Brain Networks and Topology

Matilde Marcolli and Doris Tsao

Ma191b Winter 2017
Geometry of Neuroscience

References for this lecture:

- Alex Fornito, Andrew Zalesky, Edward Bullmore, *Fundamentals of Brain Network Analysis*, Elsevier, 2016
- Olaf Sporns, *Networks of the Brain*, MIT Press, 2010
- Olaf Sporns, *Discovering the Human Connectome*, MIT Press, 2012
- Fan Chung, Linyuan Lu, *Complex Graphs and Networks*, American Mathematical Society, 2004
- László Lovász, *Large Networks and Graph Limits*, American Mathematical Society, 2012

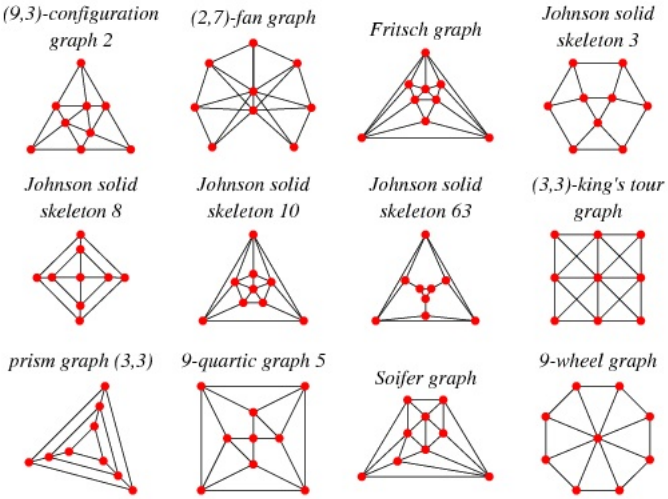
Graphs $G = (V, E, \partial)$

- $V = V(G)$ set of vertices (nodes)
- $E = E(G)$ set of edges (connections)
- boundary map $\partial : E(G) \to V(G) \times V(G)$, boundary vertices $\partial(e) = \{v, v'\}$
- directed graph (oriented edges): source and target maps

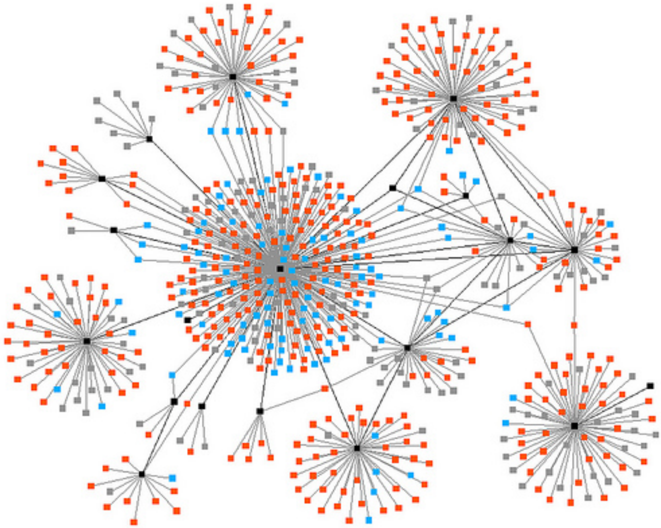
$$s : E(G) \to V(G), \quad t : E(G) \to V(G), \quad \partial(e) = \{s(e), t(e)\}$$

- *looping edge*: $s(e) = t(e)$ starts and ends at same vertex; *parallel edges*: $e \neq e'$ with $\partial(e) = \partial(e')$
- simplifying assumption: graphs G with no parallel edges and no looping edges (sometimes assume one or the other)
- additional data: label functions $f_V : V(G) \to L_V$ and $f_E : E(G) \to L_E$ to sets of vertex and edge labels L_V and L_E

Examples of Graphs

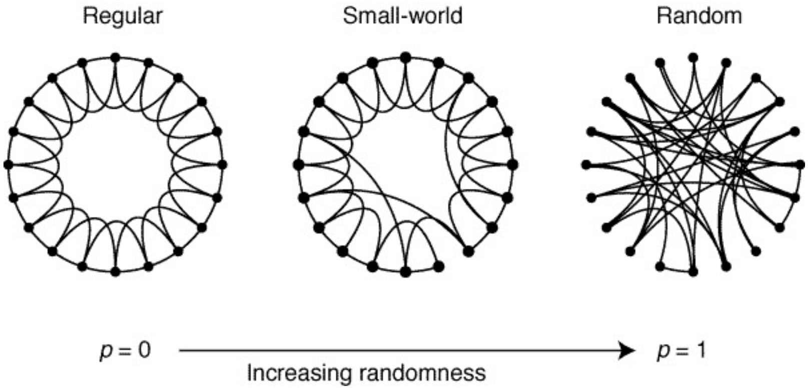


Network Graphs



(Example from Facebook)

