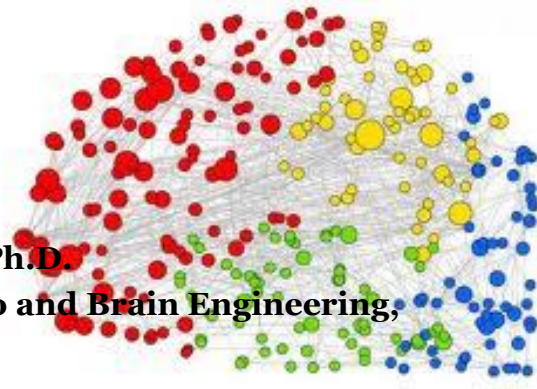
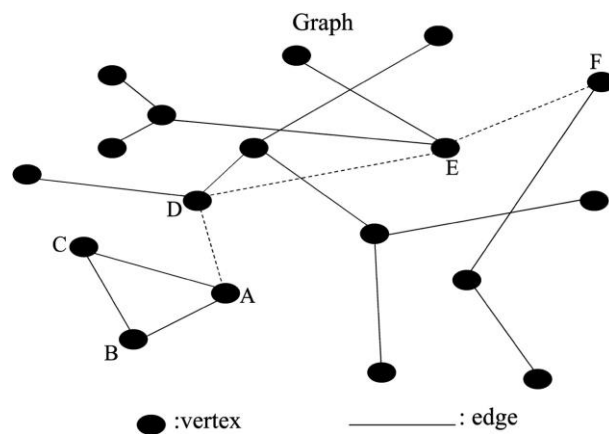


## 복잡계와 과학: Connectome 을 중심으로

**Jaeseung Jeong, Ph.D.**  
**Department of Bio and Brain Engineering,**  
**KAIST**









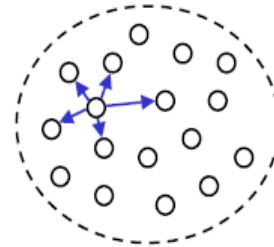
### Vertex (node) and edge (link)



Any system can be expressed as a network with nodes and links.

## Examples of complex networks: geometric, regular







Network	Nodes	Edges
 BZ reaction	<i>molecules</i>	<i>collisions</i>
 slime mold	<i>amoebae</i>	<i>cAMP</i>
 animal coats	<i>cells</i>	<i>morphogens</i>
 insect colonies	<i>ants, termites</i>	<i>pheromone</i>
 flocking, traffic	<i>animals, cars</i>	<i>perception</i>
 swarm sync	<i>fireflies</i>	<i>photons</i> <i>± long-range</i>

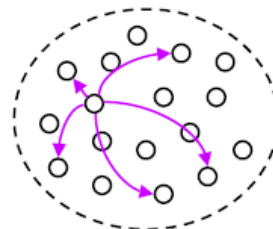


- interactions inside a local neighborhood in 2-D or 3-D geometric space
- limited "visibility" within Euclidean distance

Eileen Kraemer

## Examples of complex networks: semi-geometric, irregular

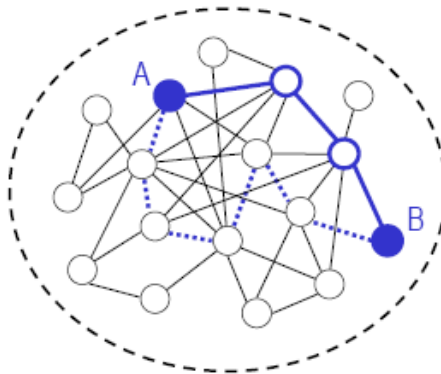
Network	Nodes	Edges
 Internet	<i>routers</i>	<i>wires</i>
 brain	<i>neurons</i>	<i>synapses</i>
 WWW	<i>pages</i>	<i>hyperlinks</i>
 Hollywood	<i>actors</i>	<i>movies</i>
 gene regulation	<i>proteins</i>	<i>binding sites</i>
 ecology web	<i>species</i>	<i>competition</i>



- local neighborhoods (also) contain "long-range" links:
  - either "element" nodes located in space
  - or "categorical" nodes not located in space
- still limited "visibility", but not according to distance

Eileen Kraemer

## Structural metrics: Average path length



The path length between A and B is 3

- the *path length* between two nodes  $A$  and  $B$  is the smallest number of edges connecting them:

$$l(A, B) = \min l(A, A_p \dots A_n, B)$$

- the *average path length* of a network over all pairs of  $N$  nodes is

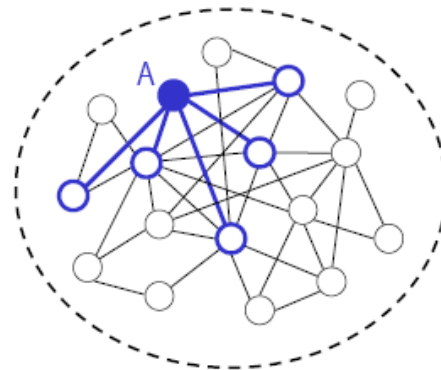
$$L = \langle l(A, B) \rangle = \frac{2N(N-1)}{N(N-1)} \sum_{A,B} l(A, B)$$

- the *network diameter* is the maximal path length between two nodes:

$$D = \max l(A, B)$$

- property:  $1 \leq L \leq D \leq N-1$

## Structural Metrics: Degree distribution(connectivity)



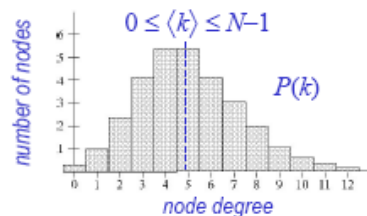
The degree of A is 5

- the *degree* of a node  $A$  is the number of its connections (or neighbors),  $k_A$

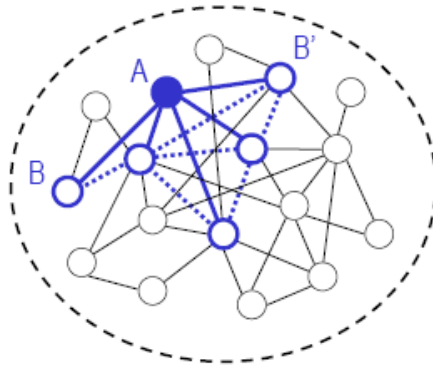
- the *average degree* of a network is

$$\langle k \rangle = \frac{1}{N} \sum_A k_A$$

- the *degree distribution* function  $P(k)$  is the histogram (or probability) of the node degrees: it shows their spread around the average value



## Structural Metrics: Clustering coefficient



The clustering coefficient of A is 0.6

A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.

- the *neighborhood* of a node  $A$  is the set of  $k_A$  nodes at distance 1 from  $A$
- given the number of *pairs* of neighbors:

$$F_A = \sum_{B, B'} 1 \\ = k_A(k_A - 1) / 2$$

- and the number of pairs of neighbors that are also *connected* to each other:

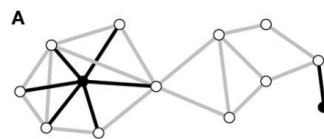
$$E_A = \sum_{B \leftrightarrow B'} 1$$

- the *clustering coefficient* of  $A$  is

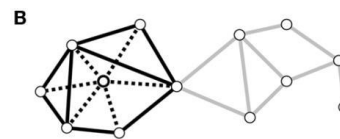
$$C_A = E_A / F_A \leq 1$$

- and the *network clustering coefficient*:

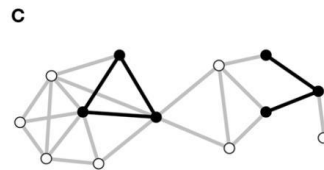
$$\langle C \rangle = 1/N \sum_A C_A \leq 1$$



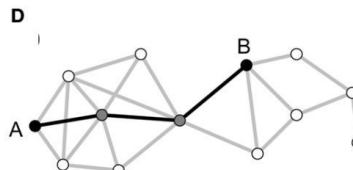
node degree



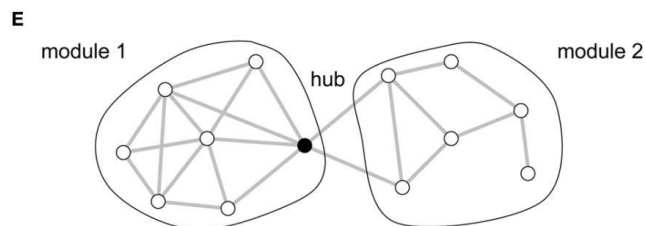
clustering coefficient



motifs



path length and distance



community structure – modules and hubs

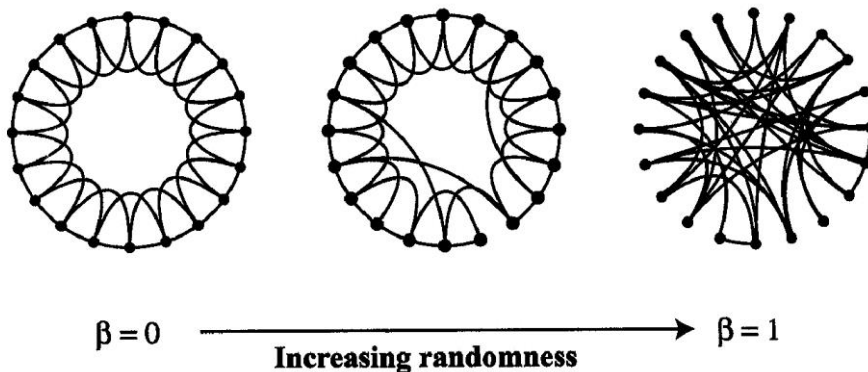
D. J. Watts and Steven Strogatz (June 1998).  
 "Collective dynamics of 'small-world' networks".  
 Nature 393 (6684): 440–442.

