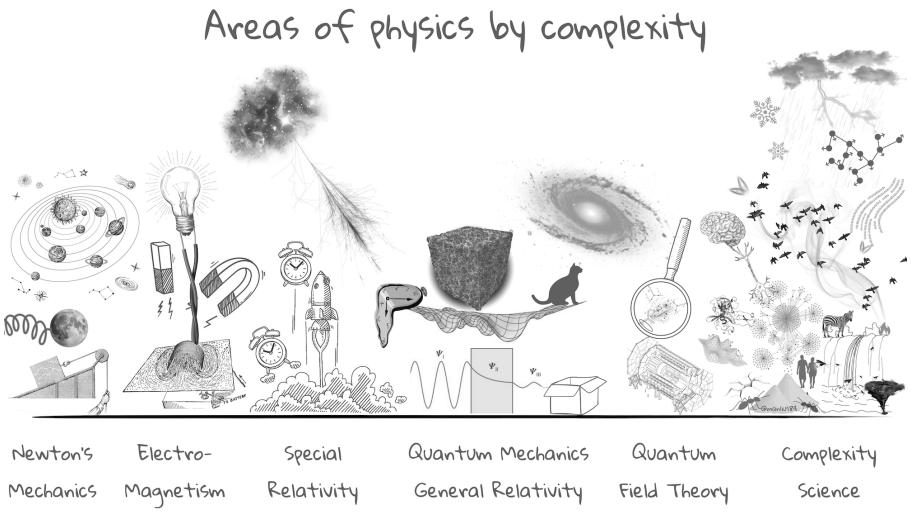
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



Project: Percolation

Team members:

Mojtaba Roshana

Roya Joulaei Vijouyeh

Maryam Feizi

Mahday Vtankhahan

Last update: July 4, 2023

Contents

1 | Explosive Percolation in Random Networks

Task leader(s): *Mojtaba Roshana*

1.1 | Introduction

Percolation is a concept in network science that studies the formation and propagation of connectivity in a network. It involves analyzing how information, signals, or substances spread through interconnected nodes and edges in a network. Percolation is important in various problems. For example, in Power Grids, Understanding how the addition or removal of power lines affects the overall connectivity and stability of the grid can help in optimizing its design and enhancing its resilience against failures.

Sometimes by adding or removing a small number of edges in the network causes a sudden and dramatic change in the overall connectivity. For instance, in a network with two connected components if we add a link from the first connected component to the other one, the network will be fully connected only by adding one link. Explosive percolation is a phenomenon that occurs in networks, where sudden and rapid changes take place from a disconnected state to a connected state.

In networks where connections are formed randomly, a percolation transition often occurs. This transition occurs near a critical point, where adding just a few connections leads to a significant portion of the network becoming interconnected. It's a remarkable phenomenon where a small number of connections can have a large impact, causing a sudden increase in network connectivity. Normally, percolation transitions behave continuously, in random networks. We will consider percolation phase transitions in models of random network formation and will show that it can be discontinuous in one of them.

1.2 | Method and Implementation

The percolation critical point in the context of percolation theory refers to the threshold at which a phase transition occurs in a random network or lattice. ER stands for Erdős-Rényi model, which is a random graph model commonly used in percolation theory.

In the Erdős-Rényi model, a network or graph is constructed by randomly connecting nodes with edges. The model assumes that each pair of nodes has a certain probability of being connected. The percolation critical point in the ER model rep-

resents the probability threshold at which a giant connected component (percolating cluster) emerges or disappears in the graph.

At the critical point, the network undergoes a phase transition from a disconnected state to a connected state (percolation transition). Below the critical point, the graph consists of small isolated clusters, whereas above the critical point, there is a giant connected component that spans a significant portion of the graph.

In the classic Erdős-Rényi (ER) model, we start with n isolated nodes and add links one by one, each link formed by picking two nodes uniformly at random and connecting them. now we should find the point that the giant connected component is forming which is known as the emergence of a giant component. The average degree for ER is :

$$\langle k \rangle = \sum(k \cdot P(k)) = p(n-1) = pn \qquad , N >> 1 \tag{1.2.1}$$

And the critical probability [4] is:

$$p_c = \frac{1}{n}, r_c = \frac{1}{2} \tag{1.2.2}$$

Which means the average degree should be equal to one for the transition point.(see Appendix A5.2)Therefore after adding rn links if $r < 1/2$ the largest component remains small and the size of it is scaling as $\log n$ which is too small for networks with large sizes. When $r > 1/2$ which means we added links equal to half of the size of network, there will be a component of size linear in n:

$$C \approx (4r - 2)n \tag{1.2.3}$$

There is a continuous phase transition at r = ½ which means after adding nodes more than half of network size we will have a sudden change in connectivity which can be observed by measuring size of largest connected component. In fig1.1 it can be seen there is a jump in r = ½ which is a continues phase transition starts from around zero.

By adding links randomly and independently we can recover ER model. However, nonrandom selection rules can delay (or accelerate) percolation. We will discuss two other selection rules. Instead of adding links completely randomly, these random selection rules will select uniformly two edges and connect one of them and leave the other one in each step. In this report, we will compare the phase transition of these three random network selection rules:

- Classic Erdős-Rényi (ER): we start with n isolated nodes and add links one by one, each link formed by picking two nodes uniformly at random and connecting them. the nodes which select to be connected are chosen randomly and independently.
- bounded-size Bohman and Frieze (BF): For K=1 In this model we randomly choose two links $\{e_1, e_2\}$ and if e_1 is connecting two connected components with size one we will choose it otherwise we will choose e_2. In BF selection rule we prefer to connect isolated nodes.

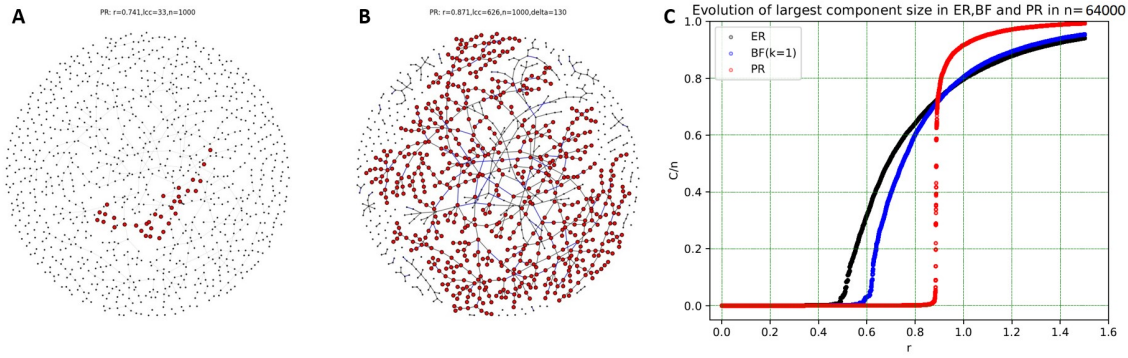


Figure 1.1: **Network Evolution.** (A) A random network with 1000 nodes after adding 741 links with the largest connected component 32 with PR model and (B) is the same network after adding overall 871 links with largest connected component equal to 626. New links shown by blue edges in (B) and red nodes belong to the largest connected component. Figure (C) shows the evolution of size of largest connected component during generating the network with three selection rules with 64000 nodes. rn is the number of links that we are adding to the network. The evolution of (A) to (B) is shown in the area that PR curve suddenly increases in (C).