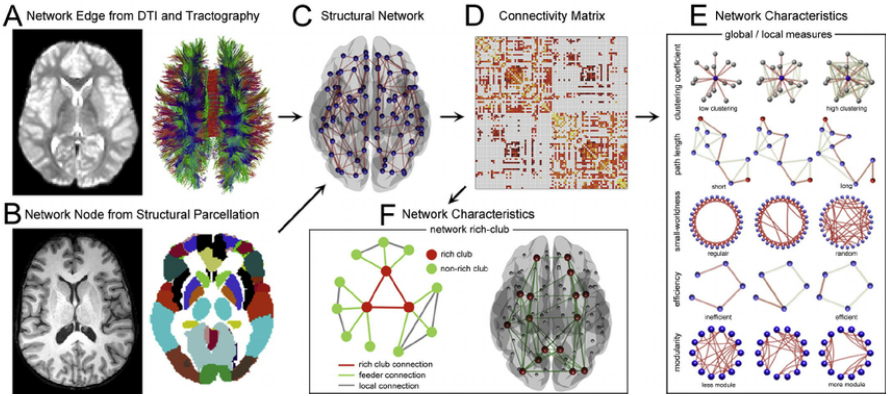
Increasing Randomness



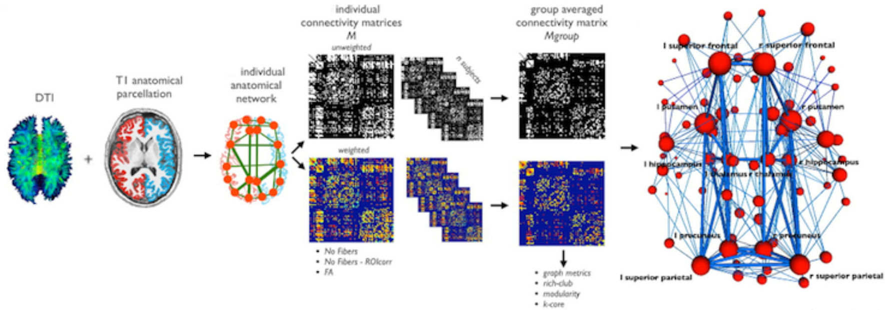
rewiring probability p

with probability p edges are disconnected and attached to a randomly chosen other vertex (Watts and Strogatz 1998)

Brain Networks: Macroscopic Scale (brain areas)

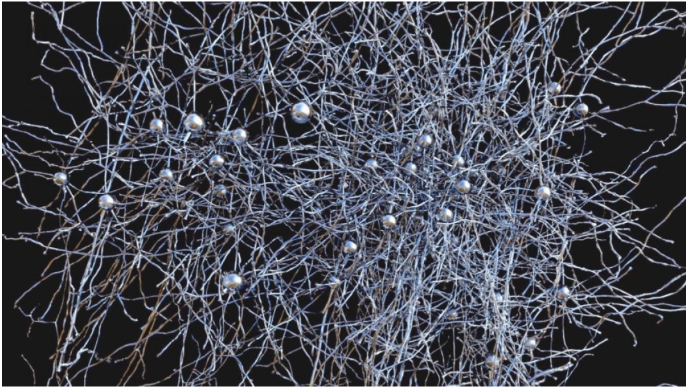


Brain Networks: Macroscopic Scale (brain areas)



Heuvel & Sporns (2011)

Brain Networks: Microscopic Scale (individual neurons)



(Clay Reid, Allen Institute; Wei-Chung Lee, Harvard Medical School; Sam Ingersoll, graphic artist; largest mapped network of individual cortical neurons, 2016)

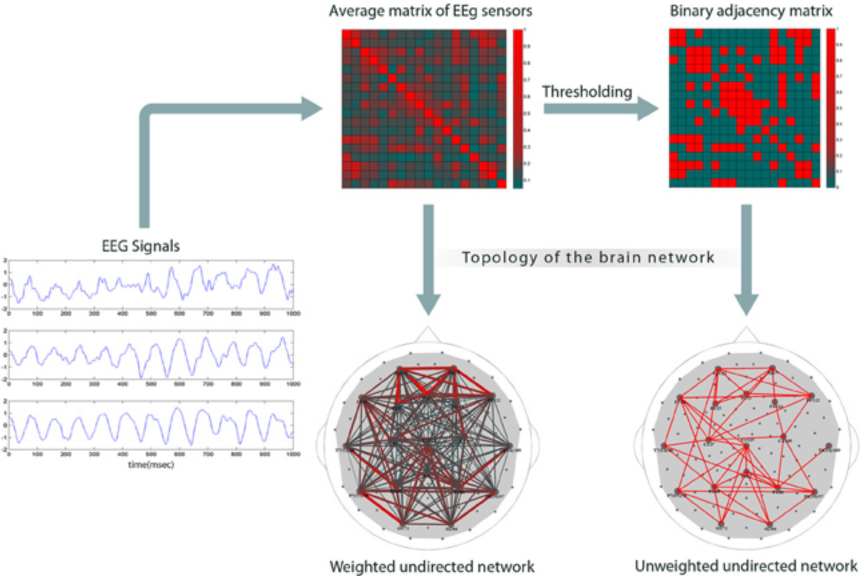
Modeling Brain Networks with Graphs

1. Spatial embeddings (embedded graphs $G \subset S^3$, knotting and linking, topological invariants of embedded graphs)
2. Vertex labels (heterogeneity of node types): distinguish different kinds of neurons/different areas
3. Edge labels (heterogeneity of edge types)
4. Orientations (directionality of connections): directed graphs
5. Weights (connection strengths)
6. Dynamical changes of network topology