**Table 1 Empirical examples of small-world networks**

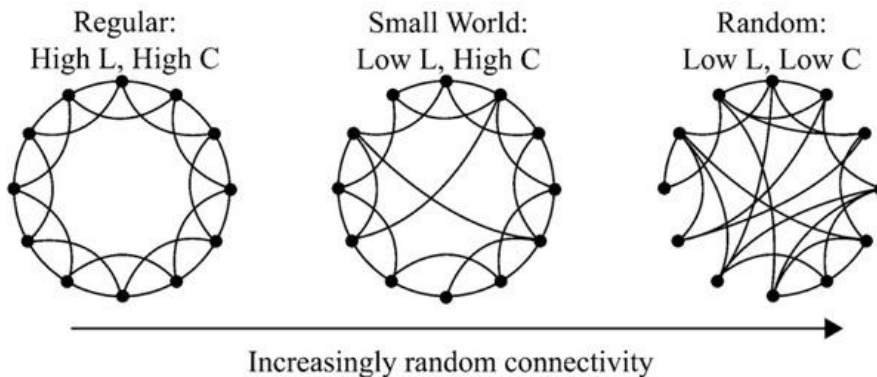
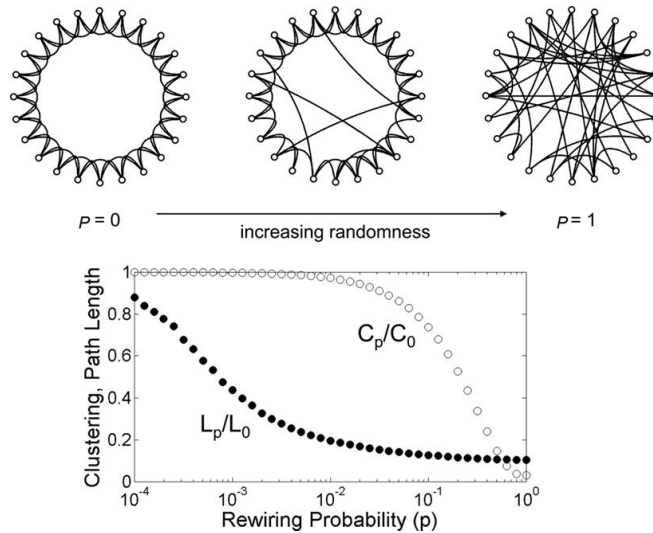
	$L_{\text{actual}}$	$L_{\text{random}}$	$C_{\text{actual}}$	$C_{\text{random}}$
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
<i>C. elegans</i>	2.65	2.25	0.28	0.05

Characteristic path length  $L$  and clustering coefficient  $C$  for three real networks, compared to random graphs with the same number of vertices ( $n$ ) and average number of edges per vertex ( $k$ ). (Actors:  $n = 225,226$ ,  $k = 61$ . Power grid:  $n = 4,941$ ,  $k = 2.67$ . *C. elegans*:  $n = 282$ ,  $k = 14$ .) The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component<sup>16</sup> of this graph, which includes ~90% of all actors listed in the Internet Movie Database (available at <http://us.imdb.com>), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction. We treat all edges as undirected and unweighted, and all vertices as identical, recognizing that these are crude approximations. All three networks show the small-world phenomenon:  $L \geq L_{\text{random}}$  but  $C \gg C_{\text{random}}$ .

D. J. Watts and Steven Strogatz (June 1998).  
 "Collective dynamics of 'small-world' networks".  
 Nature 393 (6684): 440–442.



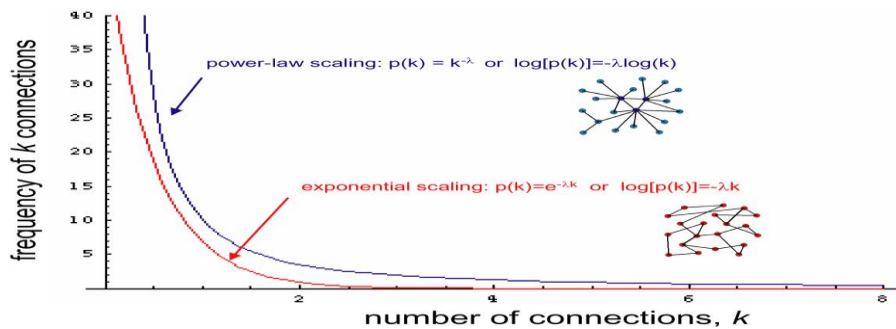
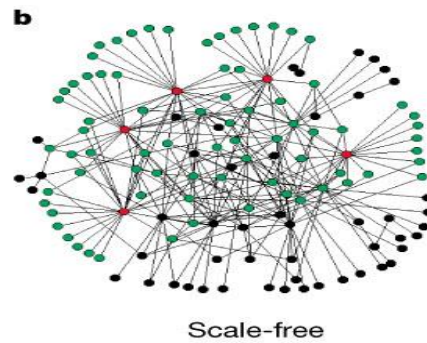
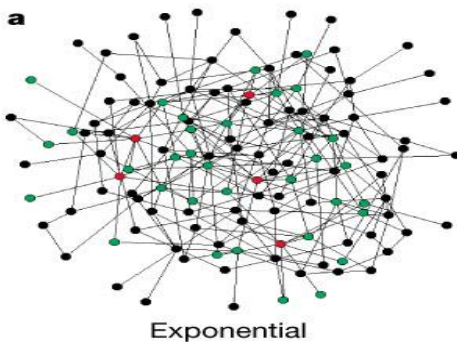
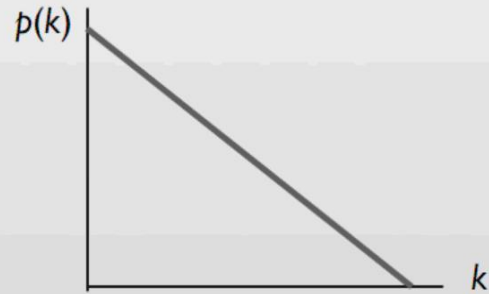
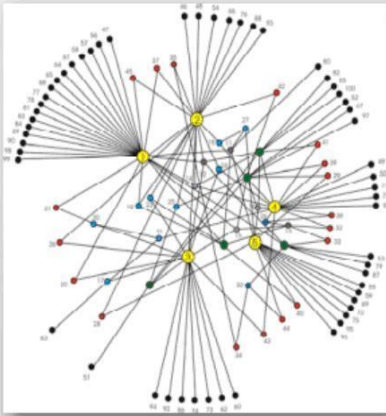
D. J. Watts and Steven Strogatz (June 1998).  
 "Collective dynamics of 'small-world' networks".  
 Nature 393 (6684): 440–442.



A **small-world network** is a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps.

## Scale-free network

- A minority of nodes have a majority of links
- $p(k)$  follows a **power law** distribution



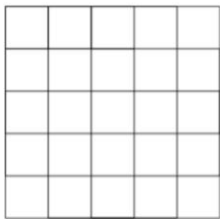






## Models

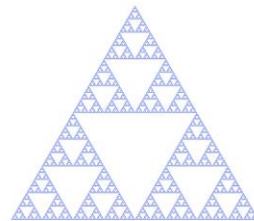
- Erdős-Rényi → Homogeneous
  - Each possible link exists with probability  $p$
- Scale-free → Heterogeneous
  - The network grows a node at a time
  - The probability  $\Pi_i$  that the new node is connected to node  $i$  is proportional to how many links node  $i$  owns (preferential attachment)



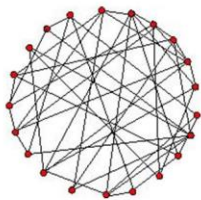
Regular lattice



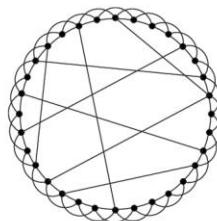
Bethe lattice



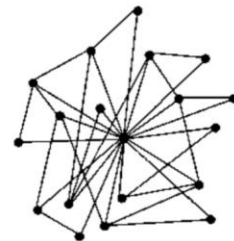
Fractal



Random network

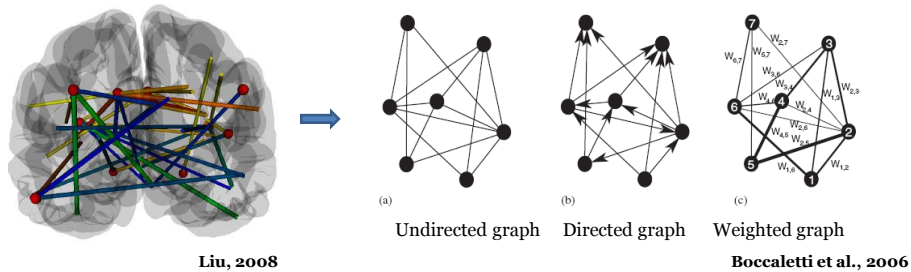


WS smallworld



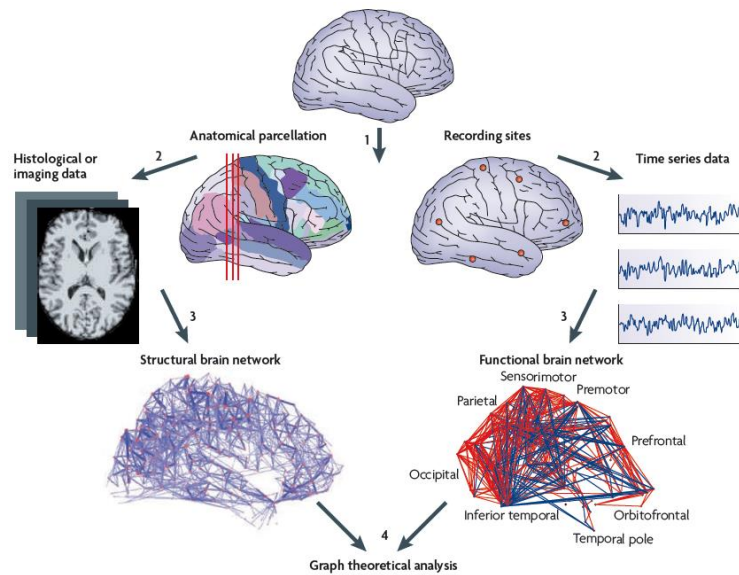
BA scale free

## Brain and complex network (graph) theory



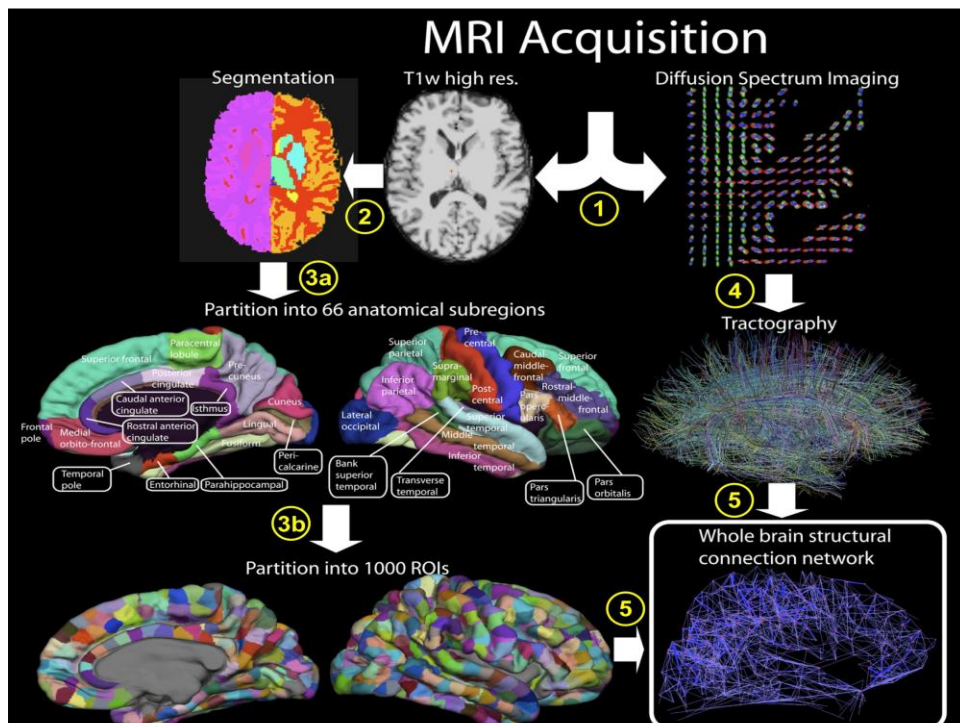
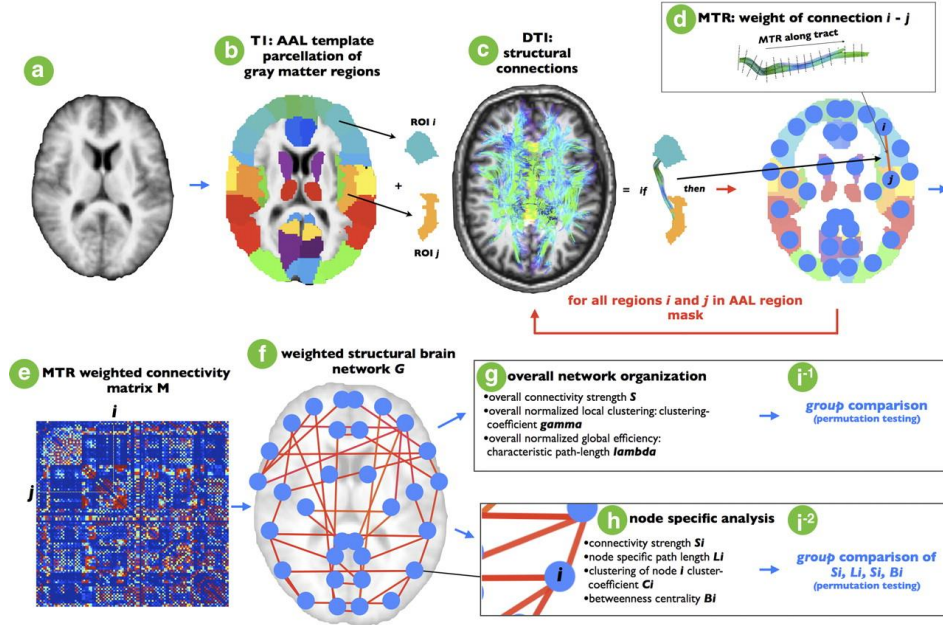
- Node (vertex) : Brain region or voxel, channel of EEG/MEG
- Link (edge) : Functional or anatomical connection between nodes
- Network analysis can reveal structural and functional organization of the brain (Liu, 2008)