- Product-Rule (PR): In this model, we begin by randomly selecting two links, denoted as $\{e_1, e_2\}$. For e_1, we calculate the product of the sizes of the connected components of its two endpoints. Similarly, we calculate the product of the sizes of the connected components for e_2's endpoints. Then, we retain the edge with the smaller product. In other words, the selected edge is the one that minimizes the product of the sizes of the components it merges.

1.3 | Results

We generated random networks and examined the behavior of the selection rules, particularly around the critical point. In fig1.1(C) we can observe the evolution of the largest connected component during the growth. Both ER and Bf model has second-order phase transition. The ER model has a jump at r=0.5 and it is a continuous transition as we expected. Besides, at r=0.6 the BF model has similar behavior and the phase transition is continuous. However, for the PR model, we can observe a first-order phase transition which is discontinuous. It has a sudden jump from C/n=0 to around 0.9 in a few steps to compare with two other models. To study the transition part, we define t_0 as the number of steps to reach $\frac{C}{n} > n^{\frac{1}{2}}$ and t_1 as the number of steps to reach $\frac{C}{n} > \frac{n}{2}$. Finally, we define $\Delta = t_1 - t_0$ and calculate it for large networks. The data presented in Figure 1.2 represents the average of 10 different random realizations obtained from simulations.

In fig1.2 (A) The process in classic ER is slower to reach from t_0 to t_1 which means needs more links or steps to reach it. In continuous transitions, the interval Δ is always extensive, meaning that it scales linearly with the network size n.

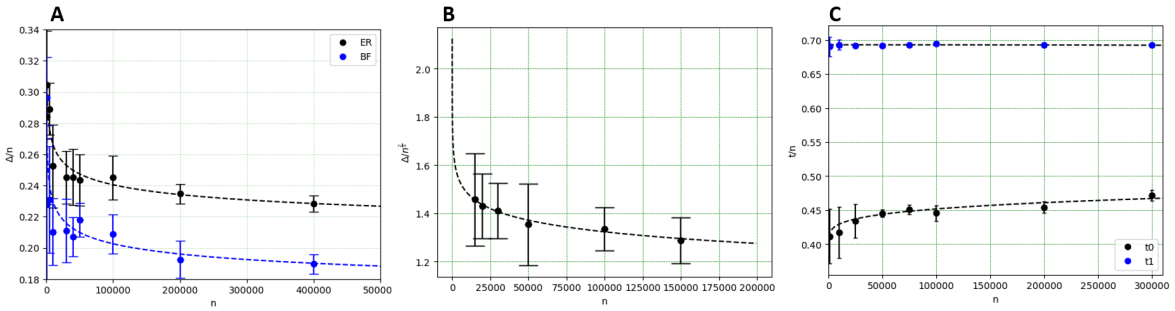


Figure 1.2: (A) The ratio Δ/n for ER and BF by increasing system sizes. (B) The ratio $\Delta/n^{2/3}$ for PR by increasing system sizes. The dashed line is a fitted curve on data in both plots. (C) The t_0 and t_1 are in different sizes of networks for the PR model. The dashed curve is a fit on data and shows t_0 and t_1 are converging in larger networks.

In fig1.2 (B) we can see the $\Delta$for the PR model in different sizes of networks. Δ is not extensive for the product rule $\Delta < 2n^{2/3}$. If calculate it for larger networks [1] it appears that $\frac{\Delta}{n^{2/3}} \to 1$. As a result, the percentage of the nodes in the largest connected component will increase suddenly by adding a few links which means there is a discontinuous transition. Also in fig1.2(C) t_0 and t_1 are converging which can be seen by the dashed curve which fitted on data.

1.4 | Conclusion

In this task, we reproduced the results of reference paper [1] and studied the phase transition phenomena in three models. We introduced the variable $\Delta = t_1 - t_0$ to analyze the behavior of the phase transition of models.

Although our ability to generate large networks was limited compared to the original paper, we were able to obtain significant results. For the Erdős-Rényi (ER) and e Bohman and Frieze (BF) models, we demonstrated a second-order phase transition, in line with the findings of the previous study. However, in the case of the Product rule, we identified a second-order phase transition instead which is the explosive percolation.

Our research highlights the significance of even small changes in edge formation, as they can fundamentally change percolation transitions. These findings suggest the need for further investigation into this phenomenon and its potential for controlling phase transitions.

2 | Assortative mixing in networks

Task leader(s): *Roya Joulaei Vijouyeh*

2.1 | Introduction

In this task, we will work on the assortativity of the complex networks that show if a network is assortative, the nodes of which tend to have a connection with ones that have the same character. In other words, hubs tend to link to each other and avoid linking to small-degree nodes. However, in disassortative networks, the hubs avoid each other, linking instead to small-degree nodes.As the table shows in [7], the social networks studied all have signifcant assortative mixing, which accords with accepted wisdom within the sociological community. By contrast, the technological and biological networks studied all have disassortative mixing. In the context of World-Wide web popular nodes (websites or servers with high connectivity or traffic) tend to link to less popular nodes, and vice versa. This characteristic can be observed in various aspects of the Internet, such as the connectivity patterns of websites, the distribution of Internet traffic, and even the peering arrangements between Internet service providers. The information about potential degree correlations are captured by the degree correlation matrix, e_{jk}, which is the probability of finding a node with degrees j and k at the two ends of a randomly selected link. In this task the correlation matrix that we are using is mentioned bellow :

$$e_{jk} = Ne^{-(j+k)/\kappa}\left[\binom{j+k}{j}p^j q^k \binom{j+k}{k}p^k q^j\right] \tag{2.1.1}$$

which consists of an exponential and a binomial distribution.In this equation

j and k are the degrees of randomly chosen edges

p and q are the binomial probabilities in which p+q = 1

N is the normalizing constant 5.4.6

and κ is the degree scale parameter.

In the degree-correlation 2.1.1, the binomial coefficients play a crucial role in determining the degree correlation. The $\binom{j+k}{j}$ and $\binom{j+k}{k}$ count the number of ways to choose j and k edges, respectively, among the total (j+k) edges that are connected to the vertices in consideration. These arrangements determine how the vertices are connected and influence the overall degree correlation in the graph.

If we wish to characterize degree correlations using a single number, we can use r which is the degree correlation coefficient Proposed by Mark Newman. the degree correlation coefficient is defined as :

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk \left(e_{jk} - q_j q_k \right) \tag{2.1.2}$$

Hence r is the Pearson correlation coefficient between the degrees found at the two end of the same link. It varies between $-1 \leq r \leq 1$: For $r > 0$ the network is assortative, for $r = 0$ the network is neutral and for $r < 0$ the network is disassortative.

2.2 | Method and Implementation

We want to show the evolution of the size of the giant component, the number of nodes in the largest connected component, for graphs of this type as a function of the degree scale parameter κ in order to evaluate the percolation in assortative mixing networks.

we know that :

$$\begin{aligned} \sum_{j,k} e_{j,k} &= 1 \\ \sum_{j} e_{j,k} &= q_k \end{aligned} \tag{2.2.1}$$

The remaining degree q_k, represents the vertex at the end of a randomly chosen edge which is proportional $(k+1)p_{k+1}$. We can connect q_k to e_{jk} 5.4.1. As mentioned before, we want to calculate the giant component of our generated graph with the degree of correlation e_{jk}. To calculate the size S of the giant component, we define u_k to be the probability that an edge connected to a vertex of the remaining degree k leads to another vertex that does not belong to the giant component which has derived from the generative model5.4.2. .

$$S = 1 - p_0 - \sum_{k=1}^{\infty} p_k u_{k-1}^k \tag{2.2.2}$$

it is usually not possible to solve for S in closed form, but we can determine it with computer simulations. according to [7] we will use the Monte Carlo algorithm which is a probabilistic approach to create random graphs.These simulations would be helpful to have a more complete picture of the properties of assortatively mixed networks. The steps to generate our graph : first, we generate a random graph with the degree sequence mentioned above. Next, we employ Metropolis dynamics as shown in Fig2.2 on the graph, where, at each step, we randomly select two edges represented by the vertex pairs (v_1, w_1) and (v_2, w_2) that they connect. We then calculate the remaining degrees (j_1, k_1) and (j_2, k_2) for these vertex pairs. Afterward, we replace the edges with two new ones (v_1, v_2) and (w_1, w_2) using a probability defined $\min\left(1, \frac{e_{j1j2}e_{k1k2}}{e_{j1k1}e_{j2k2}}\right)$.

