Social and Graph Data Management: Network Robustness

Pablo Mollá Chárlez

Contents

1 Introduction to Network Robustness: From Communities to Percolation Theory

Robustness is a cornerstone of network science, studying how networks endure disruptions. What happens when certain parts of a network are removed? This question has profound implications in diverse contexts:

- Medicine: Genetic mutations may disable parts of biological networks, affecting disease progression or treatment outcomes.
- Online Social Networks: Targeted attacks or misinformation campaigns could disconnect communities or disrupt communication.
- Global Crises: Events like pandemics, famines, or wars fragment supply chains, infrastructure, or social cohesion, challenging the survival of interconnected systems.

Understanding robustness enables us to design systems resilient to failures and adaptive under stress. This is where **percolation theory** plays a pivotal role, providing a theoretical framework to analyze network behavior under random failures or targeted attacks.

2 Percolation Theory and Its Purpose in Networks

2.1 Percolation: Describing the formation of Networks

Percolation, a concept originating from statistical physics, examines the connectivity of networks as their components (**nodes** or **edges**) are added or removed. In network analysis, percolation answers fundamental questions:

- What is the **expected size of the largest cluster** in the network?
- What is the **average cluster size**, given the presence or absence of specific connections?

2.1.1 Example: A Square Lattice

Imagine a square lattice where “pebbles” are randomly placed at intersections with a probability p. When two or more pebbles are connected, they form clusters. As p increases, the clusters grow in size. Once p surpasses a critical value p_c, a giant cluster spanning a significant portion of the lattice emerges.

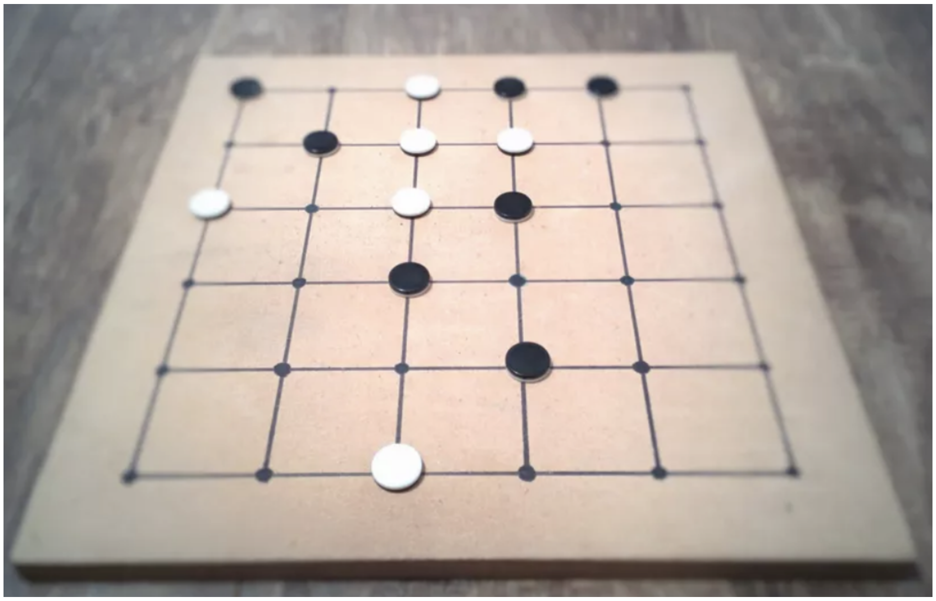

Figure 1: Lattice Board Game

2.2 Percolation Properties

Thanks to the concept of percolation, it becomes easier to identify the average cluster size, the order parameter and the correlation lenght.

1. **Average Cluster Size ($<s>$)**: Diverges near the critical threshold, scaling as below where γ_p is the critical exponent:
$$<s> \sim |p - p_c|^{-\gamma_p}$$
2. **Order Parameter (p_∞)**: Represents the probability that a pebble belongs to the largest cluster, scaling as follows where β_p is the critical exponent:
$$p_\infty \sim (p - p_c)^{\beta_p}$$
3. **Correlation Length (ξ)**: Indicates the mean distance between connected pebbles, scaling as follows where ν is the critical exponent:
$$\xi \sim |p - p_c|^{-\nu}$$

Importantly, these exponents (γ_p, β_p, ν) are universal, meaning they depend only on the system's dimensionality, not on the specific lattice or network type.