## Constructing Brain Networks

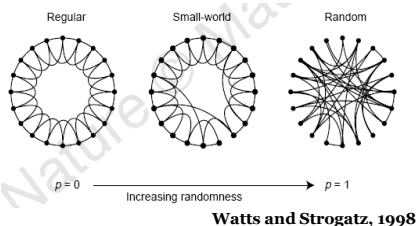


Bullmore and Sporns, 2009

## network analysis of structural brain networks



# Brain is a small-world network



high C  
high L

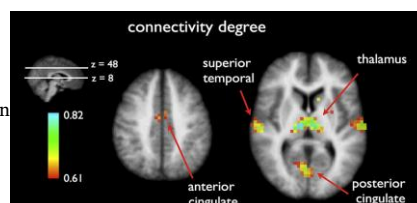
high C  
low L

low C  
low L

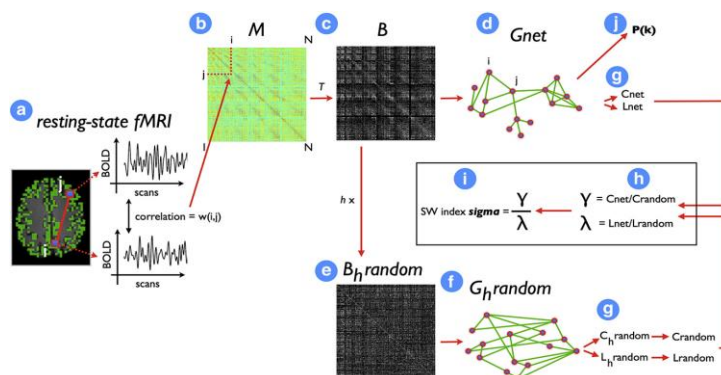
- high clustering coefficient (C) – high resilience to damage in local structures
- low average path length (L) – high level of global communication efficiency
- Brain functional network has small-world structure, while this property may be disrupted in damaged brain such as AD (vulnerability to damages, decreased communication efficiency between distant brain regions ... )

Small-world and scale-free organization of voxel-based resting state functional connectivity in the human brain  
van den Heuvel et al., Neuroimage, 2008

normal, resting-state, voxel-based (N=10,000), zero-lag temporal correlation, bandpass-filtered (0.01-0.1Hz), unweighted, **small-world and scale-free**: optimal network organization balance between maximum communication efficiency and minimum wiring



AD, damage modeling, weighted graph, efficiency



A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs (Achard et al., The Journal of Neuroscience, 2006)

MODWT (Maximal Overlap Discrete Wavelet Transform) at 6 frequency scales

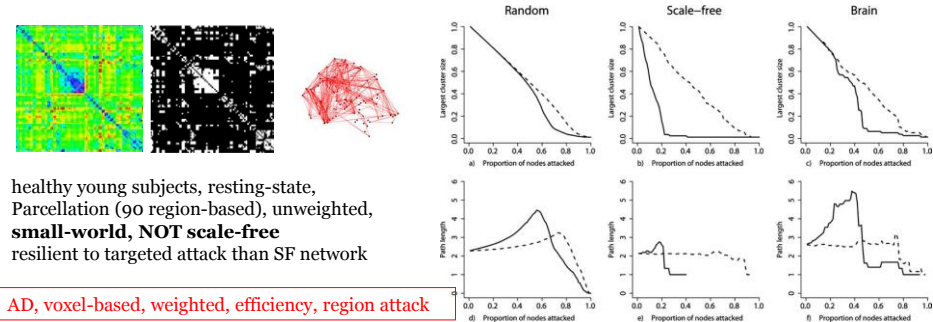
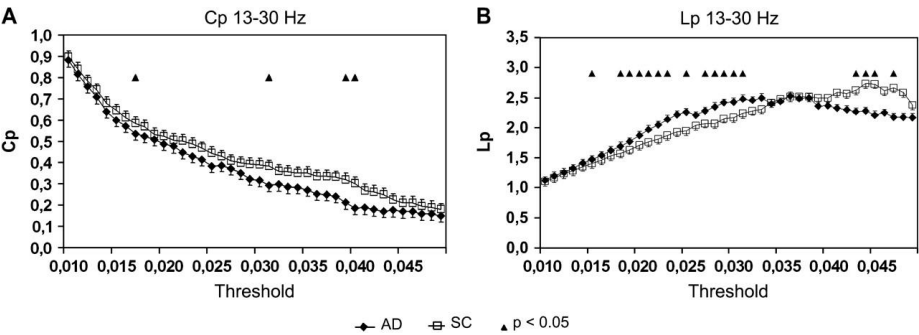


Table 1. Wavelet scale dependency of functional connectivity and small-world parameters for an entire human brain network

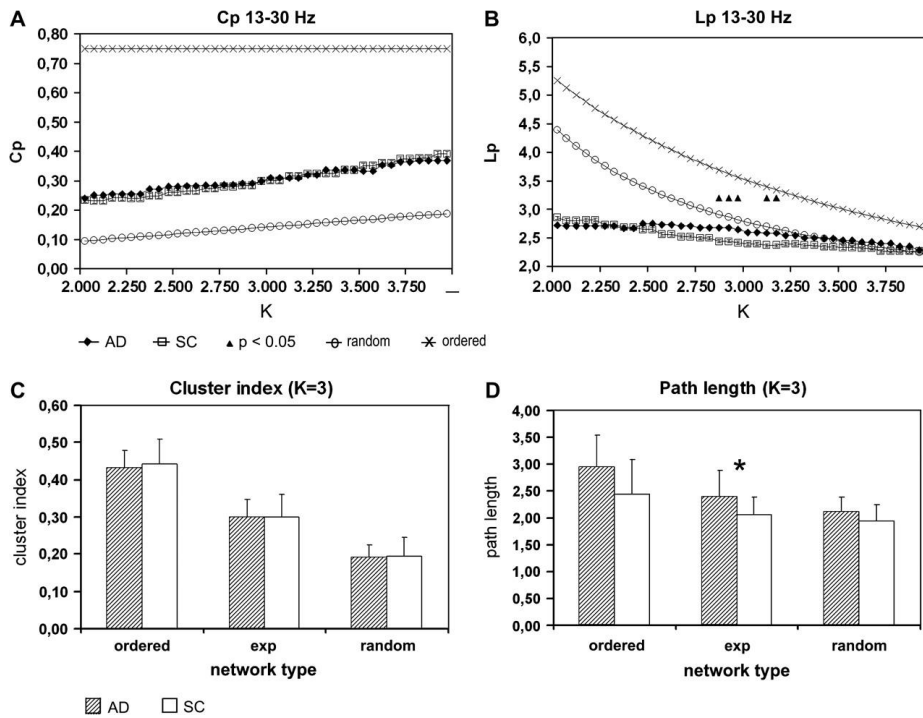
Scale	Hz	$r$	$R$	$L_{net}$	$C_{net}$	$\lambda$	$\gamma$	$\sigma$
1	0.23–0.45	0.12	0.13	2.9	0.534	1.28	1.81	1.42
2	0.11–0.23	0.21	0.2	2.6	0.566	1.12	2.14	1.92
3	0.06–0.11	0.39	0.39	2.69	0.555	1.16	2.22	1.91
4	0.03–0.06	0.45	0.44	2.49	0.525	1.08	2.38	2.19
5	0.01–0.03	0.44	0.35	2.4	0.554	1.04	2.39	2.30
6	0.007–0.01	0.41	0.17	2.65	0.515	1.15	2.15	1.88

Scales 1–6 of the MODWT denote progressively lower-frequency intervals (Hz).  $r$  is the mean inter-regional correlation, and  $R$  is the correlation threshold.  $L_{net}$  and  $C_{net}$  are the mean path length and clustering coefficient, respectively, of the thresholded network. The  $\lambda$  and  $\gamma$  are ratios of brain network path length and clustering coefficient, respectively, to comparable random network metrics. The equation  $\sigma = \gamma/\lambda$  is a scalar measure of "small-worldness."