2.3 | Results

The results can be seen in Fig2.2. We use the number of 10000 for both the number of nodes and iterations. The figure illustrates three lines for three different p. p plays as controlling the assortative mixing parameter for an assortative, neutral, and

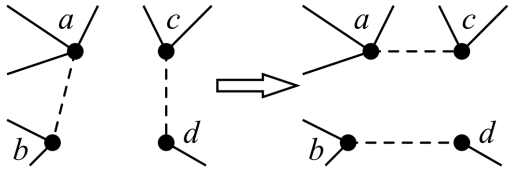


Figure 2.1: This picture shows a link rewiring process. Two links between node pairs (a, b) and (c, d) are chosen at random. They are then rewired to connect pairs (a, c) and (b, d)

disassortative graph which has been set as 0.5, 0.1466, and 0.05 respectively in 2.1.1. By analyzing the result we can see that for lower κ, these three graphs behave similarly. By contrast, as κ increases considerably, the assortative graph experiences the phase transition sooner than the two other graphs. This result is completely reasonable since in assortative networks hubs are connected to hubs, therefore, the graph percolates more easily, creating the giant component. on the other hand, the size of the giant component of this graph with higher κ will be smaller. Because in assortative graphs percolation will be restricted in the sub-network, when high-degree vertices will tend to stick together, consequently, the size of the giant component will have a smaller size. In terms of a disassortative graph, percolation is less easy because the phase-transition occurs lately.

Assortative mixing also has implications for questions of network resilience which is evaluated by using giant component size following the removal of a few nodes Which is shown in Fig2.3. Networks that maintained a high giant component size were considered to possess higher connectivity and consequently robustness. It has been found that the connectivity of many networks (i.e., the existence of paths between pairs of vertices) can be destroyed by the removal of just a few of the highest degree vertices. In assortatively mixed networks, however, we know that removing high-degree vertices is a relatively inefficient strategy for destroying network connectivity, presumably because these vertices tend to be clustered together in the core group, so that removing them is somewhat redundant. In a disassortative network with a similarly sized giant component attacks on the highest degree vertices are much more effective, these vertices being broadly distributed over the network and presumably therefore forming links on many paths between other vertices

2.4 | conclusion

we evaluate the assortatively mixed network, which is simulated using a Monte Carlo method. In this model we find that the degree correlation influences the formation and structure of the giant connected component during percolation. Assortative mixing tends to accelerate the formation of the giant component, while disassortative mixing may slow it down. Assortative networks percolate more easily and that they are also more robust to removal of their highest degree vertices, while disassortative networks percolate less easily and are more vulnerable. This suggests that social networks may be robust to intervention and attack while technological networks are not.

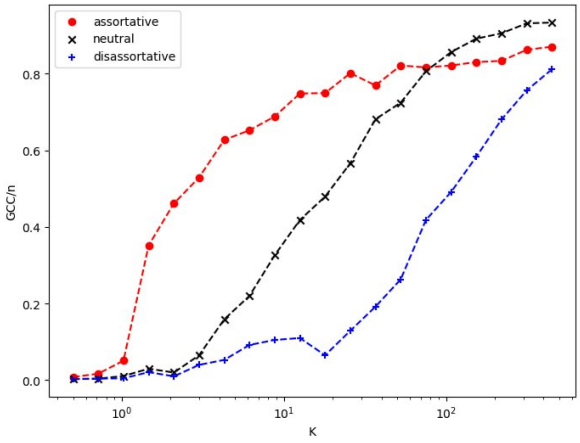


Figure 2.2: The size of the giant connected component as the fraction of κ. we set $p = p_0$=0.1466 for neutral, p=0.5 for assortative and p=0.05 for disassortative network. The number of nodes is 10000 and the number of iterations is 10000. As we can see the assortative graph experience the phase transition earlier, however; the size of the giant component for higher κ is bigger for disassortative networks

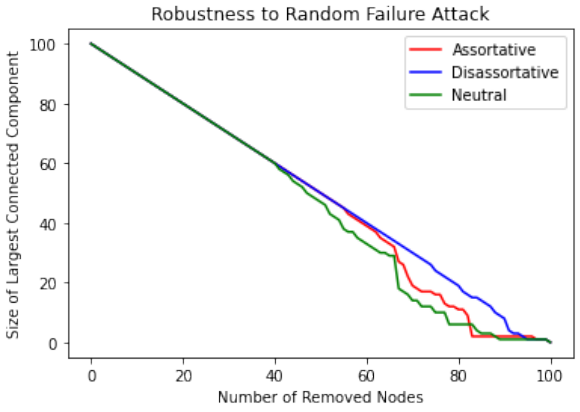


Figure 2.3: A comparison of assortative, disassortative, and neutral random networks. The plot shows the size of the largest connected component as nodes are randomly removed from each network. As we observe, the assortative network exhibits higher robustness to random node failures, maintaining a larger connected component compared to the other networks.

3 | K-Clique

Task leader(s): *Maryam Feizi*

3.1 | Introduction

In this part we will work on k-clique percolation in random graphs which introduced, where k is the size of the complete sub-graphs whose large scale organizations are analytically and numerically investigated. We want to measure the greatest k-clique percolation cluster with two different way and plot the results for many different network sizes. For this reason we must first provide some fundamental definitions before moving on to the exponent calculations. The main focus of our study, k-cliques, are complete (completely linked) subgraphs with k edges.[2].
As an illustration we should introduce a few notations specific to our problem

- k-clique adjacency: two k-cliques are adjacent if they share k-1 nodes
- k-clique chain: a subgraph, which is the union of a sequence of adjacent k-cliques.
- k-clique conceitedness: two k-clique are k-clique-connected if they are parts of a k-clique chain.
- k-clique percolation cluster(component): it is a maximal k-clique subgraph.

3.2 | The Main Work

A k-clique percolation cluster is similar to a conventional percolation cluster in the k-clique adjacency network, where the vertices reflect the original graph's k-cliques and an edge exists between two vertices if the associated k-cliques are adjacent. Rolling a k-clique template from one k-clique of the original graph to an adjacent one is analogous to moving a particle from one vertex of this adjacency graph to another along an edge. A k-clique template is an item that is isomorphic to a full graph with k vertices.A template of this type can be set on any k-clique of the original graph and rolled to a neighboring k-clique by moving one of its vertices while leaving the other $k-1$ vertices unchanged. As a result, a graph's k-clique percolation clusters are all those subgraphs that can be fully explored but cannot be abandoned by rolling a k-clique template across them. Now, we present a general result for the threshold probability (critical point) of k-clique percolation using heuristic arguments.The percolation transition of k-cliques occurs for the Erdos-Rényi network with N nodes when the chance of two nodes being connected by an edge crosses the threshold:

$$p_c(k) = \frac{1}{[(k-1)N]^{\frac{1}{k-1}}} \tag{3.2.1}$$

In the other word, a giant component appears in the ER graph where $p = p_c(k)$ and if $p < p_c(k)$ we do not have giant component.