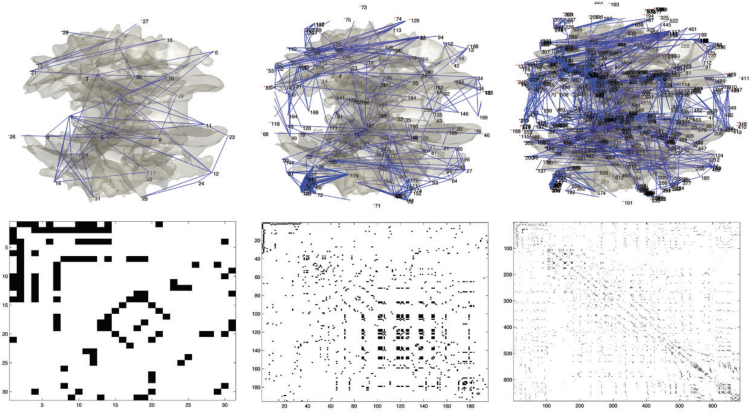
Connectivity and Adjacency Matrix

- connectivity matrix $C = (C_{ij})$ matrix size $N \times N$ with $N = \#V(G)$, with $C_{ij} \in \mathbb{R}$ connectivity strength for oriented edge from v_i to v_j
- sign of C_{ij}: excitatory/inhibitory connection
- $C_{ij} = 0$ no oriented connecting edges between these vertices
- in general $C_{ij} \neq C_{ji}$ for directed graphs, while $C_{ij} = C_{ji}$ for non-oriented
- can use $C_{ij} \in \mathbb{Z}$ for counting multiple parallel edges
- $C_{ii} = 0$ if no looping edges
- adjacency matrix $A = (A_{ij})$ also $N \times N$ with $A_{ij} = 1$ if there is (at least) an edge from v_i to v_j and zero otherwise
- $A_{ij} = 1$ if $C_{ij} \neq 0$ and $A_{ij} = 0$ if $C_{ij} = 0$
- if no parallel (oriented) edges: can reconstruct G from $A = (A_{ij})$ matrix

Connectivity and Adjacency Matrix



Filtering the Connectivity Matrix



various methods, for example pruning weaker connections: threshold

• connection density

$$\kappa = \frac{\sum_{ij} A_{ij}}{N(N-1)}$$

density of edges over choices of pairs of vertices

• total weight $W^{\pm} = \frac{1}{2}\sum_{ij} w_{ij}^{\pm}$ (for instance strength of connection positive/negative $C_{ij}^{\pm}$)

• how connectivity varies across nodes: valence of vertices (node degree), distribution of values of vertex valence over graph (e.g. most vertices with few connections, a few hubs with many connections: airplane travel, math collaborations)

• in/out degree $\iota(v) = \#\{e : v \in \partial(e)\}$ vertex valence; for oriented graph in-degree $\iota^+(v) = \#\{e : t(e) = v\}$ and out-degree $\iota^-(v) = \#\{e : s(e) = v\}$

$$\#E = \sum_v \iota^+(v) = \sum_v \iota^-(v)$$

• mean in/out degree

$\langle \iota^+ \rangle = \frac{1}{N}\sum_v \iota^+(v) = \frac{\#E}{N} = \frac{1}{N}\sum_v \iota^-(v) = \langle \iota^- \rangle$

Degree Distribution

- $\mathbb{P}(\deg(v) = k)$ fraction of vertices (nodes) of valence (degree) k

Erdös–Rényi graphs: generate random graphs by connecting vertices randomly with equal probability p: all graphs with N vertices and M edges have equal probability