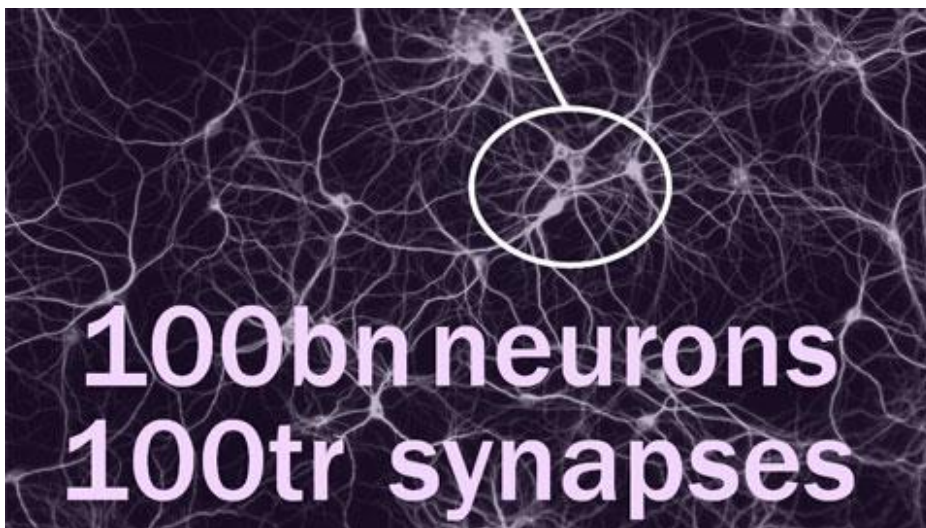
## Alzheimer vs. Healthy subjects







**Why does the brain process  
information so quickly?**



## Power grid of North America



- Nodes: generators, transmission sub-stations, distribution sub-stations
- Edges: high-voltage transmission lines
- 14,099 nodes: 1,633 generators, 2,179 distribution substations, the rest transmission substations
- 19,657 edges



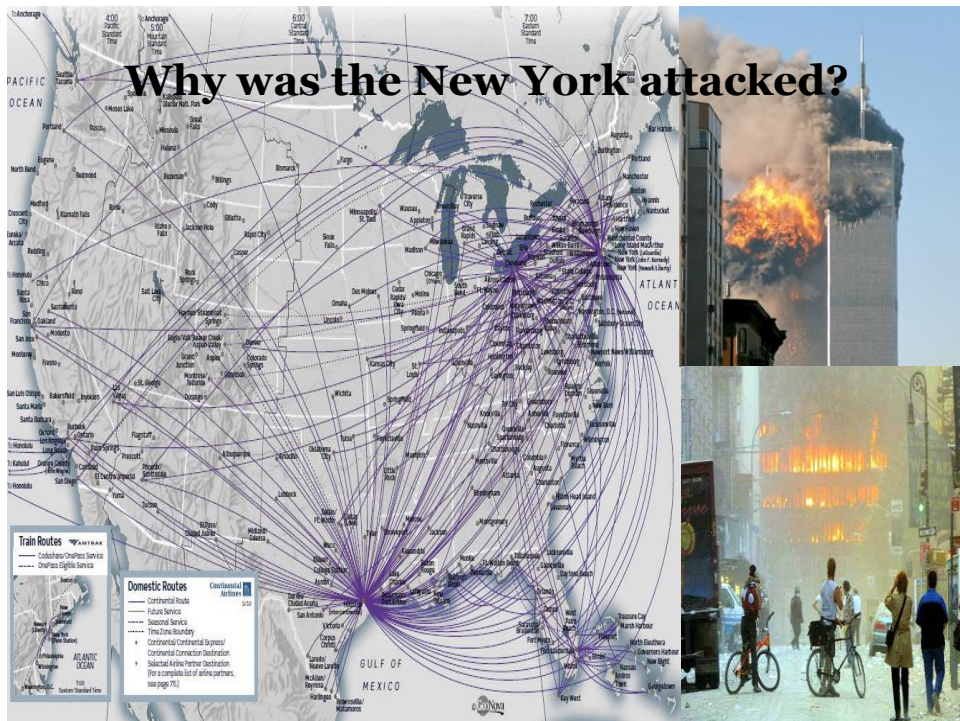
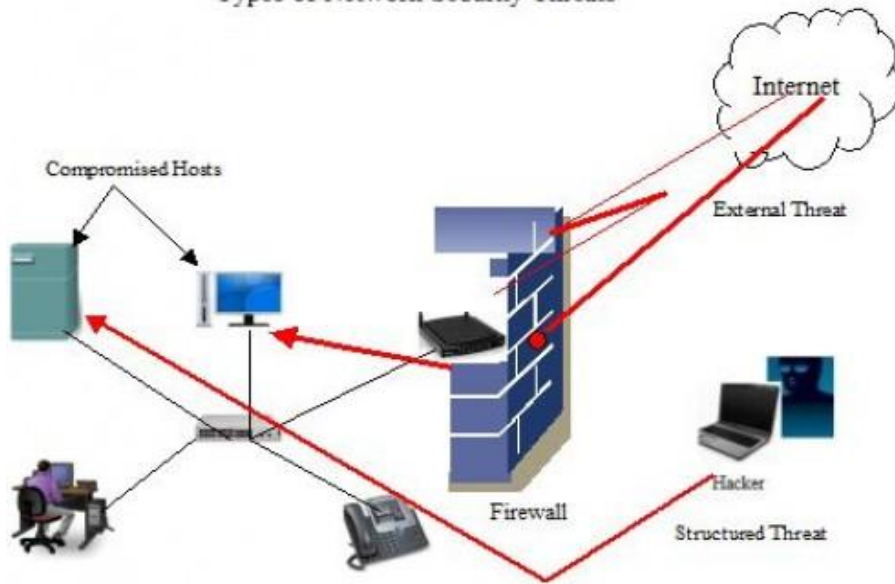
The Northeast blackout of 2003 was a widespread power outage (the Northeastern and Midwestern US and Ontario in Canada on August 14, 2003, just before 4:10 p.m.

The blackout's primary cause was a software bug in the alarm system at a control room of the FirstEnergy Corporation in Ohio.

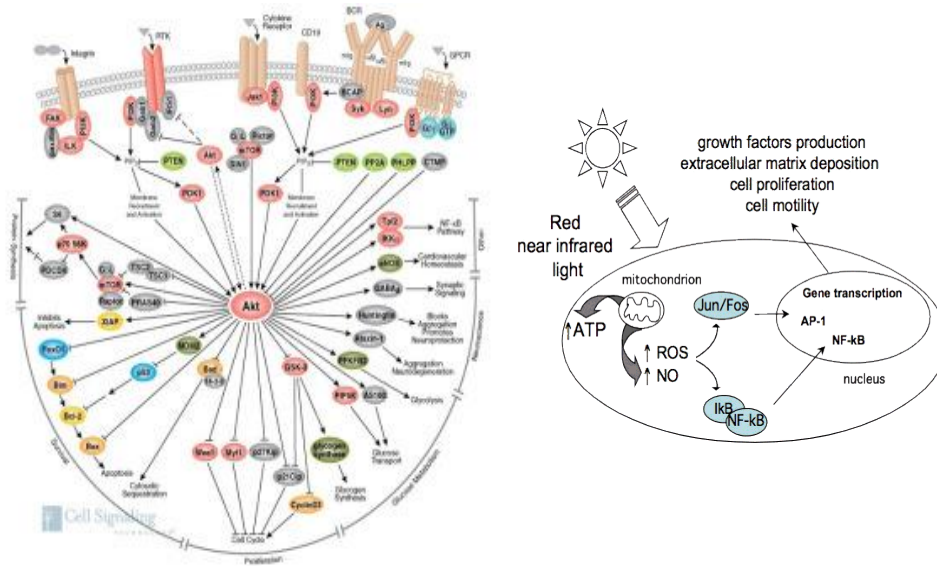
Operators were unaware of the need to re-distribute power after overloaded transmission lines hit unpruned foliage. A manageable local blackout was cascaded into widespread distress on the electric grid.



## Types of Network Security Threats



## Biological networks have critical nodes



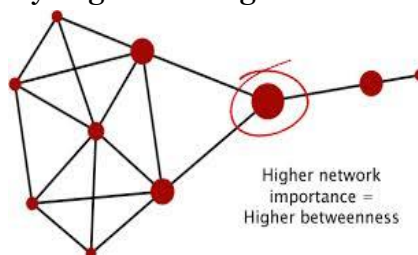
**Networks are often attacked: sometimes they are robust to this attack, and sometimes they are lethal.**

## Network robustness and resilience

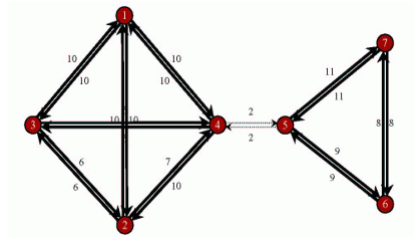
If a given fraction of nodes or edges are removed, what happens in the network? How large are the connected components? What is the average distance between nodes in the components?

### Random failure vs. targeted attack

- Edge removal
  - Random failure: each edge is removed with probability  $(1-p)$  corresponds to random failure of links.
  - targeted attack: causing the most damage to the network with the removal of the fewest edges.
    - strategies: remove edges that are most likely to break apart the network or lengthen the average shortest path
    - e.g. usually edges with high betweenness.



## Betweenness centrality



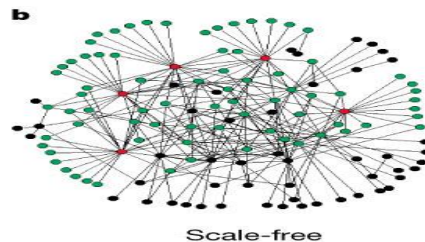
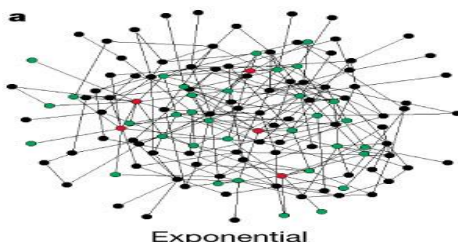
Several measures capture variations on the notion of a vertex's importance in a graph. Let  $\sigma_{st} = \sigma_{ts}$  denote the number of shortest paths from  $s \in V$  to  $t \in V$ , where  $\sigma_{ss} = 1$  by convention. Let  $\sigma_{st}(v)$  denote the number of shortest paths from  $s$  to  $t$  that some  $v \in V$  lies on. The following are standard measures of centrality:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

*betweenness centrality*  
(Freeman, 1977; Anthonisse, 1971)

## Models

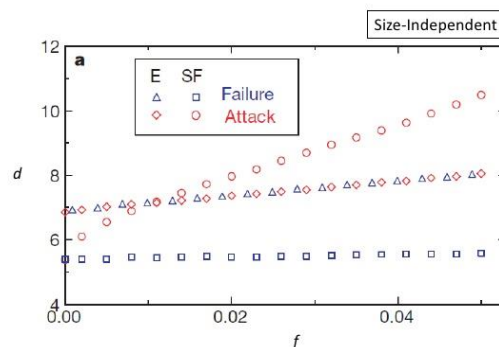
- Erdős-Rényi → Homogeneous
  - Each possible link exists with probability  $p$
- Scale-free → Heterogeneous
  - The network grows a node at a time
  - The probability  $\Pi_i$  that the new node is connected to node  $i$  is proportional to how many links node  $i$  owns (preferential attachment)



## diameter

- The interconnectedness of a network is described by its diameter  $d$ , defined as the average length of the shortest paths between any two nodes in the network.
- The diameter characterizes the ability of two nodes to communicate with each other: the smaller  $d$  is, the shorter is the expected path between them.

### Diameter Change by Failure and Attack

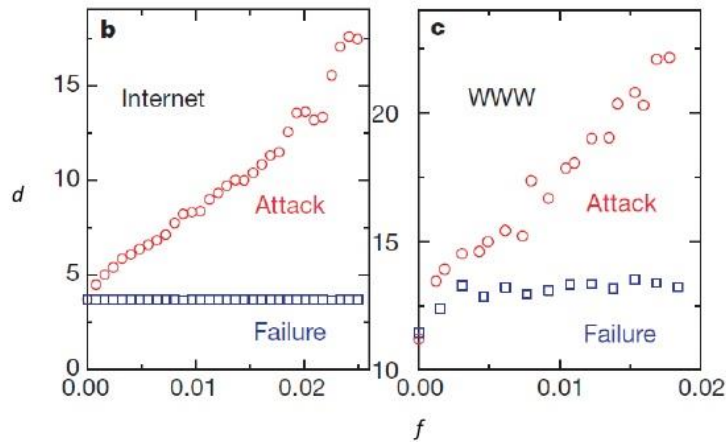


Changes in the diameter  $d$  of the network as a function of the fraction  $f$  of the removed nodes. Comparison between the exponential (E) and scale-free (SF) network models, each containing  $N^{1/4}$  10,000 nodes and 20,000 links.