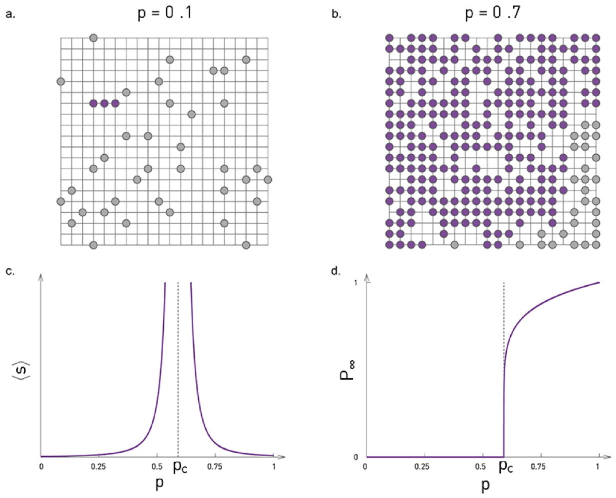

Figure 2: Formation of Giant Cluster in the Network

2.3 Inverse Percolation: Fragmenting Networks

Percolation theory not only describes how networks form (as we just discussed) but also how they fragment under stress. In **inverse percolation**, we remove a fraction f of nodes or edges, disrupting the network's structure. Initially, the network remains resilient, with a giant component still present. However, as f approaches a critical threshold f_c, the giant component collapses, and the network fragments into smaller, disconnected clusters.

This process is non-linear: the transition from a connected to a fragmented state is abrupt, occurring at the critical threshold f_c. The relationship between p_c (for forming a giant component) and f_c (for breaking the giant component) reflects their complementary nature in understanding network robustness.

Reminder The value of p_c has been already discussed in previous summaries, in particular:

- For Erdös–Rényi graphs: $p_c = 1/N$.
- For scale-free networks: Depends on the degree exponent γ.

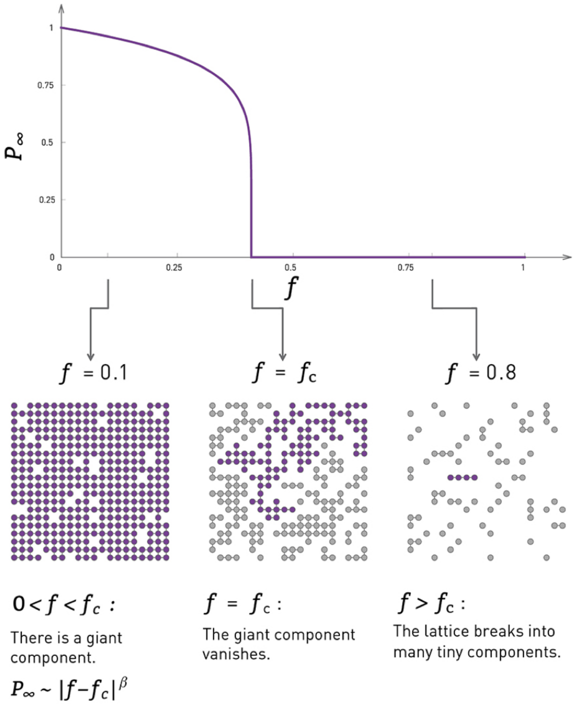

Figure 3: Fragmentation of Giant Cluster in the Network

3 Resilience of Scale-Free and Random Networks

Scale-Free Networks are resilient under random removals. In scale-free networks, the degree distribution follows a power-law, meaning that most nodes have very few connections, while a small fraction of nodes (hubs) have many connections. When nodes are randomly removed, the probability of removing a high-degree hub is low because there are very few hubs. As a result, the connectivity of the network remains largely intact, and a giant connected component persists.

Random networks are moderately resilient under random node failures because their connectivity depends on the average degree of nodes. As long as the average degree ($\langle k \rangle$) is greater than 1, a giant connected component remains (**supercritical regime**). However, their resilience decreases as nodes are removed, especially if key nodes with many connections are targeted, since all nodes have roughly similar roles in maintaining connectivity.