The greatest k-clique percolation cluster can be measured in two possible ways. The most typical one is the number of nodes that belong to this cluster, which we denote by N. The relative size of that cluster can also be defined as an order parameter linked with this selection:

$$\Phi = \frac{N^*}{N} \tag{3.2.2}$$

The other choice is the number $\mathcal{N}^*$ of k-cliques of the largest k-clique percolation cluster (or equivalently, the number of vertices of the largest component in the k-clique adjacency graph). The associated order parameter is again the relative size of this cluster:

$$\Psi = \frac{\mathcal{N}^*}{\mathcal{N}} \tag{3.2.3}$$

where $\mathcal{N}$ denotes the total number of k-cliques in the graph (or the total number of vertices in the adjacency graph). $\mathcal{N}$ can be estimated as:

$$\mathcal{N} \approx \binom{N}{k} p^{k(k-1)/2} \approx \frac{N^k}{k!} p^{k(k-1)/2} \tag{3.2.4}$$

3.3 | Coding Results

In the next step we wrote a code to simulate results for order parameters Φ and Ψ. We averaged over several runs. As it shows in the following there is four different figures. Our computer simulations indicates that two order parameters behave differently near threshold probability.

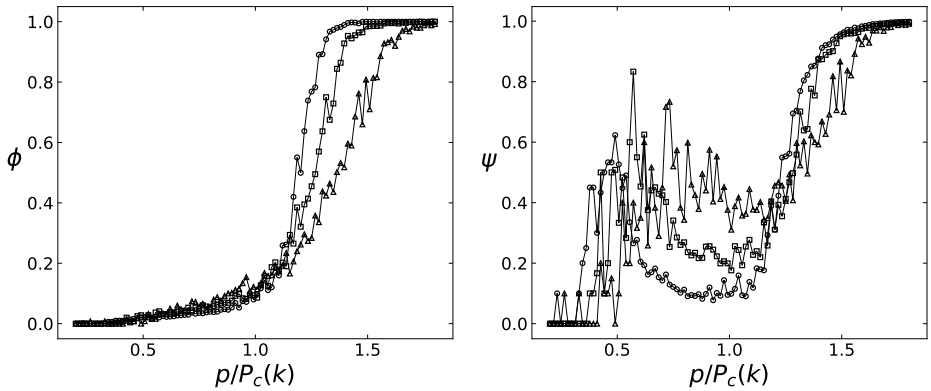
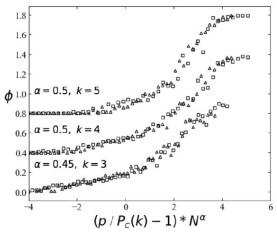


Figure 3.1: behaviour of order parameters over number of nodes in different networks

Fig. 3.1. indicate the behaviour of order parameters Φ and Ψ as a function of $p/p_c(k)$ when the k is fixed and $N \to \infty$. In addition, we measure on many networks with the same size and averaged on all of them to find each point. As you can see, we have jump near the threshold probability but they are not as same as each other

[a] 

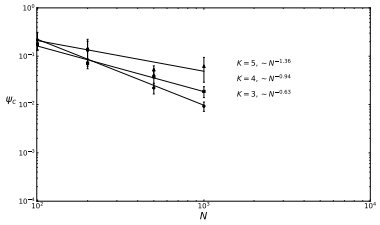
[b] 

Figure 3.2: Order parameters for different ks

In the Fig. 3.2 we try to find a behaviour of our order parameters as a function of N, but there is a difference between x-axis in two plots. The width of the steps follows a power law, $\sim N^{-\alpha}$, with some exponent α. Plotting Φ as a function of $[p/p_c(k)-1]N^{\alpha}$, when the horizontal scale is stretched by N^{α}, the data collapses into a single curve. Fig. 3.2(a) shows this for k = 3, 4, and 5. For $k = 3$, the exponent appears to be around 0.5. Although we obtained a minor departure form = 0.5 for $k = 3$, we cannot distinguish it from a hypothetical logarithmic correction. The data for k = 4 and 5 are shifted upward by 0.4 and 0.8.

One of the most important results in random graph theory is the behavior of the greatest component near the percolation threshold, where it becomes a massive (infinitely large) component in the N limit. Erdos-Rényi shown [6] that the size of the greatest component $\mathcal{N}^*$ (measured as the number of its edges) at $p = p_c \equiv 1/N$ diverges with the system size as $N^{2/3}$, or equivalently, the order parameter scales as $N^{-1/3}$. Because the threshold component has a tree-like structure, its number of vertices, N, similarly diverges as $N^{2/3}$. At the threshold probability $p_c(k)$, we will show that identical scaling behavior can be obtained for k-clique percolation.

If we assume, that the k-clique adjecancy graph is like an ER graph, then at the threshold the size of giant component $\mathcal{N}_c^*$ scales as $\mathcal{N}_c^{2/3}$. The "c" shows the system is at the percolation threshold. Plugging $p = p_c$ from 3.2.1 into 3.2.4 and omitting the N-independent factors, we get the scaling $\mathcal{N}_c \sim N^{k/2}$ for the total number of k-cliques and the order parameter Ψ_c scales as $\mathcal{N}_c^{2/3}/\mathcal{N}_c \sim N^{-k/6}$. However, this is valid only if $k \leq 3$ and for $k > 3$ it predicts that the number of vertices of the giant component in the k-clique adjacency graph, grows faster than N. On the other hand, in analogy with the structure of the giant component of the classical ER problem, we expect that the giant component in the adjacency graph also has a tree-like structure at the threshold, with very few loops. As a consequence, almost every vertex of the adjacency graph corresponds to a vertex of the original graph. Thus, in the adjacency graph the giant component should not grow faster than N at the threshold. As a result, for $k > 3$ we expect that $\mathcal{N}_c^* \sim N$, and $\Psi_c = \mathcal{N}_c^*/\mathcal{N}_c \sim N^{1-k/2}$. Also, we have determined that scaling of Ψ at p_c as a function of N numerically, and the results are in good agreement with the above heuristic arguments, as shown in Fig.3.2b.

3.4 | Conclusion

Theoretically, our community concept can be applied to actual networks thanks to the sharp percolation transition (step in) of the ER graphs. This is due to the fact that for any k at which the network is below the transition threshold, only very few and small clusters would be anticipated if the network were totally random. Large clusters

must, however, correspond to locally dense structures, or actual communities, if they do arise. Furthermore, as long as the edge density of these communities stays above the percolation threshold, their identification is resistant to random edge removal [10]. The most crucial feature of such an approach is that a single vertex can naturally be a part of multiple communities. There are no $k-1$ persons in any two groups who would all know each other and, as a result, would allow a k-clique template to roll through. This means that, in terms of an individual, they can be a part of many groups (of highly linked people). As a result, each vertex can be a member of several distinct communities that can be identified by itself, and as a result, each community can have a lot of contacts with other communities, which is what occurs in the majority of realistic situations.