R. Albert, H. Jeong, and A.-L. Barabasi, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.

## Diameter Change by Failure and Attack

### Scale-free network

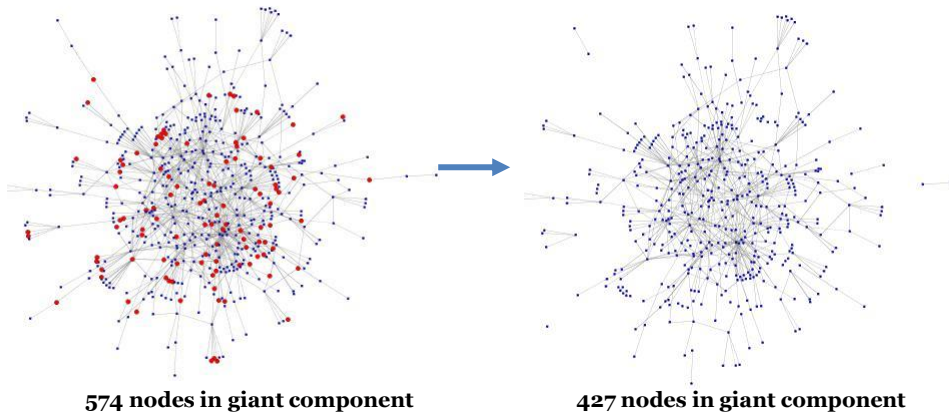


R. Albert, H. Jeong, and A.-L. Barabasi, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.

### Scale-free networks are resilient with respect to random attack

Gnutella network, 20% of nodes removed.

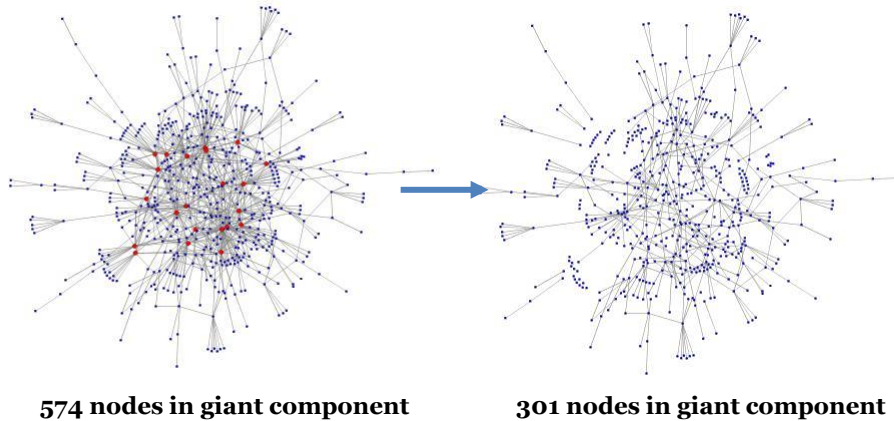
(Gnutella is the first decentralized peer-to-peer network.)





## Targeted attacks are effective against scale-free networks

Same gnutella network, 22 most connected nodes removed (2.8% of the nodes)

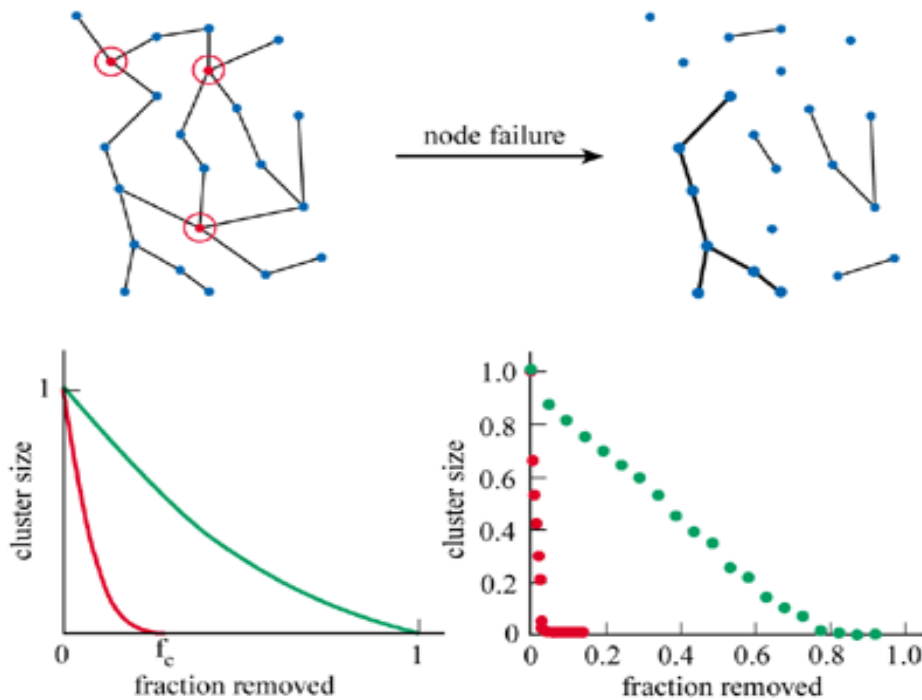


## The fragmentation process

We measure the size of the largest cluster,  $S$ , when a fraction  $f$  of the nodes are removed either randomly or in an attack mode.

We find that for the exponential network, as we increase  $f$ ,  $S$  displays a threshold-like behavior. Similar behaviour is observed when we monitor the average size of the isolated clusters (that is, all the clusters except the largest one), finding its increase rapidly until at  $f_c$ , after which it then decreases to 1.

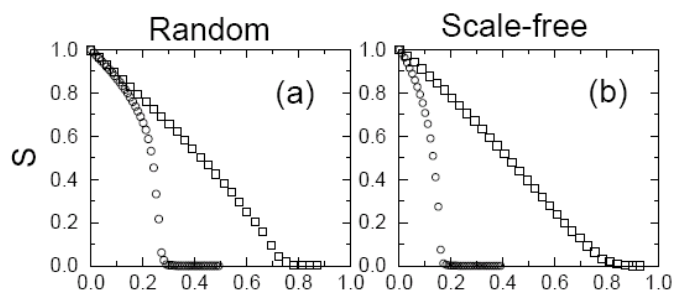




## Random network resilience to targeted attacks

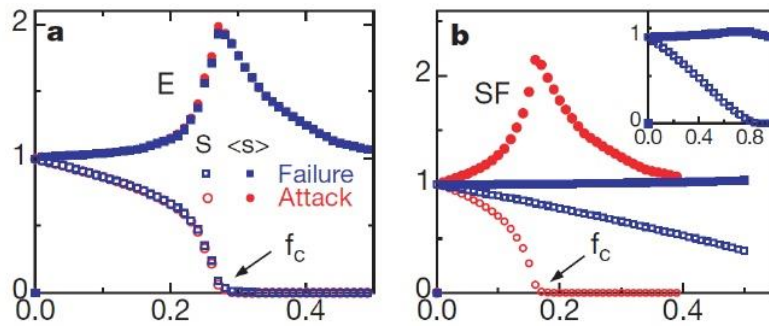
For random networks there is smaller difference between random failures and targeted attacks.

The size  $S$  is defined as the fraction of nodes contained in the largest cluster (that is,  $S = 1$  for  $f = 0$ ).



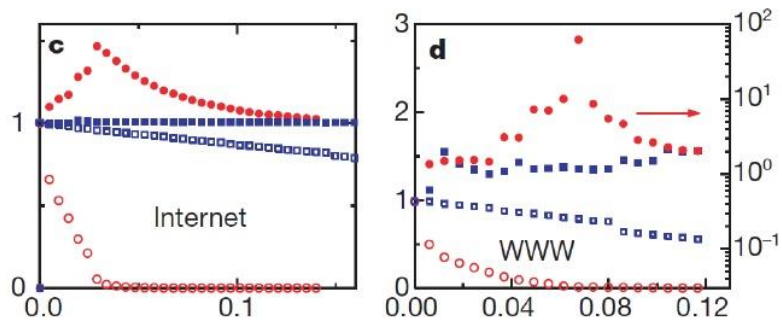
R. Albert, H. Jeong, and A.-L. Barabasi, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.

## Fragmentation by Failure and Attack



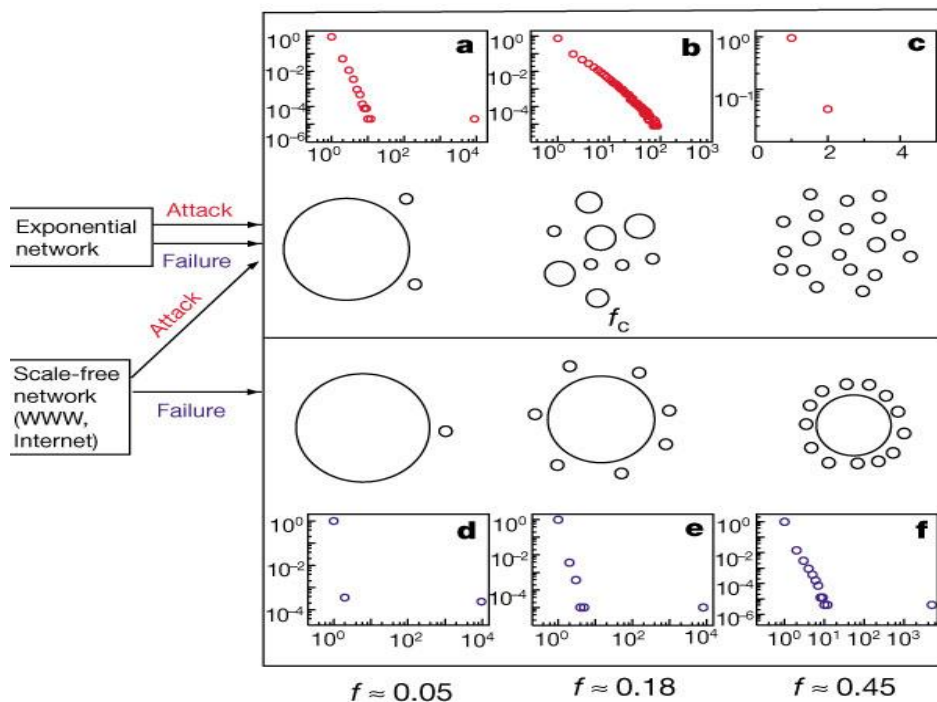
Network fragmentation under random failures and attacks. The relative size of the largest cluster  $S$  (open symbols) and the average size of the isolated clusters (filled symbols) as a function of the fraction of removed nodes.