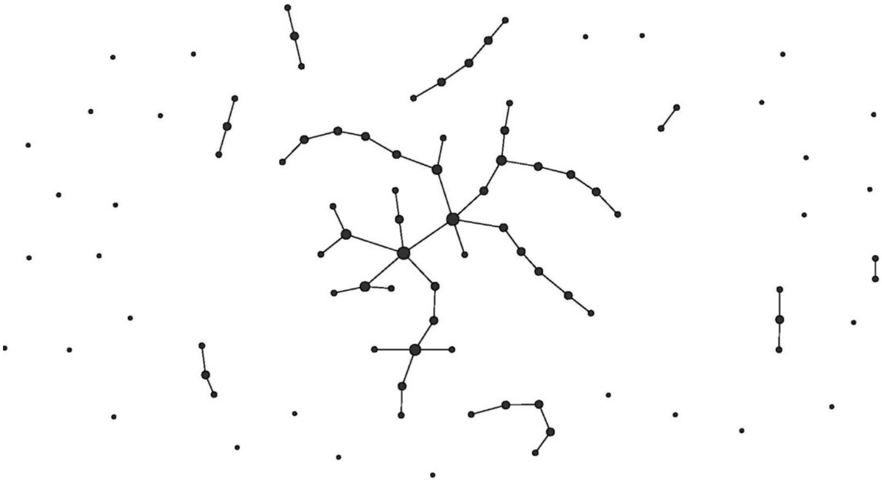
$$p^M (1-p)^{\binom{N}{2} - M}$$

- for Erdös–Rényi graphs degree distribution

$$\mathbb{P}(\deg(v) = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

second exponent $N - 1 - k$ remaining possible connection from a chosen vertex (no looping edges) after removing a choice of k edges

- $p =$ connection density of the graph (network)



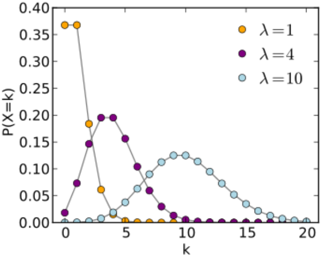
An Erdös–Rényi graph generated with $p = 0.001$

- the Erdös–Rényi degree distribution satisfies for $n \to \infty$

$$\mathbb{P}(\deg(v) = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \sim \frac{(np)^k e^{-np}}{k!}$$

- so for large n the distribution is Poisson

$$\mathbb{P}(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$



- .... but Erdös–Rényi graphs not a good model for brain networks

Scale-free networks ... power laws

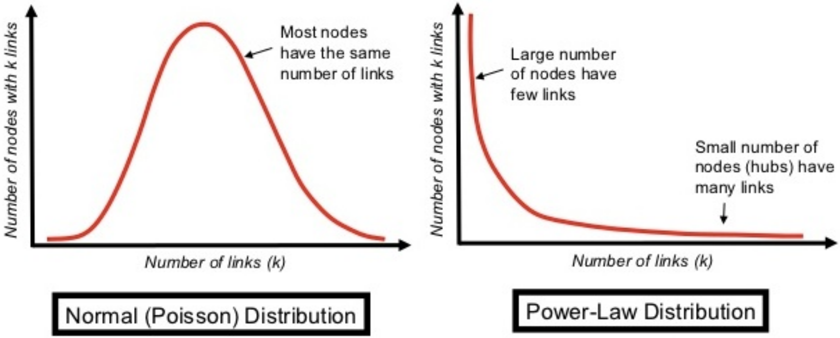
$$\mathbb{P}(\deg(v) = k) \sim k^{-\gamma} \qquad \text{for some } \gamma > 0$$

• slower decay rate than in binomial case: fat tail ... higher probability than in Erdös–Rényi case of highly connected large k nodes

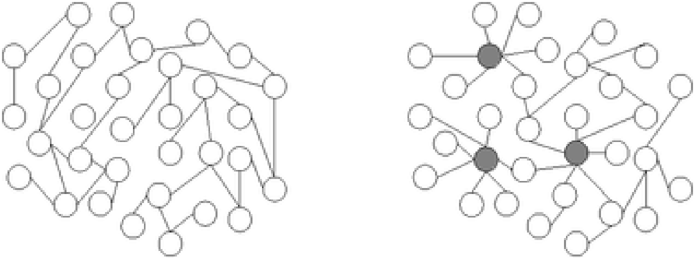
• Erdös–Rényi case has a peak in the distribution: a characteristic scale of the network

• power law distribution has no peak: no characteristic scale... scale free (typical behavior of self-similar and fractal systems)

Poisson versus Power Law degree distributions



(nodes = vertices, links = edges, number of links = valence)



(a) Random network (b) Scale-free network

Broad Scale Networks

- intermediate class: more realistic to model brain networks
- exponentially truncated power law

$$\mathbb{P}(\deg(v) = k) \sim k^{-\gamma} e^{-k/k_c}$$

- cutoff degree k_c: for small k_c quicker transition to an exponential distribution
- range of scales over which power law behavior is dominant
- so far measurements of human and animal brain networks consistent with scale free and broad scale networks

For weighted vertices with weights $w \in \mathbb{R}_+^*$

• weight distribution: best fitting for brain networks log-normal distribution

$$\mathbb{P}(\text{weight}(v) = w) = \frac{1}{w\sigma\sqrt{2\pi}} \exp\left(\frac{-(\log w - \mu)^2}{2\sigma^2}\right)$$

Gaussian in log coordimates

• why log-normal? model based on properties:

1. geometry of embedded graph with distribution of interregional distances $\sim$ Gaussian
2. distance dependent cost of long distance connections