4 | K-Core Percolation

Task leader(s): *Mahdy Vatankhahan*

4.1 | Introduction

Extracting and indexing highly interconnected parts of complex networks—communities, cliques, cores, etc. as well as finding relations between these sub structures is an issue of topical interest in network research. This decomposition helps one to describe the complex typologies of real-world networks. In this respect, the notion of k-core is of fundamental importance.
The k-core is the largest subgraph where vertices have at least k interconnection and it can be obtained using the following method. Remove all edges with degrees less than k from a graph. Some of the remaining vertices may still have less than k edges. Then delete these vertices, and so on until no more can be removed. The k-core is the result, if it exists. As a result, a network is organized as a series of successively enclosed k-cores . The formation of a huge k-core below a threshold concentration of vertices or edges eliminated at random is implied by k-core percolation.[5]

4.2 | The Main Work

In section 5.2, we have put exact equations describing the k-core organization of a randomly damaged uncorrelated network with an arbitrary degree distribution. Which allows us to obtain the sizes and other structural characteristics of k-cores in a variety of damaged and undamaged random networks and find the nature of the k core percolation in complex networks. There, we applied our general results to the classical random graphs and to scale-free networks.
In the following however, in order to validate the results of the mathematical approach, we will conduct simulations using both the R and Python environments. We will generate several graphs of the desired type and analyze k-core percolation process on them individually in order to find out the relative sizes of the k-cores and k-shells. By comparing the simulated graph properties with the mathematical approach outlined in 5.2, we can assess the accuracy and effectiveness of the proposed method. This validation process will enable us to verify the applicability and robustness of the mathematical approach through empirical experimentation.

4.3 | Implementation

For those who did not inspect section 5.2 , the concept of the k-core size, denoted by $M(k)$, is commonly used to measure the extent of connectivity within a network. It represents the number of vertices in the graph with a degree greater than or equal to

k, divided by the total number of vertices in the graph. On the other hand, the k-shell is defined as the difference between the k-core sizes of consecutive levels, denoted by $S(k)$, and can be expressed mathematically as:

$$S(k) = M(k) - M(k+1) \tag{4.3.1}$$

This equation allows us to quantify the variation in the k-shell size as we move from one level of the k-core to the next.
Our objective is to generate two plots. The first plot demonstrates the decay of k-cores for $k \in [3, 7]$ in a randomly generated Erdős-Rényi graph with an average degree of $z_1 = 10$. The second plot showcases the fluctuations in core and shell sizes as k varies across various networks, primarily comprising Erdős-Rényi graphs and a Scale-free graph characterized by a known degree distribution denoted as (see section 5.3):

$$P(q) \propto (q + c)^{-\gamma} \tag{4.3.2}$$

In addition to that, we consider the Router-Level Internet as our real world network model.
In the initial stage, the networks were constructed. Table 4.1 provides information on the criteria and methods used to create each network:

Graph type	Key property	Graph details	Construction Method
ER	$z_1 = 10$	$N = 10^6$	A
ER	$z_1 = 20$	$N = 10^6$	A
Scale-free	$\gamma = 2.5$, $c = 1$, $q_0 = 1$, $q_{\text{cut}} = 2000$	$N = 22125$, $E = 60906$	B
Scale-free	$\gamma = 4$, $c = 30$, $q_0 = 1$, $q_{\text{cut}} = 2000$	$N = 294475$, $E = 2331298$	B
Scale-free	$\gamma = 7$, $c = 50$, $q_0 = 1$, $q_{\text{cut}} = 447$	$N = 45226$, $E = 241489$	B
Real-World	Router-Level Internet	$N = 192244$, $E = 609066$	Edgelist

Table 4.1: Graph properties and construction details, involve the total number of nodes N and edges E in the network. Method A employs the `Graph.Erdos_Renyi()` function from the `igraph` library to construct the graph. However, Method B utilizes the R environment and the `sample_degseq()` function from the `igraph` library to generate a degree sequence based on a specified degree distribution. This degree sequence is then used to define the network's edge list, which is subsequently imported into the Python environment using the `Graph.Read_Edgelist()` function. Additionally, the size of the scale-free networks was estimated using Equation 5.3.5 in 5.3

In the next phase, we implemented two essential functions. The first function is designed to remove a fraction, denoted as Q, of randomly selected vertices from the network. Here, the range of Q is specified as $Q \in [0, 0.8)$ for obvious reasons. The second function is responsible for determining the size of the k-core at any stage of the network for any value of k.
By employing these functions also applying built-in functions available in Igraph library, we were able to successfully generate Figures 4.1 , 4.2 in section 4.4.

4.4 | Results and Conclusion

In summary, using mathematical approach 5.2, we have developed the theory of k-core percolation in damaged uncorrelated networks. We have found that if the second moment of the degree distribution of a network is finite, the k-core transition has a hybrid nature. In contrast, in networks with infinite z_2, instead of the hybrid transition, we have observed an infinite order transition, similarly to the ordinary percolation in this situation.
Here we present the simulation that confirms the validity of the mathematical approach we discussed earlier. It is noteworthy that each data point in plot 4.1 represents the average of 15 calculations performed with different random seeds. Our objective was to maximize the number of nodes in the networks while minimizing time complexity. The entire process of generating the graphs and producing the plots required approximately one day. Consequently, we observe a slight discrepancy for the Scale-free network with $\gamma = 2.5$ in Fig. 4.2. However, apart from that, the results align perfectly with the findings in [5].
All in all, we validated the statement *"In networks with a finite mean number z_2 of the second-nearest neighbors, the emergence of a k-core is a hybrid phase transition. In contrast, if z_2 diverges, the networks contain an infinite sequence of k-cores that are ultra-robust against random damage"*.

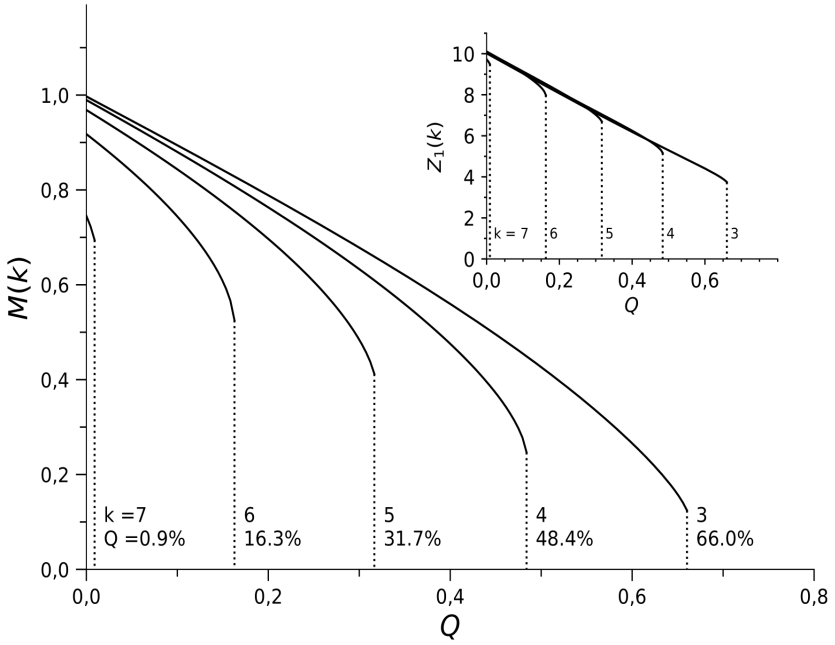


Figure 4.1: The size of the k-core, M_k, in the Erdős-Rényi random graph with the mean degree $z_1 = 10$ versus the concentration Q of vertices removed at random. The highest core disappears at a very low concentration $Q \approx 1\%$ in contrast to the ordinary percolation threshold $Q \approx 90\%$ [5]. The inset shows the mean degree $z_{1,k}$, the mean degree of vertices, in the k-core.