## Fragmentation by Failure and Attack



## Summary of the response of a network to failures or attacks I

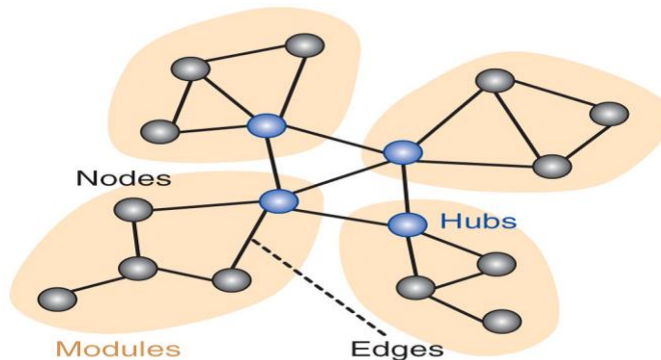
- The cluster size distribution for various values of  $f$  when a scale-free network of parameters is subject to random failures or attacks.
- Upper panels, exponential networks under random failures and attacks and scale-free networks under attacks behave similarly. For small  $f$ , clusters of different sizes break down, although there is still a large cluster. This is supported by the cluster size distribution: although we see a few fragments of sizes between 1 and 16, there is a large cluster of size 9,000 (the size of the original system being 10,000).
- At a critical  $f_c$ , the network breaks into small fragments between sizes 1 and 100 (b) and the large cluster disappears. At even higher  $f$  (c) the clusters are further fragmented into single nodes or clusters of size two.

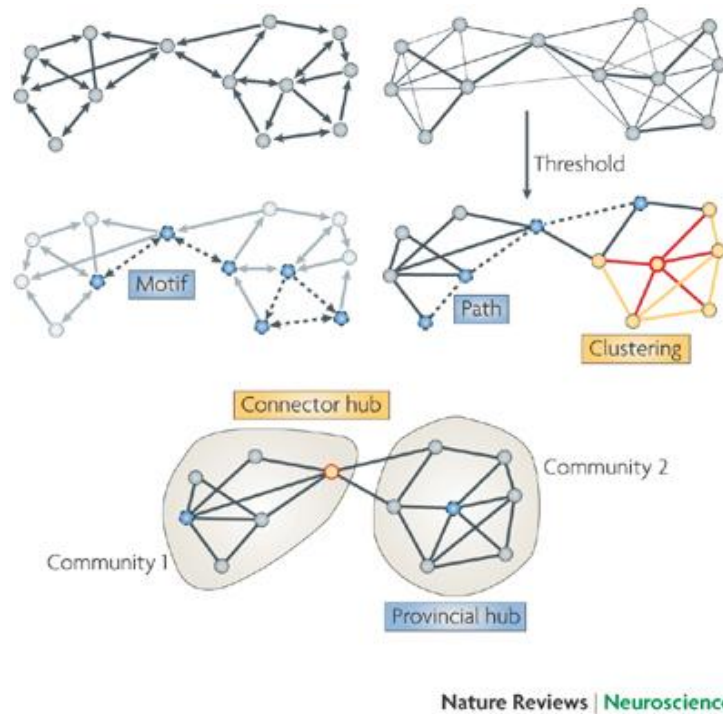


It can be assumed that the cost of attacking is not identical for all nodes. More important nodes are better defended, thus attacking a more important node should be more difficult (= cost more in the model).

## Brain Hubs

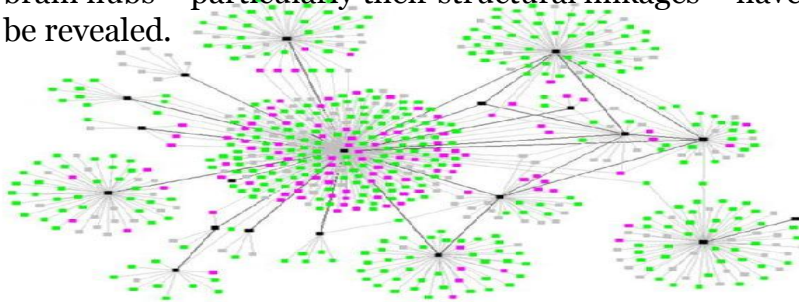
- It has been noted that some of these brain regions play a central role in the overall network organization, as indexed by a high degree, low clustering, short path length, high centrality and participation in multiple communities across the network, identifying them as “brain hubs.”

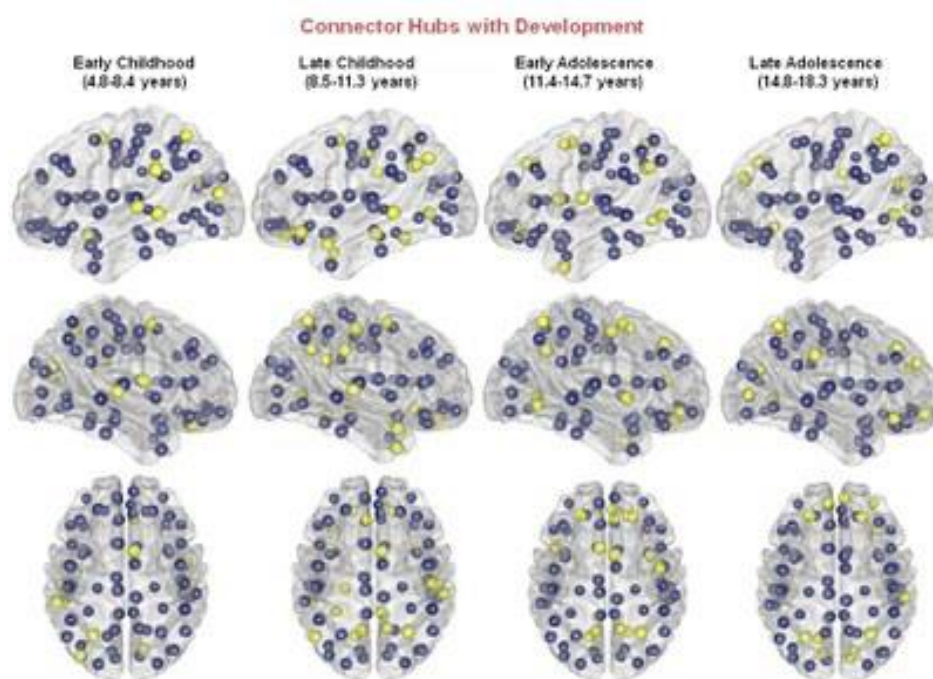
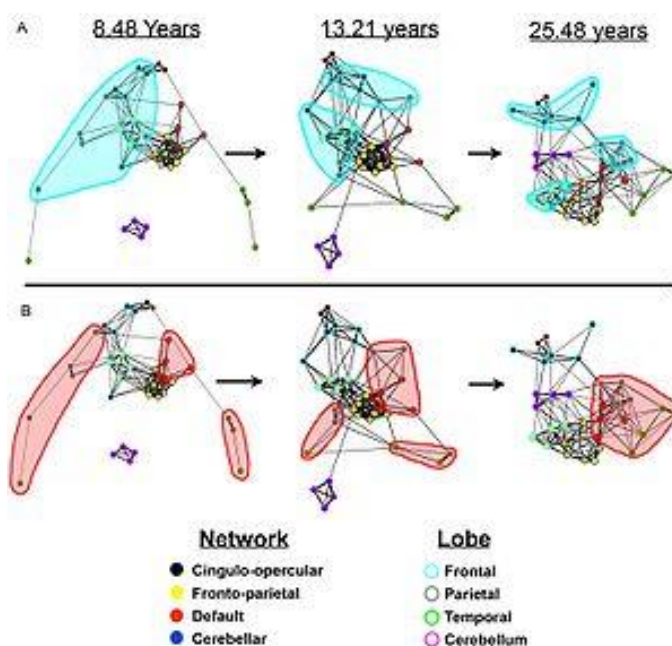




## Brain Hubs

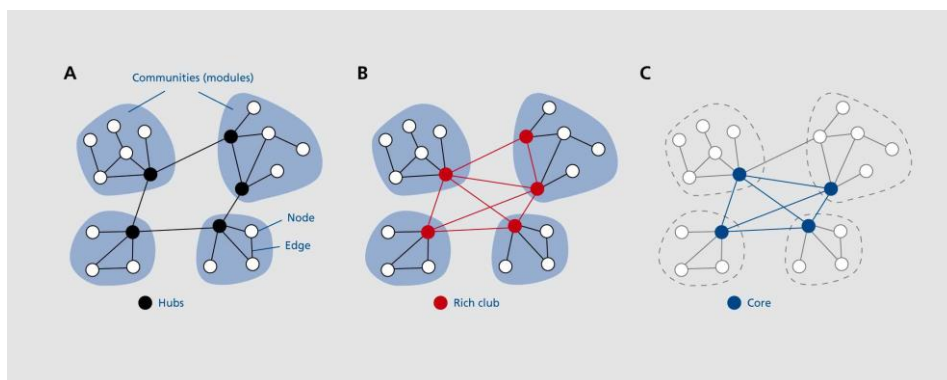
- Examining the function and role of these hubs is of special interest as they play a central role in establishing and maintaining efficient global brain communication, a crucial feature for healthy brain functioning.
- First studies have identified a number of key cortical hubs (Hagmann et al., 2008) but many organizational properties of brain hubs— particularly their structural linkages— have yet to be revealed.





## Rich club phenomena

- The so-called “rich-club” phenomenon in networks is said to be present when the hubs of a network tend to be more densely connected among themselves than nodes of a lower degree.
- The name arises from the analogy with social systems, where highly central individuals— being “rich” in connections—often form a highly interconnected club.
- The presence, or absence, of rich-club organization can provide important information on the higher-order structure of a network, particularly on the level of resilience, hierarchal ordering, and specialization.





## Rich club phenomena

- The strong rich-club tendency of power grids, for example, is related to the necessity of the network to easily distribute the load of one station to the other stations, reducing the possibility of critical failure.
- On the other hand, the absence of rich-club organization in protein interaction networks has been suggested to reflect a high level of functional specialization.

## Global efficiency of the network

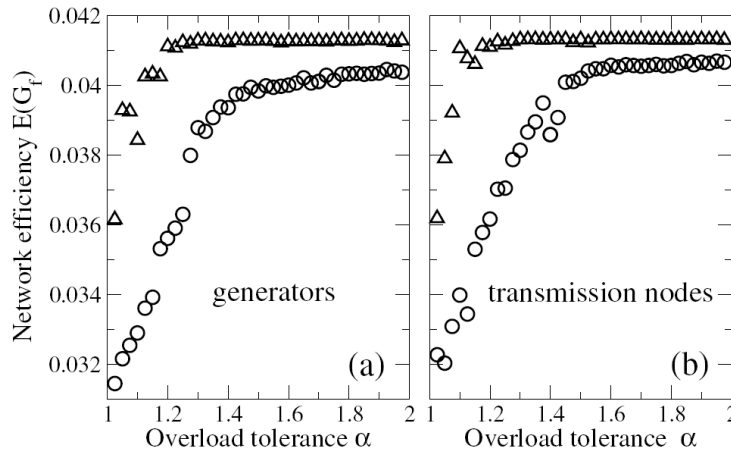
- The inverse of the mean of the minimum path length between each pair of nodes,  $L_{ij}$ , is a measure of the *global* efficiency of parallel information transfer  $E_{global}$  in the network

$$E_{glob}(G) = \frac{\sum_{i \neq j \in G} E_{ij}}{n(n-1)} \\ = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