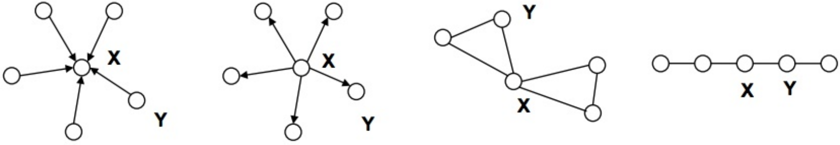
drop in probability of long distance connections with strong weights

Centrality

- a node is more "central" to a network the more
  - it is highly connected (large valence) – degree
  - it is located on the shortest path between other nodes – betweenness
  - it is close to a large number of other nodes (eg via highly connected neighbors) – closeness

- valence $\deg(v)$ is a measure of centrality (but not so good because it does not distinguish between highly or sparsely connected neighbors)

In each of the following networks, X has higher centrality than Y according to a particular measure



indegree outdegree betweenness closeness

Perron–Frobenius centrality

- Perron–Frobenius theorem (version for non-negative matrices)
  - $A = (A_{ij})$ non-negative $N \times N$ matrix: $A_{ij} \geq 0, \forall i, j$
  - A is primitive if $\exists k \in \mathbb{N}$ such that A^k is positive
  - A irreducible iff $\forall i, j, \exists k \in \mathbb{N}$ such that $A^k_{ij} > 0$ (implies $I + A$ primitive)
  - Directed graph G_A with N vertices and edge from v_i to v_j iff $A_{ij} > 0$: matrix A irreducible iff G_A strongly connected (every vertex is reachable through an oriented path from every other vertex)
  - Period h_A: greatest common divisor of lengths of all closed directed paths in G_A

Assume A non-negative and irreducible with period h_A and spectral radius ρ_A, then:

1. $\rho_A > 0$ and eigenvalue of A (Perron–Frobenius eigenvalue); simple
2. Left eigenvector V_A and right eigenvector W_A with all positive components (Perron–Frobenius eigenvector): only eigenvectors with all positive components
3. h_A complex eigenvectors with eigenvalues on circle $|\lambda| = \rho_A$
4. spectrum invariant under multiplication by $e^{2\pi i/h_A}$

Take $A =$ adjacency matrix of graph G

• $A =$ adjacency matrix of graph G

• vertex $v = v_i$: PF centrality

$$\mathcal{C}_{PF}(v_i) = V_{A,i} = \frac{1}{\rho_A} \sum_j A_{ij} V_{A,j}$$

ith component of PF eigenvector V_A

• high centrality if high degree (many neighbors), neighbors of high degree, or both

• can use V_A or W_A, left/right PF eigenvectors: centrality according to in-degree or out-degree

Page Rank Centrality (google)

- $D =$ diagonal matrix $D_{ii} = \max\{\deg(v_i)^{out}, 1\}$
- α, β adjustable parameters

$$\mathcal{C}_{PR}(v_i) = ((I - \alpha A D^{-1})^{-1} \beta \mathbf{1})_i$$

with $\mathbf{1}$ vector of N entries 1

- this scales contributions of neighbors of node v_i by their degree: dampens potential bias of nodes connected to nodes of high degree

Delta Centrality

• measure of how much a topological property of the graph changes if a vertex is removed

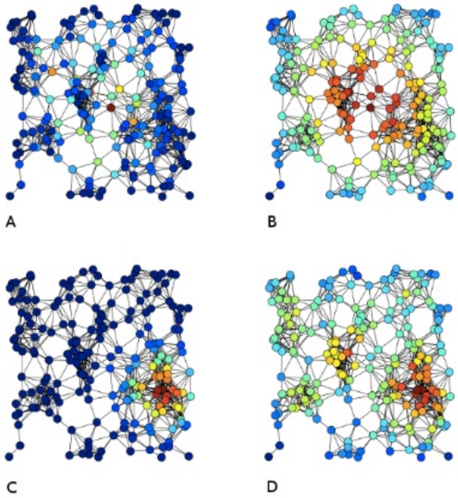
• graph G and vertex $v \in V(G)$: remove v and star $S(v)$ of edges adjacent to v

$$G_v = G \smallsetminus S(v)$$

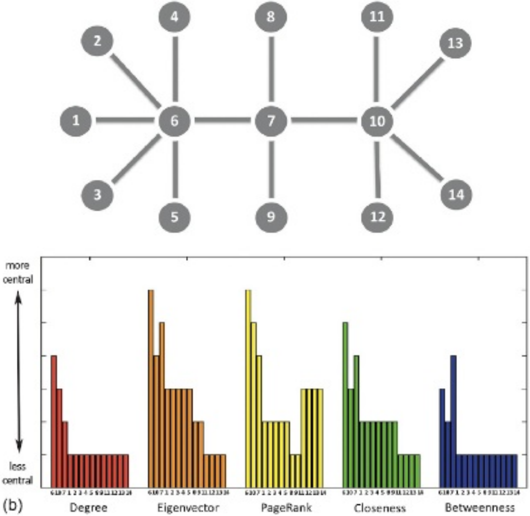
• topological invariants of (embedded) graph $M(G)$ (with integer or real values)

