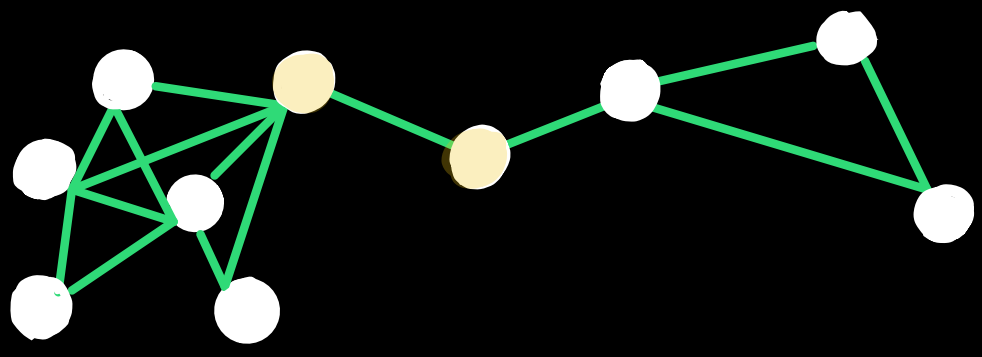
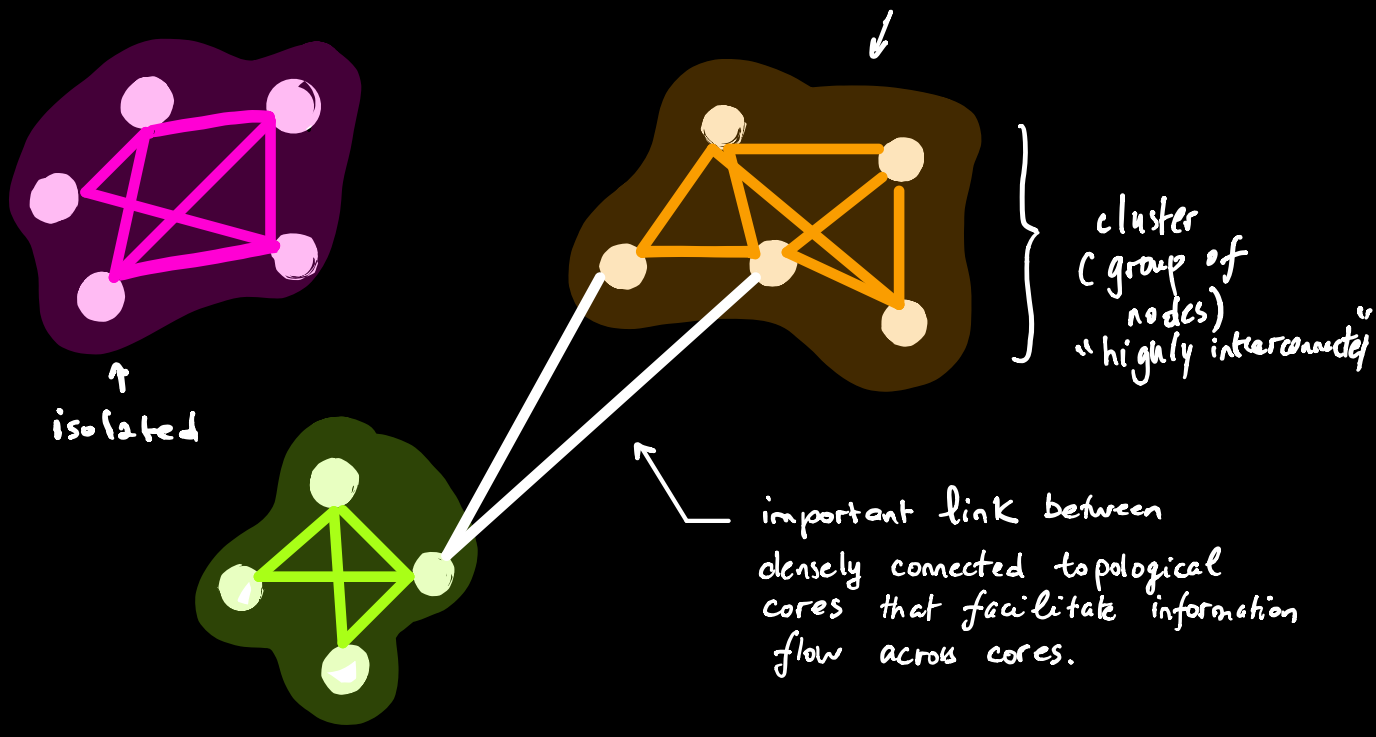
# $GraphT_3$