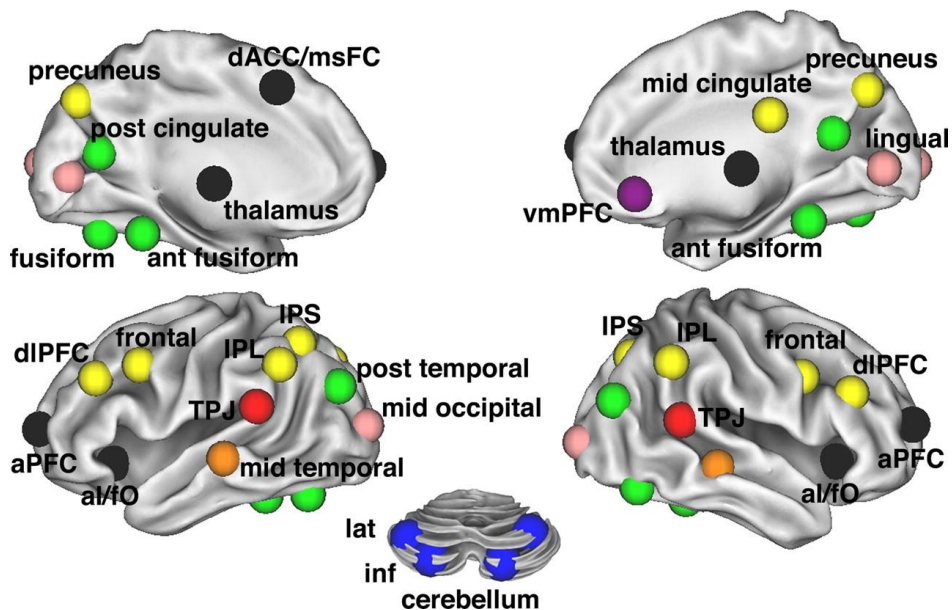
$$L^{-1} = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} \frac{1}{d_{i,j}}$$

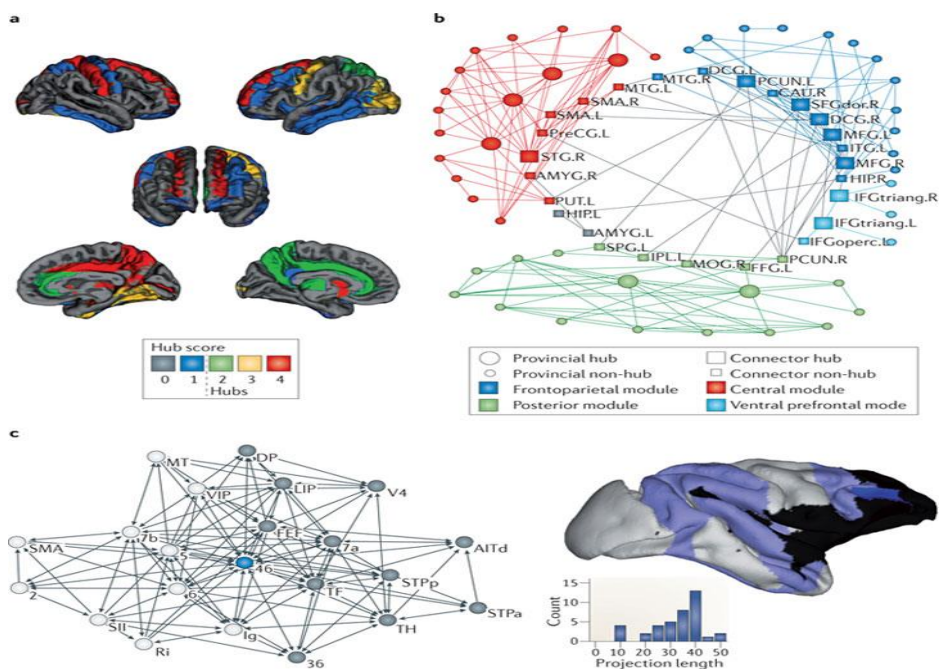
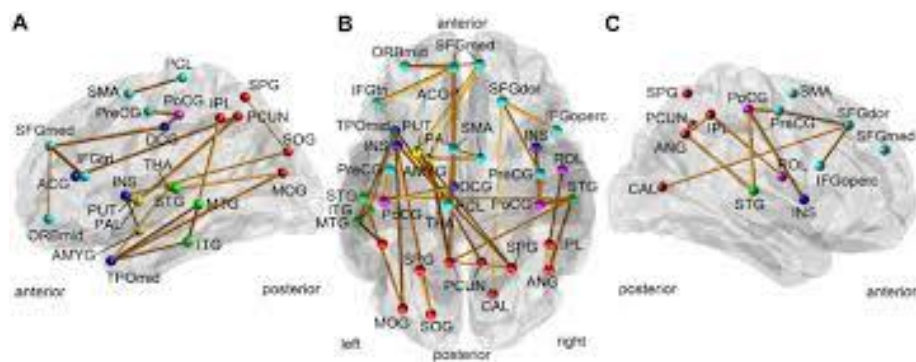
## power grid structural resilience

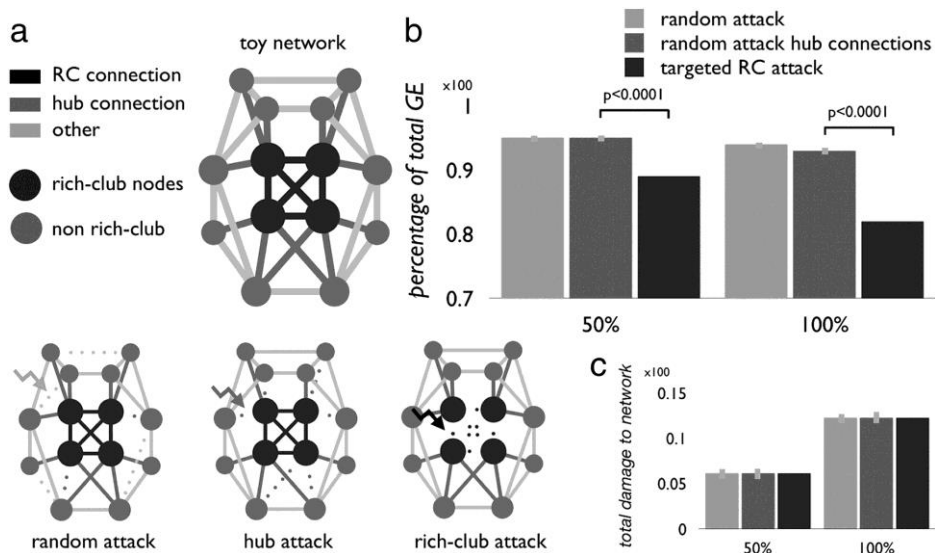
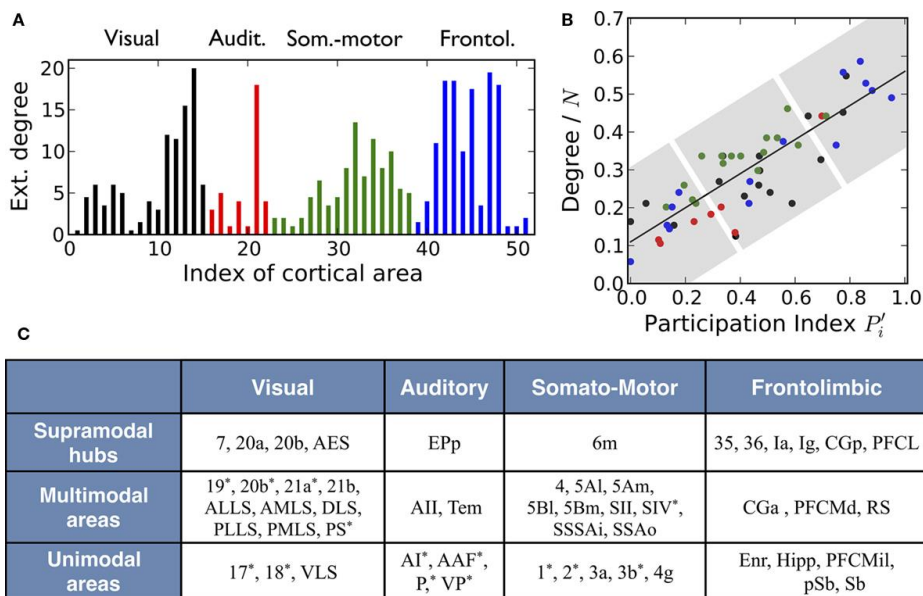
- Efficiency is impacted the most if the edge removed is the one with the highest load (e.g., links of connector hubs).



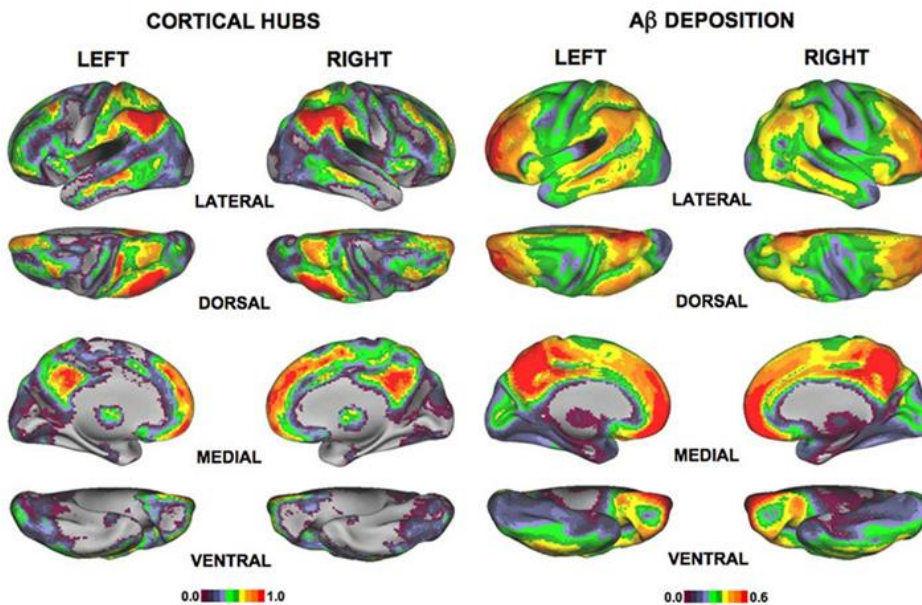
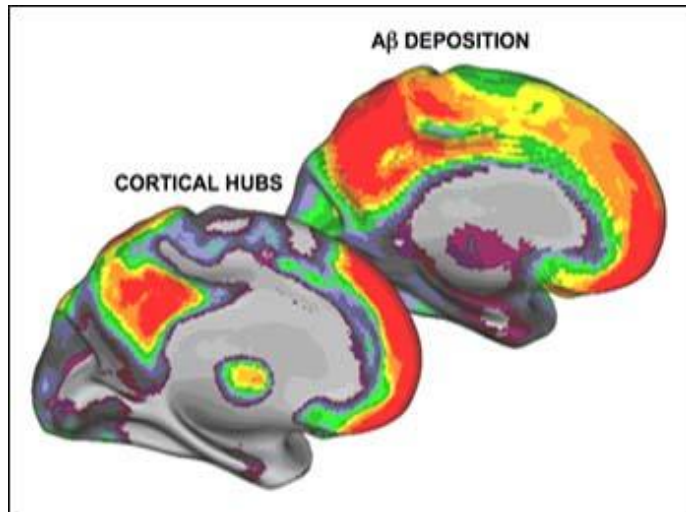
## Cortical hubs