• delta centrality with respect to M

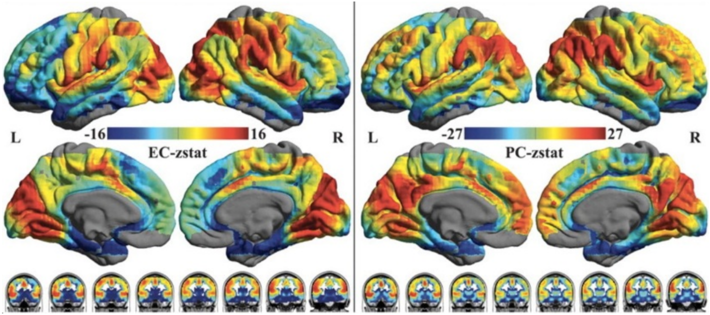
$$\mathcal{C}_M(v) = \frac{M(G) - M(G_v)}{M(G)}$$



(A) betweenness; (B) closeness; (C) eigenvector (PF); (D) degree


1
2
3
4
5
6
7
8
9
10
11
12
13
14
more central
less central
(b)
Degree
Eigenvector
PageRank
Closeness
Betweenness

Eigenvalue and PageRank centrality in brain networks



X.N.Zuo, R.Ehmke, M.Mennes, D.Imperati, F.X.Castellanos, O.Sporns, M.P.Milham, *Network Centrality in the Human Functional Connectome*, Cereb Cortex (2012) 22 (8): 1862-1875.

Connected Components

• what is the right notion of "connectedness" for a large graph? small components breaking off should not matter, but large components becoming separated should

• is there one large component?

• Erdös–Rényi graphs: size of largest component ($N = \#V(G)$)

- sharp increase at $p \sim 1/N$
- graph tends to be connected for $p > \frac{\log N}{N}$
- for $p < \frac{1}{N}$ fragmented graph: many connected components of comparable (small) size
- for $\frac{1}{N} \leq p \leq \frac{\log N}{N}$ emergence of one giant component; other components still exist of smaller size

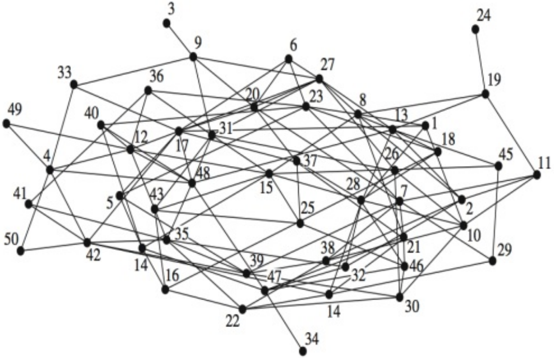


Figure: Emergence of connectedness: a random network on 50 nodes with $p = 0.10$.

(from Daron Acemoglu and Asu Ozdaglar, Lecture Notes on Networks)

How to prove the emergence of connectedness?

(Argument from Daron Acemoglu and Asu Ozdaglar, Lecture Notes on Networks)

• threshold function $\tau(N)$ for a property $\mathcal{P}(G)$ of a random graph G with $N = \#V(G)$, with probability $p = p(N)$:

$$\mathbb{P}(\mathcal{P}(G)) \to 0 \quad \text{when } \frac{p(N)}{\tau(N)} \to 0$$

$$\mathbb{P}(\mathcal{P}(G)) \to 1 \quad \text{when } \frac{p(N)}{\tau(N)} \to \infty$$

with $\mathbb{P}(\mathcal{P}(G))$ probability that the property is satisfied

• show that $\tau(N) = \frac{\log N}{N}$ is a threshold function for the property $\mathcal{P}$ =connectedness

• for $\mathcal{P}$ =connectedness show that for $p(N) = \lambda \frac{\log N}{N}$:

$$\mathbb{P}(\mathcal{P}(G)) \to 1 \quad \text{for } \lambda > 1$$

$$\mathbb{P}(\mathcal{P}(G)) \to 0 \quad \text{for } \lambda < 1$$

• to prove graph disconnected for $\lambda < 1$ show growing number of single node components

• in an Erdös–Rényi graph probability of a given node being a connected component is $(1-p)^{N-1}$; so typical number of single node components is $N \cdot (1-p)^{N-1}$

• for large N this $(1-p)^{N-1} \sim e^{-pN}$

• if $p = p(N) = \lambda \frac{\log N}{N}$ this gives

$$e^{-p(N)\,N} = e^{-\lambda \log N} = N^{-\lambda}$$

• for $\lambda < 1$ typical number of single node components

$$N \cdot (1-p)^{N-1} \sim N \cdot N^{-\lambda} \to \infty$$

• for $\lambda > 1$ typical number of single vertex components goes to zero, but not enough to know graph becomes connected (larger size components may remain)

• probability of a set S_k of k vertices having no connection to the rest of the graph (but possible connections between them) is $(1-p)^{k(N-k)}$