3.1 Molloy-Reed Criterion: Formation and Fragmentation of the Giant Component

The Molloy-Reed criterion provides a mathematical condition for the existence of a giant component in a network:

$$\boxed{\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} > 2}$$

Where the parameters are:

- $\langle k \rangle$ is the average degree of the network.
- $\langle k^2 \rangle$ is the second moment of the degree distribution.

Analogously, the critical fraction f_c of nodes that must be removed to destroy the giant component is:

$$\boxed{f_c = 1 - \frac{1}{\kappa - 1}}$$

The closer f_c gets to 1, the more nodes need to be removed, meaning that if $f_c \to 1$, all nodes need to be removed.

3.1.1 Molloy-Reed Criterion for Random Networks

Random networks follow a Poisson Distribution for the degree distribution, and one of its features is that the variance of such is equal to the mean (therefore $Var(k) = \langle k \rangle$) and we know that $Var(k) = \langle k^2 \rangle - \langle k \rangle^2$, then we have that $\langle k \rangle = \langle k^2 \rangle - \langle k \rangle^2 \longleftrightarrow \langle k^2 \rangle = \langle k \rangle \cdot [1 + \langle k \rangle]$. Then, for random networks (e.g., Erdös-Rényi graphs), the Molloy-Reed criterion simplifies to:

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle \cdot [1 + \langle k \rangle]}{\langle k \rangle} = 1 + \langle k \rangle \longleftrightarrow \boxed{\kappa = 1 + \langle k \rangle} > 2 \longleftrightarrow \langle k \rangle > 1$$

We can conclude that a giant component exists when $\langle k \rangle > 1$. In terms of the critical fraction of nodes to remove (f_c), the formula for random networks would be:

$$\boxed{\mathrm{f}_c = 1 - \frac{1}{\langle k \rangle}}$$

This implies that, for dense random networks ($\langle k \rangle$ is large) it is required the removal of a higher fraction of nodes to disconnect them.

3.1.2 Molloy-Reed Criterion in Scale-Free Networks

For scale-free networks, the degree distribution $P(k)$ follows a power-law distribution:

$$p_k \sim k^{-\gamma} \quad \& \quad \kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}, \; f_c = 1 - \frac{1}{\kappa - 1}$$

Let's remind the results from the summary of Graph Formation:

γ	$\langle k \rangle$	$\langle k^2 \rangle$	$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}$	Resilience
$\gamma \leq 2$	∞	∞	Undefined	Extremely resilient (dominated by hubs)
$2 < \gamma < 3$	Finite	∞	∞	Highly resilient (hubs dominate)
$\gamma \geq 3$	Finite	Finite	Finite	Less resilient (random network-like)

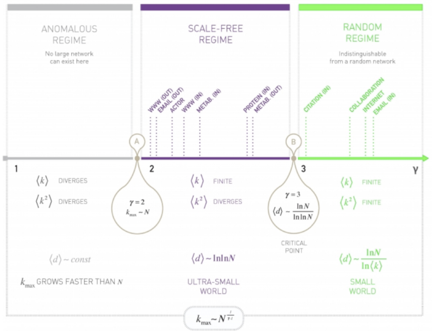

Figure 4: Scale-Free Networks

In scale-free networks with $\gamma < 3$, the resilience increases because the hubs maintain the network's connectivity. The critical fraction f_c is extremely high, meaning that almost all nodes must be removed to destroy the giant component. To be more specific with precise values, the critical fraction f_c can be obtained as:

$$f_c = \begin{cases} 1 - \frac{1}{\frac{\gamma-2}{3-\gamma} k_{\min}^{\gamma-2} k_{\max}^{3-\gamma} - 1} & \text{if } 2 < \gamma < 3, \\ 1 - \frac{1}{\frac{\gamma-2}{\gamma-3} k_{\min} - 1} & \text{if } \gamma > 3. \end{cases}$$

4 Attack Robustness

Hubs play a crucial role in holding together scale-free networks, leading to an important question:

$$\boxed{\text{What happens if nodes are not removed randomly but instead in a targeted manner?}}$$

In targeted attacks, the most connected hub is removed first, followed by the next most connected hub, and so on. While such an order of removal is highly unlikely under normal circumstances, it simulates an intentional attack on the network by exploiting its topology. Removing a single hub does not immediately fragment the network, as the remaining hubs can maintain connectivity. However, as more hubs are removed, large portions of the network break away, forming isolated clusters. This behavior contrasts with random failures, where the network remains robust due to the redundancy provided by the hubs.