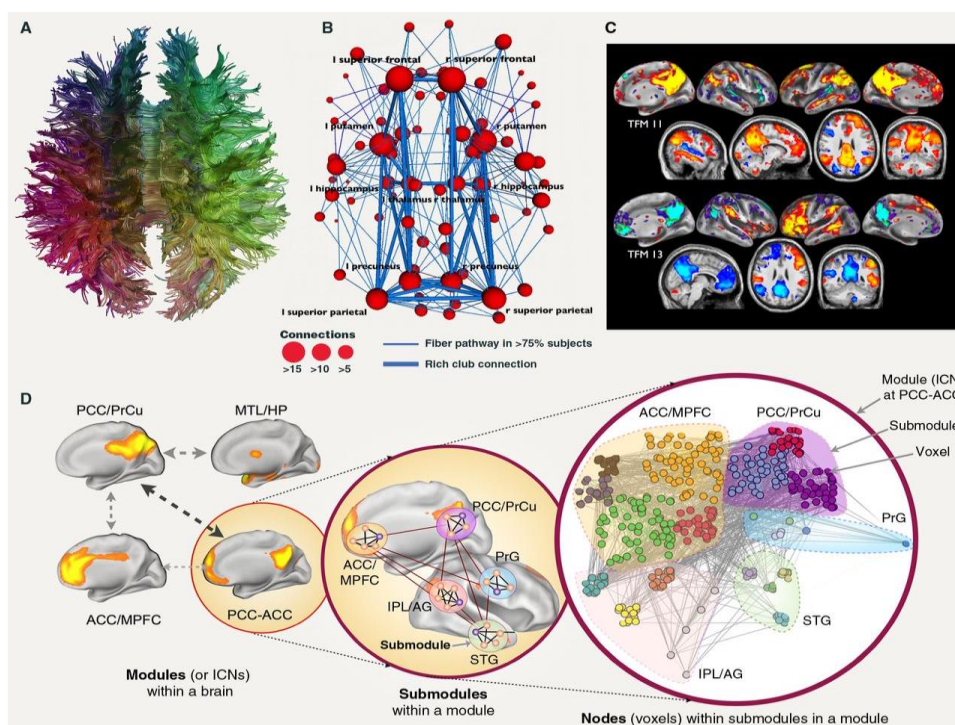
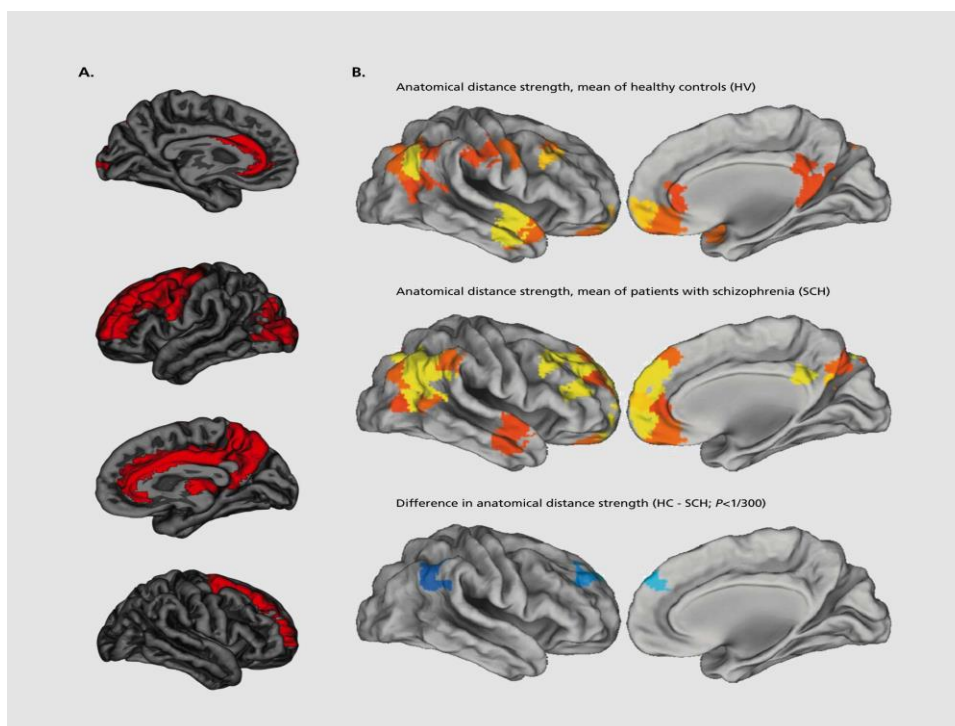


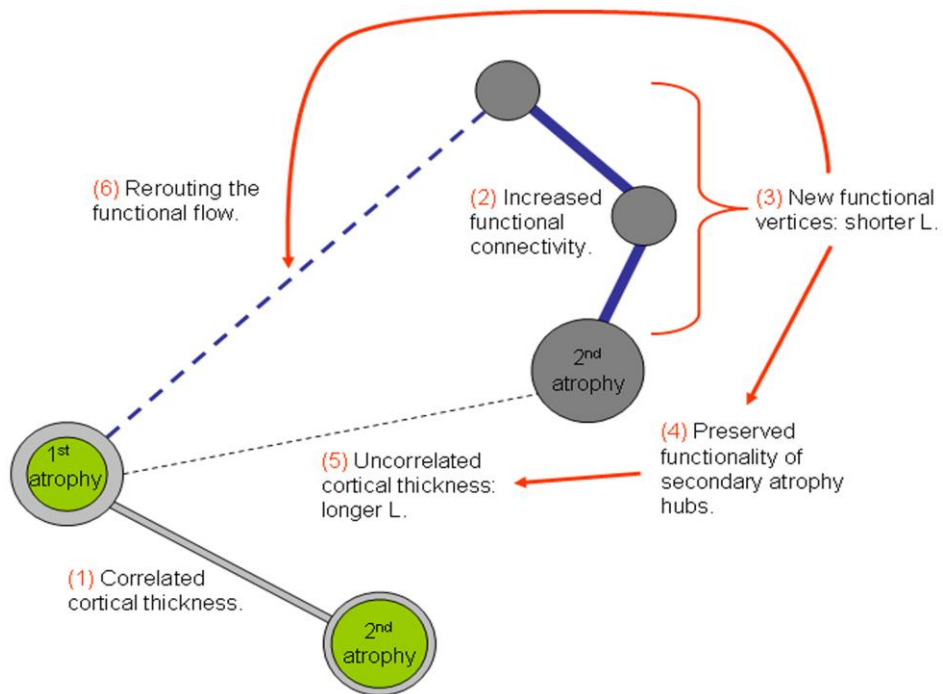


## Relationship between cortical hubs and Amyloid beta deposition?

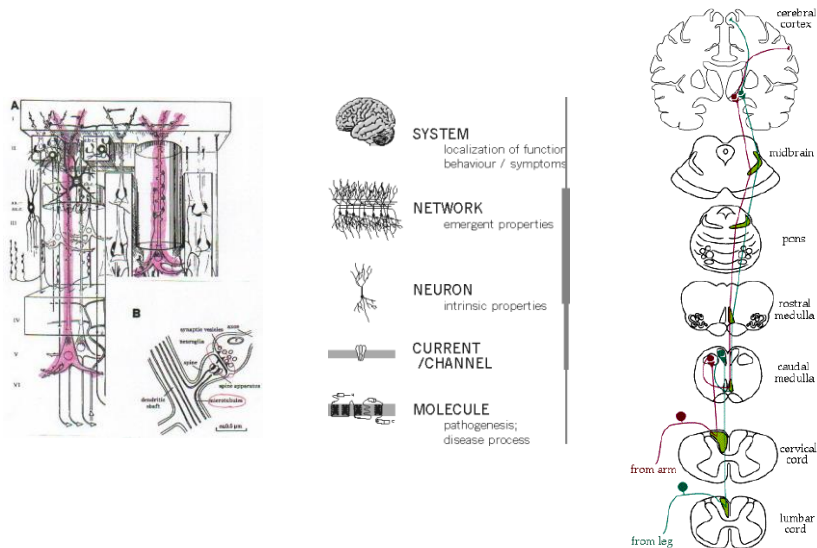








## Hierarchical structure of the brain





## Why complexity?

- Why does complexity exist in the first place, especially among biological systems? A definitive answer to this question remains elusive.
- One perspective is based on the evolutionary demands biological systems face. The evolutionary success of biological structures and organisms depends on their ability to capture information about the environment, be it molecular or ecological.
- Biological complexity may then emerge as a result of evolutionary pressure on the effective encoding of structured relationships which support differential survival.

## Why complexity?

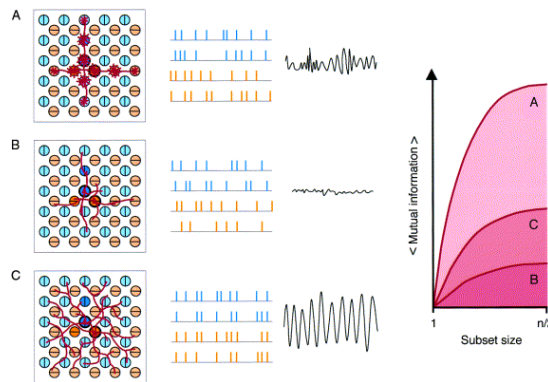
- Another clue may be found in the emerging link between complexity and network structure.
- Complexity appears very prominently in systems that combine segregated and heterogeneous components with large-scale integration.
- Such systems become more complex as they more efficiently integrate more information, that is, as they become more capable to accommodate both the existence of specialized components that generate information and the existence of structured interactions that bind these components into a coherent whole.
- Thus reconciling parts and wholes, complexity may be a necessary manifestation of a fundamental dialectic in nature (Scholapedia).

## **Functional segregation and integration**

- While the evidence for regional specialization in the brain is overwhelming, it is clear that the information conveyed by the activity of specialized groups of neurons must be functionally integrated in order to guide adaptive behavior
- Like functional specialization, functional integration occurs at multiple spatial and temporal scales.
- The rapid integration of information within the thalamocortical system does not occur in a particular location but rather in terms of a unified neural process.

## **How does the brain 'bind' together the attributes of objects to construct a unified conscious scene?**

- Neurons can integrate frequently co-occurring constellations of features by convergent connectivity. However, convergence is unlikely to be the predominant mechanism for integration.
- First, no single ('master') brain area has been identified, the activity of which represents entire perceptual or mental states.
- Second, the vast number of possible perceptual stimuli occurring in ever changing contexts greatly exceeds the number of available neuronal groups (or even single neurons), thus causing a combinatorial explosion.
- Third, convergence does not allow for dynamic ('on-the-fly') conjunctions in response to novel, previously unencountered stimuli.



- **(A)** Connections between groups are arranged such that groups with similar response selectivity are preferentially connected, are arranged anisotropically along the axis of their orientation selectivity, and connection density falls off with distance. This produces spike patterns with significant correlations between some groups and not others, as well as a temporally varying EEG that reflects a mixture of synchronization and desynchronization. **Segregation and integration** are balanced and complexity is high. **(B)** Connection density is reduced. No statistically significant correlations exist, and a flat EEG results. **(C)** Connections are of the same overall density as in (A), but are spread out uniformly and randomly over the network. The system is fully integrated but functional specialization is lost, and complexity is low.