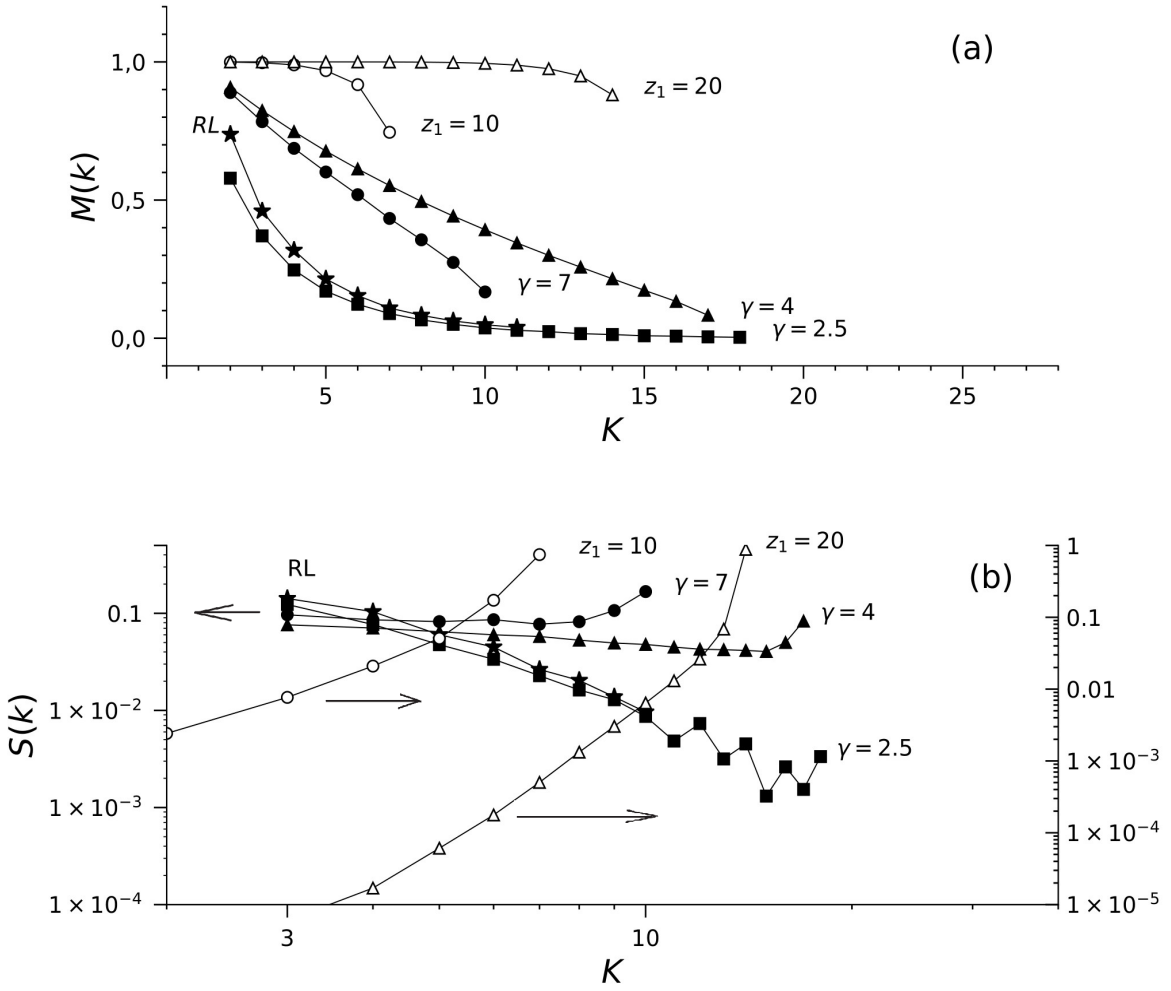


Figure 4.2: The relative sizes of the k-cores, M_k (panel (a)), and k-shells, S_k (panel (b)), in the Erdős-Rényi graphs with $z_1 = 10$ and 20; scale-free networks with $\lambda = 2.5$, 4, and 7; and an uncorrelated network with the degree distribution of the router-level IR. The minimum degree in the scale-free networks is $q_0 = 1$. In the case $\lambda = 2.5$, the maximum degree in the network is $q_{\text{cut}} = 2000$, and $c = 2$; for $\lambda = 4$ and 7, $c = 30$ and 50, respectively. Additonal information could be found in Table 4.1

5 | Appendixes

5.1 | Appendix A: Erdős-Rényi Degree Distribution and Emergence of a Giant Connected Component

The ER model is characterized by two parameters: the number of nodes N and the edge probability p. The choice of these parameters determines the properties of the generated random graph.

The probability distribution of the number of edges connected to a node in an ER graph follows a binomial distribution. The probability $P(k)$ of a node having exactly k edges is given by:

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

where $\binom{N-1}{k}$ represents the binomial coefficient, p is the edge probability, and $1-p$ is the probability of no edge existing between two nodes.

The average degree in an ER graph can be calculated by summing the product of the number of edges k and their corresponding probabilities $P(k)$ over all possible values of k.

π is the probability of finding a node that does not belong to the GCC, and $S = 1 - \pi$ is the probability of finding a node in the GCC.

The probability that i does not belong to the GCC is obtained by assuming that it is not connected to any other node (from others $N-1$) connected to the GCC, i.e., $\pi = (1 - p + p\pi)^{N-1}$. Since $p \approx \langle k \rangle / N$, we can rewrite such a probability as:

$$\pi = \left(1 - (1-\pi)\frac{\langle k \rangle}{N}\right)^{N-1}$$

Taking the logarithm:

$$\log \pi = (N-1) \log \left(1 - \frac{(1-\pi)\langle k \rangle}{N}\right) \approx -(N-1)(1-\pi)\frac{\langle k \rangle}{N} \approx (1-\pi)\langle k \rangle$$

Since $1 - \pi = S$, it follows :

$$S = 1 - e^{-S\langle k \rangle} \quad (3.14)$$

This cannot be solved in closed form but can be solved graphically or numerically. We can have the estimate that $p_c = \frac{1}{N}$: therefore, Erdős-Rényi-Gilbert networks exhibit a GCC only if $p > p_c$, corresponding to $\langle k \rangle > 1$ on average each node must have at least one connection. Also, N/2 links are needed in the network.[4]

5.2 | *Fundamental Equations of K-Core Analysis*

In this part, we aim to explain and illustrate the mathematical aspects of k-core definitions.
We consider an uncorrelated network — a maximally random graph with a given degree distribution $P(q)$ — the so-called configuration model. We assume that a fraction $Q \equiv 1 - p$ of the vertices in this network are removed at random.
Taking into account the treelike structure of the infinite sparse configuration model shows that the k-core coincides with the infinite $(k-1)$-ary subtree. (The m-ary tree is a tree where all vertices have branching at least m). Let R be the probability that a given end of an edge of a network if it is not the root of an infinite $(k-1)$-ary subtree, then a vertex belongs to the k-core if at least k of its neighbors are roots of infinite $(k-1)$-ary subtrees. Therefore, the probability that a vertex is in the k-core is:

$$M(k) = p \sum_{p \geq k} P(q) \sum_{n=k}^{q} C_n^q R^{q-n} (1-R)^n \tag{5.2.1}$$

where $C_n^m = m!/(m-n)!n!$. Note that for the ordinary percolation we must set $k = 1$ in this equation.
An end of an edge is not a root of an infinite $(k-1)$-ary subtree if at most $k-2$ its children branches are roots of infinite $(k-1)$-ary subtrees. This leads to the following equation for R :

$$R = 1 - p + p \sum_{n=0}^{k-2} [\sum_{i=n}^{\infty} \frac{(i+1)P(i+1)}{z_1} C_n^i R^{i-n} (1-R)^n] \tag{5.2.2}$$