1. TOPOLOGICAL CENTRALITY IS A MULTIFACETED CONCEPT → THERE ARE MANY GRAPH THEORETIC MEASURES TO QUANTIFY THE CENTRALITY OF A NODE.

2. DIFFERENT MEASURES OF CENTRALITY MAKE DIFFERENT ASSUMPTIONS ABOUT HOW INFORMATION FLOWS ON THE NETWORK.

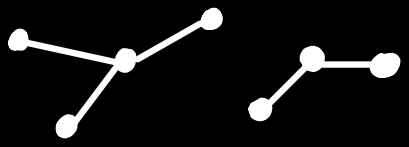
3. MEASURES BASED ON THE SHORTEST PATHS ASSUME THAT INFORMATION IS ROUTED ALONG THE SHORTEST PATH.



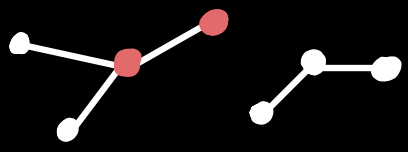

"what if we extend the idea of centrality to consider collections of nodes and edges which play a central role in network organization and dynamics?

# GraphT$_3$

① Broad scale
CONNECTED COMPONENTS IN A GRAPH



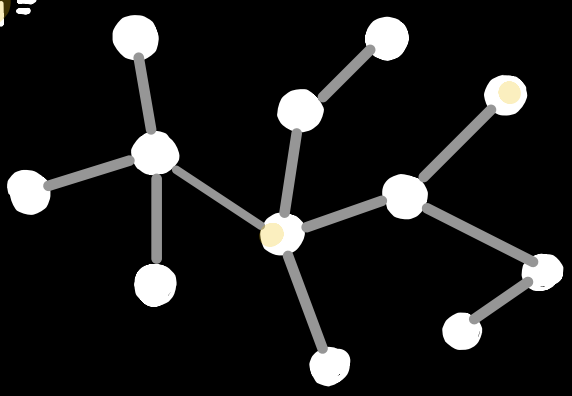
② PERCOLATION AND GRAGH ROBUSTNESS TO ATTACKS



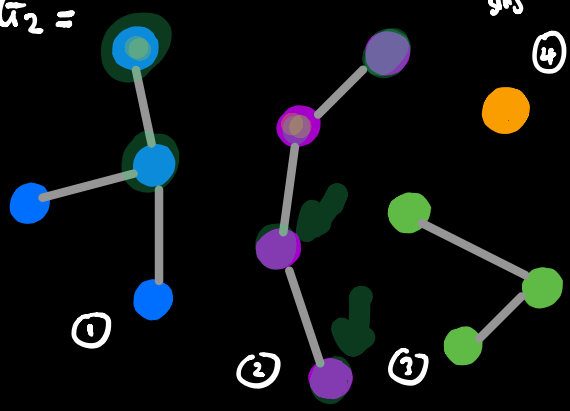
## CONNECTED COMPONENTS

* Is all the constituent nodes are interconnected such that they form a single component?
* A node-connected graph is one in which a path can be traced between any pair of nodes by traversing the edges of the graph.

$G_1$ =



$G_2$ =




Graphs containing subsets of nodes that cannot be linked by a traversable path are called fragmented or disconnected.

→ each subset is called a connected component.

