follow power laws.



Collaboration network, protein interaction network, metabolic network of eukaryotes, etc.



Internet, metabolic network of archaea, WWW, etc.

Goh et al, PNAS 2002

$$P(g) \sim g^{-\delta}$$

The end