Let us explain this equation. (i) The first term, $1 - p \equiv Q$, is the probability that the end of the edge is unoccupied. (ii) $C_n^i R^{i-n}(1-R)^n$ is the probability that if a given end of the edge has i children (i.e., other edges than the starting edge), then exactly n of them are roots of infinite $(k-1)$-ary subtrees. $(i+1)P(i+1)/z_1$ is the probability that a randomly chosen edge leads to a vertex with branching i.$z_1 = \sum_q qP(q)$ is the mean number of the nearest neighbors of a vertex in the graph. Thus, in the square brackets, we present the probability that a given end of the edge has exactly n edges, which are roots of infinite $(k-1)$-ary subtrees. (iii) Finally, we take into account that n must be at most $k-2$. The sum $\sum_{n=0}^{k-2}$ in 5.2.2 may be rewritten as:

$$\Phi_k(R) = \sum_{n=0}^{k-2} \frac{(1-R)^n}{n!} \frac{d^n}{dR^n} G_1(R) \tag{5.2.3}$$

where $G_1(x) = z_1^{-1} \sum_a P(q) q x^{q-1} = z_1^{-1} dG_0(x)/dx$, and $G_0(x) = \sum_q P(q) x^q$. Then 5.2.2 takes the form:

$$R = 1 - p + p\Phi_k(R) \tag{5.2.4}$$

Let us define a function

$$f_k(R) = [1 - \Phi_k(R)] / (1-R) \tag{5.2.5}$$

This function is positive in the range $R \in [0, 1)$ and, in networks with a finite mean number of the second neighbors of a vertex, $z_2 = \sum_q q(q-1)P(q)$, it tends to zero in

the limit $R \to 1$ as $f_k(R) \propto (1-R)^{k-2}$. In terms of the function $f_k(R)$, Eq. 5.2.2 is especially simple:

$$p f_k(R) = 1 \tag{5.2.6}$$

Depending on $P(q)$, with increasing R , $f_k(R)$ either (i) monotonously decreases from $f_k(0) < 1$ to $f_k(1) = 0$, or (ii) at first increases, then approaches a maximum at $R_{\max} \in (0,1)$, and finally tends to zero at $R \to 1$. Therefore 5.2.6 has a non-trivial solution $R < 1$ if

$$p \max_{R\in[0,1)} f_k(R) \geq 1 \tag{5.2.7}$$

This is the criterion for the emergence of the giant k core in a randomly damaged uncorrelated network. The equality in 5.2.7 takes place at a critical concentration $p_c(k)$ when the line $y(R) = 1/p_c(k)$ touches the maximum of $f_k(R)$. Therefore the threshold of the k-core percolation is determined by two equations:

$$p_c(k) = 1/f_k\left(R_{\max}\right), \quad 0 = f_k'\left(R_{\max}\right) \tag{5.2.8}$$

$R_{\max}$ is the value of the order parameter at the birth point of the k-core. At $p < p_c(k)$ there is only the trivial solution $R = 1$.

At $k = 2$, 5.2.4 describes the ordinary percolation in a random uncorrelated graph [3] [9]. In this case, in infinite networks we have $R_{\max} \to 1$, and the criterion 5.2.7 is reduced to the standard condition for existence of the giant connected component: $pG_1'(1) = z_2/z_1 \geq 1$.

Let us find R near the $k \geq 3$-core percolation transition in a network with a finite z_2. We examine 5.2.4 for $R = R_{\max} + r$ and $p = p_c(k) + \epsilon$ with $\epsilon, |r| \ll 1$. Note that at $k \geq 3, \Phi_k(R)$ is an analytical function in the range $R \in [0,1)$. It means that the expansion of $\Phi_k(R+r)$ over r contains no singular term at $R \in [0,1)$. Substituting this expansion into 5.2.4, in the leading order, we find

$$R_{\max} - R \propto \left[p - p_c(k)\right]^{1/2} \tag{5.2.9}$$

i.e., the combination of a jump and the square root critical singularity. The origin of this singularity is an intriguing problem of the hybrid phase transition.

The structure of the k-core is essentially determined by its degree distribution which we find to be

$$P_k(q) = \frac{p}{M(k)} \sum_{q' \geq q} P\left(q'\right) C_q^{q'} R^{q'-q} (1-R)^q \tag{5.2.10}$$

The mean degree of the k-core vertices is $z_1(k) = \sum_{q\geq k} P_k(q) q$. The k-core of a given graph contains the $(k+1)$-core as a subgraph. Vertices which belong to the k core, but do not belong to the $(k+1)$-core, form the k-shell of the relative size $S(k) = M(k) - M(k+1)$. We apply our general results to two basic networks.

Erdős-Rényi (ER) graphs — These random graphs have the Poisson degree distribution $P(q) = z_1^q \exp\left(-z_1\right)/q!$, where z_1 is the mean degree. In this case, $G_0(x) = G_1(x) = \exp\left[z_1(x-1)\right]$. In 5.2.4, $\Phi_k(R) = \Gamma\left[k-1, z_1(1-R)\right]/\Gamma(k-1)$, where $\Gamma(n,x)$ is the incomplete gamma function. From 5.2.1 we get the size of the k-core:

$$M(k) = p\left\{1 - \Gamma\left[k, z_1(1-R)\right]/\Gamma(k)\right\} \tag{5.2.11}$$

where R is the solution of 5.2.4. The degree distribution in the k-core is:

$$P_k(q \geq k) = p z_1^q \frac{(1-R)^q e^{-z_1(1-R)}}{[M(k)q!]} \tag{5.2.12}$$

Our numerical calculations revealed that at $p = 1$, the highest k-core increases almost linearly with z_1, namely, $k_h \approx 0.78 z_1$ at $z_1 \lesssim 500$. Furthermore, the mean degree $z_1(k)$ in the k-core weakly depends on k: $z_1(k) \approx z_1$. See Fig. 4.1

Scale-free networks —We consider uncorrelated networks with a degree distribution $P(q) \propto (q + c)^{-\gamma}$, where q is the degree and c is a constant. Let us start with the case of $\gamma > 3$, where z_2 is finite. It turns out that the existence of k-cores is determined by the complete form of the degree distribution, including its low-degree region. It was proved that there is no $(k \geq 3)$-core in a graph with the minimal degree $q_0 = 1$, $\gamma \geq 3$, and $c = 0$. We find that the k-cores emerge as c increases.

The case $2 < \gamma \leq 3$ is realized in most important real-world networks. With γ in this range, z_2 diverges if $N \to \infty$. In the leading order in $1 - R \ll 1$, 5.2.5 gives $f_k(R) \cong (q_0/k)^{\gamma-2} (1 - R)^{-(3-\gamma)}$. From 5.2.6 we find the order parameter R. Substituting this solution into 5.2.1, in the leading order in $1 - R$ we find that the size of the k-core decreases with increasing k:

$$M(k) = p \left[q_0(1 - R)/k\right]^{\gamma-1} = p^{2/(3-\gamma)} (q_0/k)^{(\gamma-1)/(3-\gamma)} \tag{5.2.13}$$

The divergence of $f_k(R)$ at $R \to 1$ means that the percolation threshold $p_c(k)$ tends to zero as $N \to \infty$. The k-core percolation transition in this limit is of infinite order, similarly to the ordinary percolation. As $k_h(N \to \infty) \to \infty$, there is an infinite sequence of successively enclosed k-cores. One has to remove at random almost all vertices in order to destroy any of these cores.