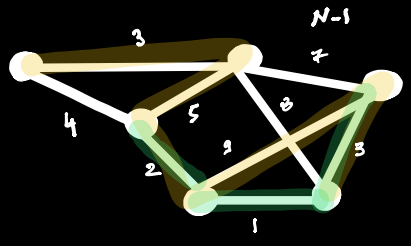
A singleton is a connected component with a zero degree node.

**Definition** MSP = minimum spanning tree

The minimum set of edges required to form a node connected network.

- **Binary graph** → MST is a subgraph comprising the minimum number of edges that ensures a path can be found between all pairs of nodes.

  Since we require (N-1) edges to connect N nodes, the MST will always comprise (N-1) edges.




- **Weighted graph** → Each graph is assigned a distance penalty (a function of its weight).

  MST is a subset of edges that minimizes the total sum of these penalty values subject to a path existing between all node pairs.

→ MST : foundational backbone of a graph.

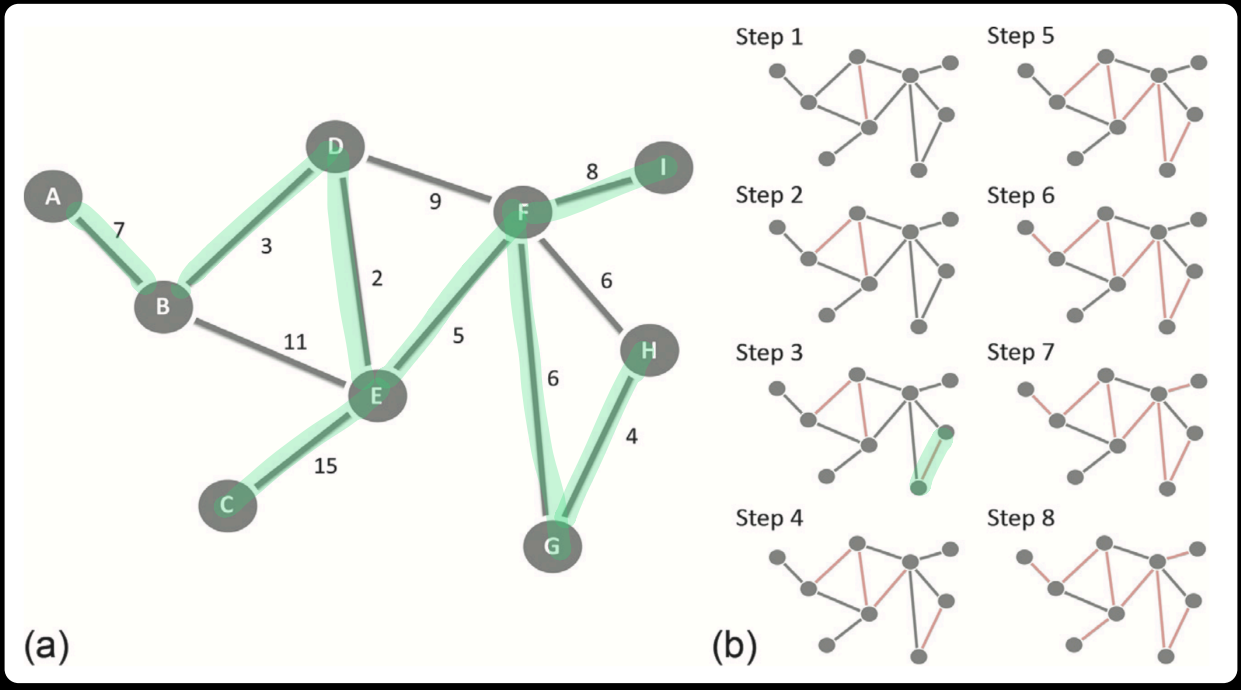
"Think of an algorithm to find the MSP in a weighted graph"!

---

## MST KRUSKAL'S ALGO

**Step 1** The lowest weighted edge is selected first

**Step 2** Add edges in ascending order without forming any cycles.




[Chapter 6 - FBNA]

**Giant Component** is a connected component that contains a large proportion of the total number of nodes, and whose size grows in proportion to N.