The following graph illustrates this vulnerability clearly: under random failures, the largest component (**Y-Axis)** remains large even after removing many nodes. However, under targeted attacks, even the removal of a small fraction of hubs is enough to reak the network into tiny clusters. This highlights the network's high resilience to random failures but significant vulnerability to deliberate attacks.

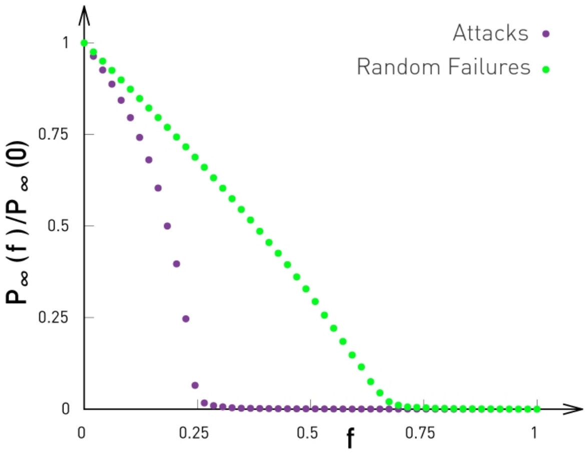

Figure 5: Targeted Attacks vs. Random Attacks

4.1 Critical Threshold under Attacks

When γ (the degree distribution exponent) is very large ($\gamma \to \infty$), a scale-free network starts to resemble a random network. This happens because the degree distribution becomes very narrow—almost all nodes have the same degree, close to the minimum degree $k_{\min}$. In such a network, there are **no prominent hubs** (nodes with significantly higher connections). As a result, the impact of targeted attacks (removing the most connected nodes) becomes similar to the impact of random failures (removing nodes randomly), since all nodes are roughly equivalent in importance.

The failure threshold (f_c, the fraction of nodes that need to be removed to break the network) and the attack threshold (how many nodes need to be removed to cripple the network) converge as $\gamma \to \infty$. In the extreme case where $\gamma \to \infty$, all nodes have exactly $k_{\min}$ connections (degree becomes a delta function). In this scenario, the fraction of nodes that need to be removed to break the network is given by $f_c \to 1 - \frac{1}{k_{\min}-1}$. This reflects the lack of hubs and the network's uniform structure. Essentially, as γ grows, the network loses its resilience to targeted attacks and behaves like a random network.

4.2 Cascading Failures

Previously, we discussed how scale-free networks are highly resilient to random failures but vulnerable to targeted attacks, especially when hubs are deliberately removed. This vulnerability becomes critical when we consider cascading failures, a phenomenon that amplifies the initial damage caused by an attack.

Cascading failures occur when the failure of one or more nodes triggers a chain reaction that causes other nodes to fail as well. This happens because networks are often interdependent: the removal or breakdown of one node affects its neighbors, which in turn impacts their neighbors, and so on. This cascading process can lead to widespread disruptions. Cascading failures are pervasive in real-world networks:

- **Power grids**: The failure of one substation can overload nearby substations, leading to widespread blackouts.
- **Denial of service (DoS) attacks**: Overloading a critical server can cause dependent systems to fail.
- **Social networks**: In information cascades, one viral post can quickly spread misinformation or trigger large-scale behavioral changes.
- **Financial systems**: The collapse of one financial institution can lead to a domino effect, resulting in widespread market crises.

Interestingly, cascading failures often follow power-law distributions. This means the frequency and size of cascading events are highly skewed:

- Small cascades are common.
- Large, catastrophic cascades are rare but possible, leading to blackouts, viral outbreaks, or systemic financial crises.

This characteristic aligns with the broader behavior of scale-free networks, where hubs play a critical role in both stability and fragility. For instance, in random failures, hubs typically survive, containing cascades and in targeted attacks, the removal of hubs can initiate large-scale cascades that cripple the network.