Equation 5.2.10 allows us to find the degree distribution of k-cores in scale-free networks. For $\gamma > 2$ and $k \gg 1$, we have $P_k(q \gg k) \approx (\gamma - 1)k^{\gamma-1}q^{-\gamma}$. The mean degree $z_1(k)$ in the k-core grows linearly with k: $z_1(k) \approx k z_1/q_0$, in contrast to the Erdős-Rényi graphs.

Finite-size effect —The finiteness of the scale-free networks with $2 < \gamma < 3$ essentially determines their k-core organization. We introduce a size-dependent cutoff $q_{\text{cut}}(N)$ of the degree distribution. Here $q_{\text{cut}}(N)$ depends on details of a specific network. For example, for the configuration model without multiple connections, the dependence $q_{\text{cut}}(N) \sim \sqrt{N}$ is usually used if $2 < \gamma < 3$. It is this function that must be substituted into Eqs. (14)-(16) below. A detailed analysis of Eq. 5.2.5 shows that the cutoff dramatically changes the behavior of the function $f_k(R)$ near $R = 1$. $f_k(R)$ has a maximum at $R_{\max} \cong 1 - (3 - \gamma)^{1/(\gamma-2)} k/q_{\text{cut}}$ and tends to zero as $R \to 1$ instead of divergence. As a result, the k-core percolation again becomes the hybrid phase transition. The cutoff determines the highest k-core:

$$k_h \cong p(\gamma - 2)(3 - \gamma)^{\frac{3-\gamma}{\gamma-2}} q_{\text{cut}} \left(\frac{q_0}{q_{\text{cut}}}\right)^{\gamma-2} \tag{5.2.14}$$

The sizes of the k-core at $q_0 \ll k \ll k_h$ are given by 5.2.13 . The relative size of the highest k-core is

$$M\left(k_h\right) \cong p \left[(3 - \gamma)^{-(\gamma-1)/(\gamma-2)} - 1\right] \left(\frac{q_0}{q_{\text{cut}}}\right)^{\gamma-1}. \tag{5.2.15}$$

Finally, the threshold of the k-core percolation is:

$$p_c(k) = \frac{1}{f_k\left(R_{\max}\right)} \cong \frac{k}{k_h}. \tag{5.2.16}$$

If $k \to k_h$, then $p_c(k) \to 1$, i.e., even minor random damage destroys the highest k_h-core. See Fig. 4.2, these curves agree with asymptotic expressions 5.2.13 and 5.2.15.

5.3 | *Scale-Free Degree Distribution*

Here, we present the derivation of the estimation for the total number of nodes in a scale-free network with a discrete degree distribution. The probability function for a node having degree q is defined as:

$$P(q) = A(q + c)^{-\gamma} \tag{5.3.1}$$

Here, c represents a constant term, and γ denotes the power-law coefficient of the distribution. Also, A is a constant that can be determined by satisfying the condition:

$$\sum_{q=q_{\min}}^{\infty} P(q) = 1 \tag{5.3.2}$$

To estimate the total number of vertices, we consider the requirement that there must be at least one node with a degree equal to or greater than q_{cut}. This leads us to the following equation:

$$N \sum_{q=q_{\text{cut}}}^{\infty} P(q) = 1 \tag{5.3.3}$$

To evaluate this expression, we introduce the concept of the incomplete zeta function:

$$\zeta(\gamma, c) = \sum_{x=1}^{c} \frac{1}{x^{\gamma}} \tag{5.3.4}$$

By solving 5.3.3 and using 5.3.4, our estimation for the total number of nodes, denoted as N, can be expressed as:

$$N = \left\lceil \frac{\zeta(\gamma, \infty) - \zeta(\gamma, c)}{\zeta(\gamma, \infty) - \zeta(\gamma, q_{\text{cut}} - 1)} \right\rceil \tag{5.3.5}$$

In the above equation, the symbol $\lceil \cdot \rceil$ represents the ceiling function, which rounds up to the nearest integer. Note that this estimation was used to produce Scale-free random networks described in Table 4.1.

5.4 | Assoratative mixing network

The correctly normalized distribution qk of the remaining degree is then

$$q_k = \frac{(k+1)p_{k+1}}{\sum_j j p_j} \tag{5.4.1}$$

5.4.1 Gnerative function

Turning now to theoretical developments, Newman proposed a simple model of an assortatively mixed network, which is exactly solvable for many of its properties in the limit of large graph size. Consider the ensemble of graphs in which the distribution e_{jk} takes a specified value. This defines a random graph model similar in concept to therandom graphs with specified degree sequence, except for the added element of assortative mixing. Consider a typical member of this ensemble in the limit of large graph size, and consider a randomly chosen edge in that graph, one end of which is attached to a vertex of degree j. We ask what the probability distribution is of the number of other vertices reachable by following that edge. Let this probability distribution be generated by a generating function G_j (x), which depends in general on the degree j of the starting vertex.

$$G_j(x) = x\frac{\sum_k e_{jk}\left[G_k(x)\right]^k}{\sum_k e_{jk}} \tag{5.4.2}$$

Acoording to [8] G_j(x) must satisfy a self-consistency condition of the form5.4.3

$$H(x) = xp_0 + x\sum_{k=1}^{\infty} p_k\left[G_{k-1}(x)\right]^k \tag{5.4.3}$$

The average size of the component to which such a vertex belongs is given by the derivative of H which is shown in 5.4.4.

$$\langle s\rangle = H'(1) = 1 + \sum_k k p_k G'_{k-1}(1) \tag{5.4.4}$$

To calculate the size S of the giant component, we define u_k to be the probability that an edge connected to a vertex of the remaining degree k leads to another vertex that does not belong to the giant component which has drived from the generative model

$$S = 1 - p_0 - \sum_{k=1}^{\infty} p_k u_{k-1}^k \tag{5.4.5}$$

$$N = \frac{1}{2}\left(1 - e^{-\frac{1}{k}}\right) \tag{5.4.6}$$

6 | Bibliography

[1] Dimitris Achlioptas, Raissa M. D'Souza, and Joel Spencer. Explosive percolation in random networks. *Science*, 323(5920):1453–1455, 2009. doi: 10.1126/science.1167782.

[2] Béla Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001. doi: 10.1017/CBO9780511814068.

[3] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Physical Review Letters*, 85(21):4626–4628, nov 2000. doi: 10.1103/physrevlett.85.4626. URL https://doi.org/10.1103%2Fphysrevlett.85.4626.

[4] Manlio De Domenico. *PHYSICS OF COMPLEX NETWORKS: STRUCTURE DYNAMICS*. 2023.

[5] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. mml:math xmlns:mml="http://www.w3.org/1998/math/MathML" display="inline"mml:mik/mml:mi/mml:math-core organization of complex networks. *Physical Review Letters*, 96(4), feb 2006. doi: 10.1103/physrevlett.96.040601. URL https://doi.org/10.1103%2Fphysrevlett.96.040601.

[6] Paul Erdos and Alfred Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.*, 5:17–61, 1960.

[7] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89 (20):208701, 2002.

[8] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001. doi: 10.1103/PhysRevE.64.026118.

[9] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3):2566–2572, 2002. ISSN 00278424. URL http://www.jstor.org/stable/3057594.

[10] John Scott. *Social Network Analysis: A Handbook*. Sage Publications, London, 2nd edition, 2000. doi: 10.1002/cpe.607.