## COMPONENTS IN UNDIRECTED GRAPHS

- A component of an undirected graph is a subgraph of vertices in which each node can be linked to every other node via one or more paths.
- The size of a CC can be measured in terms of the number of nodes or edges it contains.
- When the size of a CC is N → G is node-connected.
- When the size of the largest component < N → G is fragmented.

" Think of an algorithm to identify the connected components in a graph. "

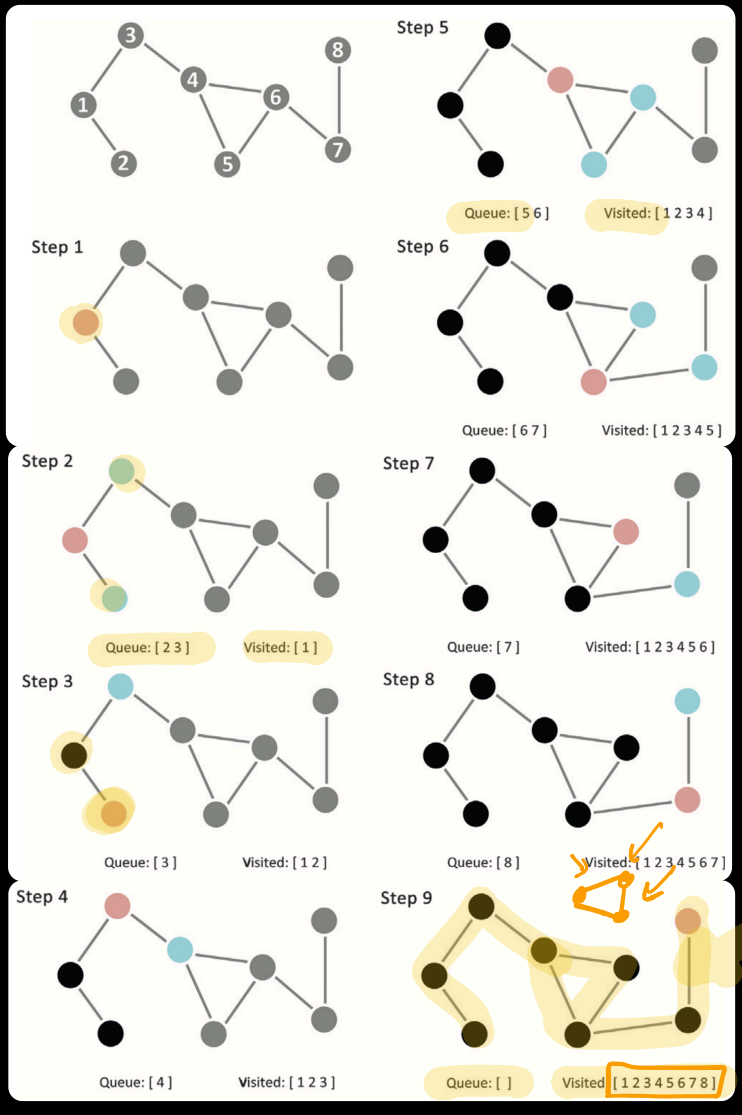
# Breadth-first search (BFS) algo

- Iterative technique of visiting all connected components in G.