• typical number of sets of k nodes not connected to the rest of the graph

$$\binom{N}{k}(1-p)^{k(N-k)}$$

• Stirling's formula $k! \sim k^k e^{-k}$ gives for large N and k

$$\binom{N}{k}(1-p)^{k(N-k)} \sim (\frac{N}{k})^k e^k e^{-k\lambda \log N} = N^k N^{-\lambda k} e^{k(1-\log k)} \to 0$$

for $p = p(N) = \lambda \frac{\log N}{N}$ with $\lambda > 1$

Phase transitions:

- at $p = \frac{\log N}{N}$: graph becomes connected
- at $p = \frac{1}{N}$: emergence of one giant component

• use similar method: threshold function $\tau(N) = \frac{1}{N}$ and probabilities $p(N) = \frac{\lambda}{N}$ with either $\lambda > 1$ or $\lambda < 1$

Case $\lambda < 1$:

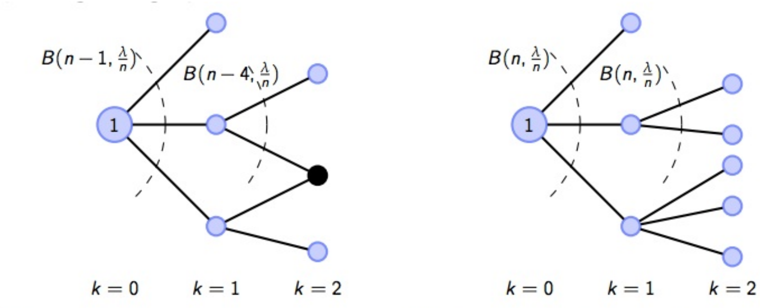
• starting at a vertex approximate counting of connected vertices in an Erdös–Rényi graph with a branching process $B(N, \frac{\lambda}{N})$

• replaces graph by a tree (overcounting of vertices) with typical number of descendants $N \times \frac{\lambda}{N}$ so in k steps from starting vertex expected number of connections λ^k

• so typical size of the component connected to first vertex is bounded above by size obtained from branching process

$$\sum_k \lambda^k = \frac{1}{1-\lambda}$$

small sized components



(a) Erdos-Renyi graph process. (b) Branching Process Approx.

Branching process approximation to an Erdös–Rényi graph process (from Daron Acemoglu and Asu Ozdaglar, Lecture Notes on Networks)

Case $\lambda > 1$:

• process $B(N, \frac{\lambda}{N})$ asymptotically Poisson with probability (k steps)

$$e^{-\lambda} \frac{\lambda^k}{k!}$$

• probability ρ that tree obtained via this process is finite: recursive structure (overall tree finite if each tree starting from next vertices finite)

$$\rho = \sum_k e^{-\lambda} \frac{\lambda^k}{k!} \rho^k$$

fixed point equation $\rho = e^{\lambda(\rho-1)}$

• one solution $\rho = 1$ but another solution inside interval $0 < \rho < 1$

• however... the branching process $B(n, p)$ produces trees, but on the graph G fewer vertices...

• after δN vertices have been added to a component via the branching process starting from one vertex, to continue the process one has $B(N(1-\delta), p)$ correspondingly changing $\lambda \mapsto \lambda(1-\delta)$ (to continue to approximate same p of Erdös–Rényi process)

• progressively decreases $\lambda(1-\delta)$ as δ increases so branching process becomes more likely to stop quickly

• typical size of the big component becomes $(1-\rho)N$ where $\rho = \rho(\lambda)$ as above probability of finite tree, solution of $\rho = e^{\lambda(\rho-1)}$

Robustness to lesions

• Remove a vertex $v \in V(G)$ with its star of edges $S(v) = \{e \in E(G) : v \in \partial(e)\}$

$$G_v = G \smallsetminus S(v)$$

measure change: e.g. change in size of largest component

• when keep removing more $S(v)$'s at various vertices, reach a threshold at which graph becomes fragmented

• in opposite way, keep adding edges with a certain probability, critical threshold where giant component arises, as discussed previously