**Step 1** Select an arbitrary (index) node in G.

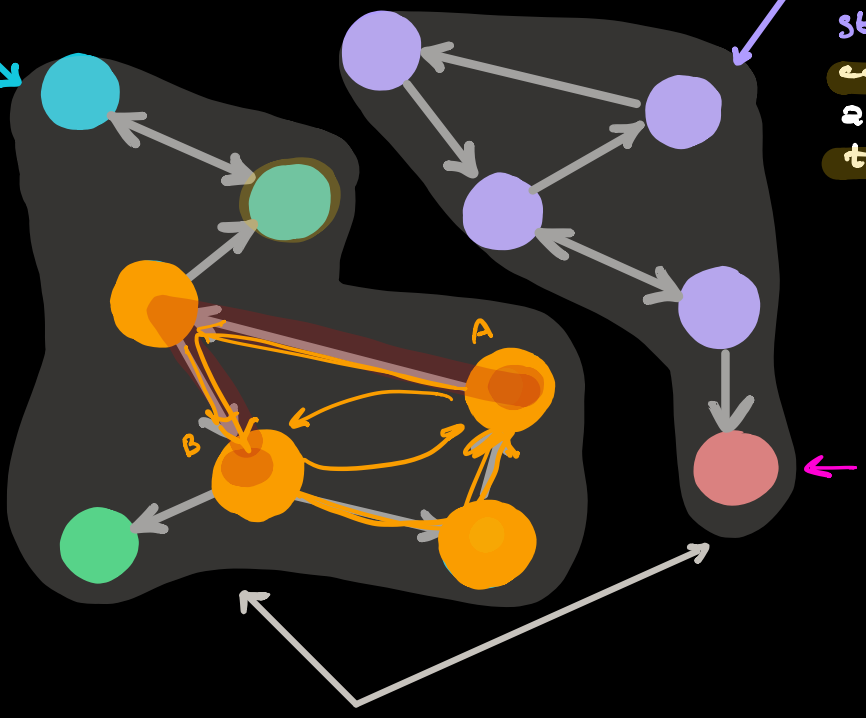
**Step 2**
- Mark the node as visited and place its neighbors in a queue Q.
- The enqueued nodes are sequentially selected as the next working nodes.

**Step 3** Repeat until all nodes in the components have been visited.



## WEAKLY & STRONGLY CC in DIRECTED G

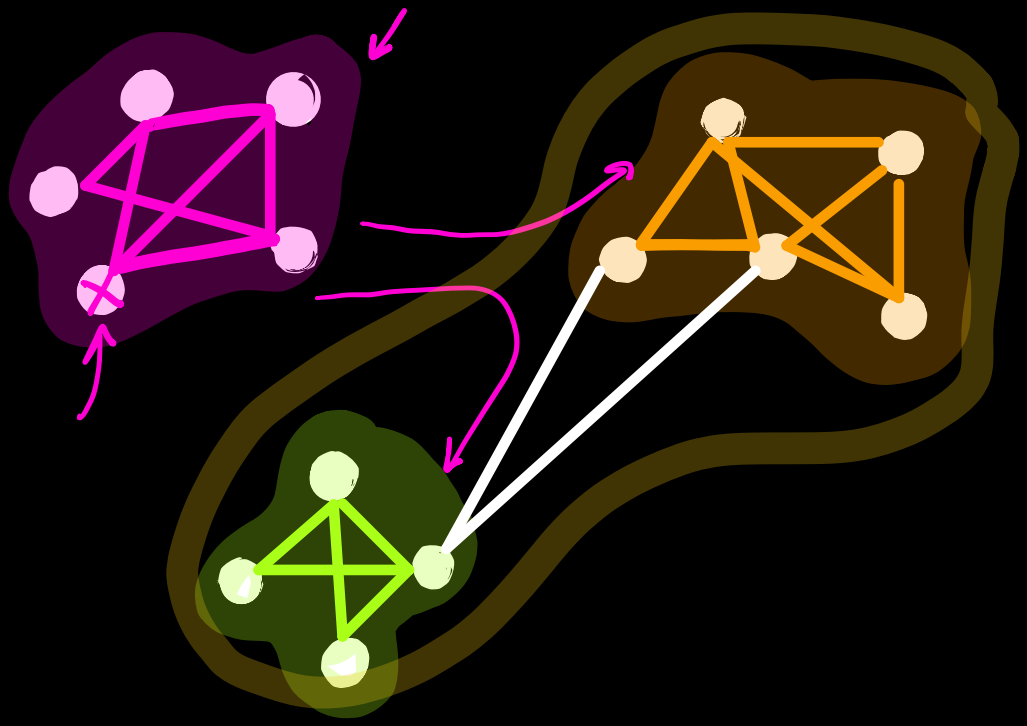
**Strongly** connected component = a **subset of nodes** in which there exists a **directed path that runs in both directions** between **every pair** of two nodes.




Each node in a **strongly** connected **component** should at least belong **to one cycle.**

Two **weakly** connected components ( there is a connection between each pair of node **regardless of the direction** of the edge.

## PERCOLATION & ROBUSTNESS OF A GRAPH



① The size of connected components within a graph can be used **to examine the robustness** of the system **to node damage or failure**

→ This can be done by **node / edge isolation** (or "**lesioning**" in brain connectomes).

→ **remove** a node at a time and **recompute** the size of **the largest component**

→ examine the capacity of the graph to maintain its **connectedness** under different kinds of topological attacks.

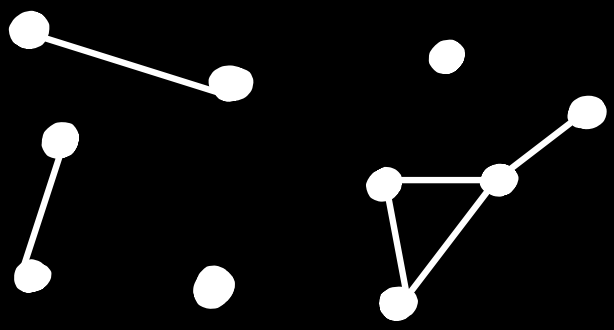
graph topological perturbation

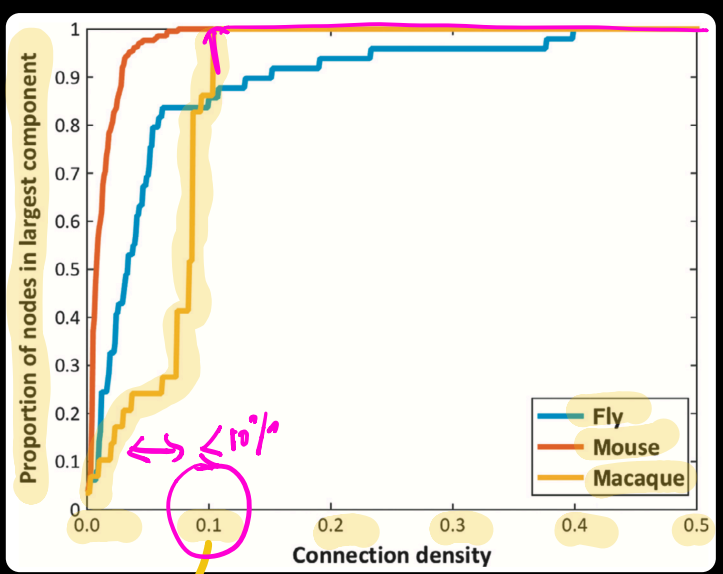
- In graph theory, percolation refers to processes that occur on a graph as nodes / edges are being added or removed.
- Percolation theory emerged from the physical study of how fluid passes, or percolates, through a porous material.

- As we remove nodes / edges from a graph, we find a percolation threshold $p_c$ at which the network fragments.

  $> p_c \Rightarrow$ a large component emerges

  $< p_c \Rightarrow$ fragmentation into multiple smaller components.






[ Chap 6. FBNA ]

* Connectomes of the fruit fly, mouse and macaque.
* We start with an empty adjacency matrix and gradually add edges in descending order of connectivity strength.

percolation threshold $\leq 10\%$
$\Rightarrow$ when 10% of connections are added, a giant CC emerges.

The rapid shift from fragmentation to the emergence of a large component at the percolation threshold implies a phase transition: a rapid change in the state of a system at a critical value.

- Analyzing percolation processes on graphs help simulate the effect of a progressive damage to individual edges / nodes.
- Graph topologies that withstand fragmentation may explain the resilience of a network to random or targeted attacks.

Eg. The first wi-Fi network, called ALOHAnet, built in 1971, was designed as a star-like topology with a central hub allowing efficient transmission between nodes at the periphery.

→ Now we avoid this since the central node represents a single point of failure.