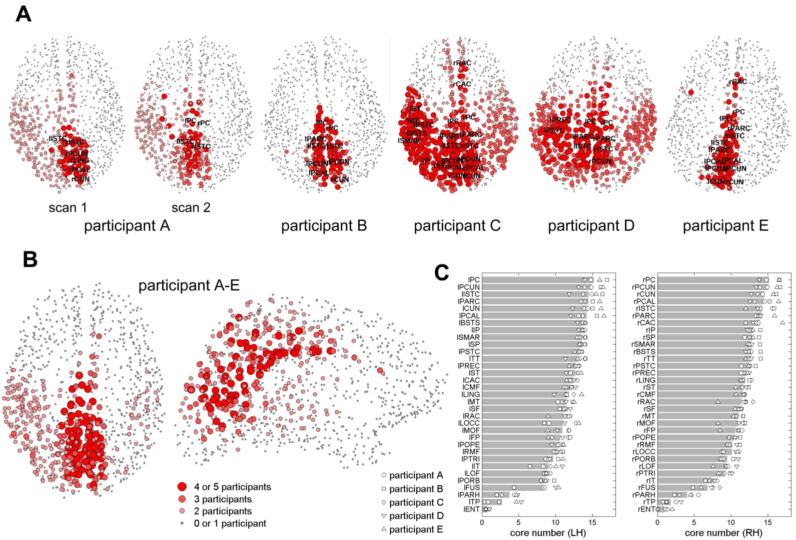
Core/Periphery

- *core*: subset of vertices highly connected to each other (hubs)
- *periphery*: nodes connected to core vertices but not with each other
- *maximal cliques*: maximally connected subsets of nodes

k-core decomposition: remove all $S(v)$ with $\deg(v) < k$, remaining graph $G^{(k)}$ *k-core*

- *core index* of $v \in V(G)$: largest k such that $v \in V(G^{(k)})$

s-core $G^{(s)}$: remove all $S(v)$ of vertices with weight $w(v) < s$



k-core decomposition, from P.Hagmann, L.Cammoun, X.Gigandet, R.Meuli, C.J.Honey, V.J.Wedeen, O.Sporns, "Mapping the Structural Core of Human Cerebral Cortex", PLOS bio 2008

Topology and flow of information along directed graphs

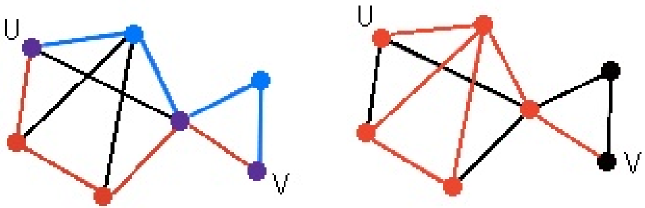
• networks with characteristic path length (close to min in a random graph)

• path length = min number of oriented edges or minimum total weight of these edges

walks, trails, paths

- *walk*: sequence of oriented edges (target of one source of next): revisiting vertices and edges allowed
- *trail*: same but no edge repetitions (edges visited only once)
- *path*: no edge and no vertex repetitions (both vertices and edges visited only once)

Warning: terminology varies in the literature



two paths from U to V and a trail from U to V

- shortest path = geodesics (with edge weights as metric)
- average of path length over all shortest path in the graph
  - at vertex v_i average ℓ_i of lengths $\ell_{ij} = \ell(v_i, v_j)$ of shortest paths starting at v_i (and ending at any other vertex v_j)
  - average over vertices

$$L = \frac{1}{N} \sum_i \ell_i = \frac{1}{N(N-1)} \sum_{i \neq j} \ell_{ij}$$

• main idea: brain networks with the shortest average path length integrate information better

• in case of a graph with several connected components, for v_i and v_j not in the same component usually take $\ell_{ij} = \infty$, then better to use harmonic mean

$$N(N-1)\left(\sum_{i\neq j} \ell_{ij}^{-1}\right)^{-1}$$

or its inverse, the global efficiency:

$$\frac{1}{N(N-1)}\sum_{i\neq j} \ell_{ij}^{-1}$$

The Graph Laplacian measuring flow of information in a graph

- $\delta = (\delta_{ij})$ diagonal matrix of valencies $\delta_{ii} = \deg(v_i)$
- adjacency matrix A (weighted with w_{ij})
- Laplacian $\Delta_G = \delta - A$
- normalized Laplacian $\hat{\Delta}_G = I - A\delta^{-1}$ or symmetrized form $\hat{\Delta}^s_G = I - \delta^{-1/2} A \delta^{-1/2}$
- w_{ij}/δ_{ii} = probability of reaching vertex v_j after v_i along a random walk search
- dimension of $\mathrm{Ker}\hat{\Delta}_G$ counts connected components
- eigenvalues and eigenvectors give decomposition of the graph into "modules"

Dynamics

• in brain networks fast dynamics $\sim$ 100 millisecond timescale: variability in functional coupling of neurons and brain regions, underlying functional anatomy unchanged; also slow dynamics: long lasting changes in neuron interaction due to plasticity... growth, rewiring

• types of dynamics: diffusion processes, random walks, synchronization, information flow, energy flow

• topology of the graph plays a role in the spontaneous emergence of global (collective) dynamical states

• multiscale dynamics: dynamics at a scale influenced by states and dynamics on larger and smaller scales

• mean field model: dynamics at a scale averaging over effects at all smaller scales

• synchronization of a system of coupled oscillators at vertices with coupling along edges

• if graph very regular (eg lattice) difficult to achieve synchronized states

• small-world networks are easier to synchronize

• fully random graphs synchronize more easily but synchronized state also very easily disrupted: difficult to maintain synchronization

• more complex types of behavior when non-identical oscillators (different weights at different vertices) and additional presence of noise

• still open question: what is the role of topology and topology changes in the graph in supporting self-organized-criticality