↳ if it gets attacked, the whole network is disabled.

$$S(D) = -\int f(D)\log[f(D)]\mathrm{d}D \quad (5)$$

where the integral is taken over all values of $D$, that is, from 0 to $2\pi$. The use of $D$, rather than $\phi$ itself, to define entropy is one way of accounting for the lack of translation invariance of $\phi$, a problem that was missed in previous attempts to quantify phase entropy[16]. A uniform distribution of $D$ is a state of maximum entropy (minimum information), corresponding to gaussian initial conditions (random phases). This maximal value of $S_{max} = \log(2\pi)$ is a characteristic of gaussian fields. As the system evolves, it moves into states of greater information content (that is, lower entropy). The scaling of $S$ with clustering growth displays interesting properties[5], establishing an important link between the spatial pattern and the physical processes driving clustering growth. This phase information is a unique 'fingerprint' of gravitational instability, and it therefore also furnishes statistical tests of the presence of any initial non-gaussianity[17–19]. □

Received 17 January; accepted 19 May 2000.

1. Saunders, W. *et al.* The density field of the local Universe. *Nature* **349,** 32–38 (1991).
2. Shectman, S. *et al.* The Las Campanas redshift survey. *Astrophys. J.* **470,** 172–188 (1996).
3. Smoot, G. F. *et al.* Structure in the COBE differential microwave radiometer first-year maps. *Astrophys. J.* **396,** L1–L4 (1992).

# Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

*Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, Indiana 46556, USA*

**Many complex systems display a surprising degree of tolerance against errors. For example, relatively simple organisms grow, persist and reproduce despite drastic pharmaceutical or environmental interventions, an error tolerance attributed to the robustness of the underlying metabolic network[1]. Complex communication networks[2] display a surprising degree of robustness: although key components regularly malfunction, local failures rarely lead to the loss of the global information-carrying ability of the network. The stability of these and other complex systems is often attributed to the redundant wiring of the functional web defined by the systems' components. Here we demonstrate that error tolerance is not shared by all redundant systems: it is displayed only by a class of inhomogeneously wired networks,**



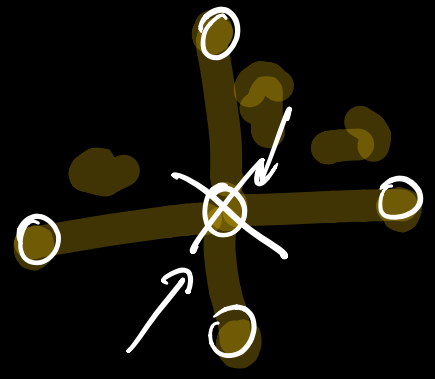


**called scale-free networks, which include the World-Wide Web[3–5], the Internet[6], social networks[7] and cells[8]. We find that such networks display an unexpected degree of robustness, the ability of their nodes to communicate being unaffected even by unrealistically high failure rates. However, error tolerance comes at a high price in that these networks are extremely vulnerable to attacks (that is, to the selection and removal of a few nodes that play a vital role in maintaining the network's connectivity). Such error tolerance and attack vulnerability are generic properties of communication networks.**

The increasing availability of topological data on large networks, aided by the computerization of data acquisition, had led to great advances in our understanding of the generic aspects of network structure and development[9–16]. The existing empirical and theoretical results indicate that complex networks can be divided into two major classes based on their connectivity distribution $P(k)$,

The inhomogeneous connectivity distribution of many real networks is reproduced by the scale-free model[17,18] that incorporates two ingredients common to real networks: growth and preferential attachment. The model starts with $m_0$ nodes. At every time step $t$ a new node is introduced, which is connected to $m$ of the already-existing nodes. The probability $\Pi_i$ that the new node is connected to node $i$ depends on the connectivity $k_i$ of node $i$ such that $\Pi_i = k_i/\Sigma_j k_j$. For large $t$ the connectivity distribution is a power-law following $P(k) = 2m^2/k^3$.

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. Networks with a very large number of nodes can have quite a small diameter; for example, the diameter of the WWW, with over 800 million nodes[20], is around



"disabled"