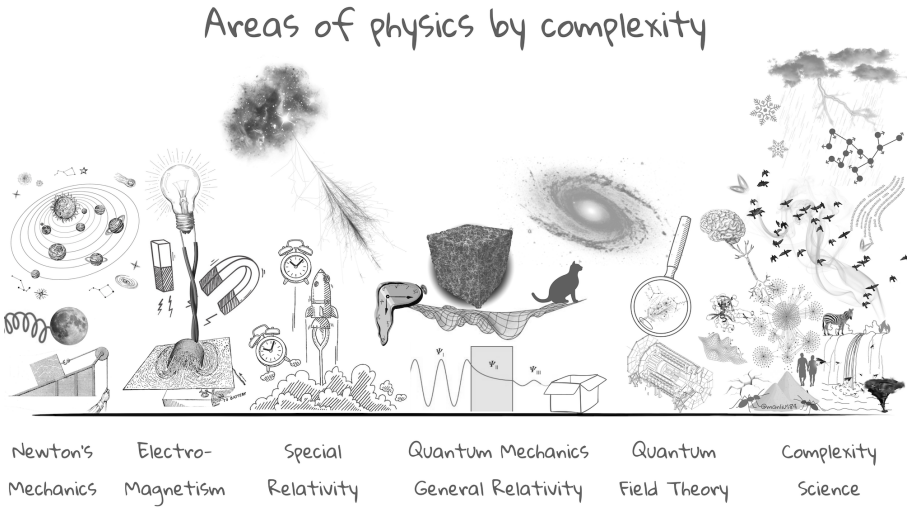
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



Project 13: Popularity versus Similarity in growing networks
Project 27: Robustness of Noisy Quantum Networks
Project 39: Subways I

Roccatello Mattia

Last update: January 23, 2025

Contents

1 | Project 13: Popularity versus Similarity in growing networks

Task leader(s): *Roccatello Mattia*

1.1 | Networks generation: three models

The report *'Popularity versus similarity in growing networks'* [2] describes the implementation of a network generation model that integrates the concepts of popularity and similarity. The model evolves by adding nodes iteratively, connecting new nodes to existing ones based on their popularity and similarity. Popularity is proportional to the logarithmic growth of time, while similarity depends on the angular coordinates assigned to nodes at their creation. Three variations of the model, referred to as **Model 1**, **Model 2**, and **Model 2'**, are implemented and described below. In **Model 1**, a new node t is introduced at each time step with coordinates:

$$r_t = \log(t), \quad \theta_t \sim U(0, 2\pi),$$

while all the others are effected by the popularity fading phenomenon: $r_s = \beta r_s + (1 - \beta) r_t, \forall s < t$. To determine the connections, the hyperbolic distance between the new node t and each existing node s is approximated as:

$$x_{s,t} = \log(s \cdot t \cdot \theta_{s,t}/2),$$

where $\theta_{s,t}$ is the angular distance given by:

$$\theta_{s,t} = \pi - |\pi - |\theta_s - \theta_t||.$$

The new node is connected to the m nodes with the smallest hyperbolic distances $x_{s,t}$. If $m + 1$ nodes are not yet available, the new nodes connect to all the present nodes. **Model 2** introduces a probabilistic wiring process influenced by a temperature parameter T. The connection probability between node t and an existing node s is computed (when $T > 0$) using:

$$P_{s,t} = \frac{1}{1 + \exp\left(\frac{x_{s,t} - R_t}{T}\right)}$$

Here, R_t represents a threshold hyperbolic radius calculated based on r_t and model parameters:

$$R_t = r_t - \log\left(\frac{2T(1 - e^{-(1-\beta)r_t})}{\sin(T\pi)m(1-\beta)}\right),$$

for $T \neq 0$ and $\beta \neq 1$. Boundary cases for $T = 0$ or $\beta = 1$ are handled separately and explained in the Appendix .1. Again, for $t < m + 1$, new nodes are connected to all existing nodes. **Model 2'** builds on Model 2 but iterates over all existing nodes just once, and creates a connection with probability $P_{s,t}$ as above, in order to create *on average* m connections - this can lead to less or more than exactly m connections with a new node.

1.2 | Results

1.2.1 Analysis of Model 1

As in the main article, in order to study some key properties of this ensemble, a popularity × similarity network has been grown reaching $N = 10^5$ nodes, with $m = 2$ and $\gamma = 2.1$ or $\beta = \frac{1}{\gamma-1} = \frac{1}{1.1}$. The first quantity investigated is the attraction probability, defined as the probability that a node with degree k attracts a connection from a new node (see fig. 1.1). Before simulating the introduction of new nodes, the final popularity fading phenomenon was applied to all existing nodes in the network. After, 10000 new nodes were virtually introduced, each with a radial coordinate $r_t = \ln t$ and θ_t chosen from of the interval $(0, 2\pi)$, evenly spaced. For each new node, the hyperbolic distance to all existing nodes in the network has been computed and the degrees of the $m = 2$ closest nodes were recorded. The total number of connections made to nodes of degree k was binned into a histogram. To compute the attraction probability for a given degree k, the total number of connections to nodes of degree k was divided by the total number of nodes in the network with degree k, to ensure that the attraction probability is normalized to account for the varying abundance of nodes with different degrees. Another quantity to investigate is the connection probability as a function of the hyperbolic distance, on the same network grown to $N = 10^5$ nodes. For each pair of connected nodes, their distance is computed using exact formula reported in the Appendix .1.3. The hyperbolic distances are then grouped into discrete bins and, for each bin, the connection probability is obtained as

$$P(x) = \frac{\text{Number of connected pairs with } x \in [x_i, x_{i+1})}{\text{Total pairs with } x \in [x_i, x_{i+1})},$$

where the interval $[x_i, x_{i+1})$ represents the current bin of the hyperbolic distances histogram. This process involves counting the number of connected node pairs within a given range of hyperbolic distances and dividing it by the total number of pairs in the same range; the results are shown in figure 1.2.

Furthermore, in the Supplementary Notes of the paper, the degree distribution and the average clustering distribution - here meant as the average clustering of nodes having a certain degree - of the same graph are computed. These results can be found in figure 1.3 and figure 1.4.

1.2.2 Analysis of Model 2 and Model 2'

As for Model 1, the attraction probability is computed. Following Model 2, a network with $N = 10^4$ nodes has been grown and, 10000 new nodes have been created to

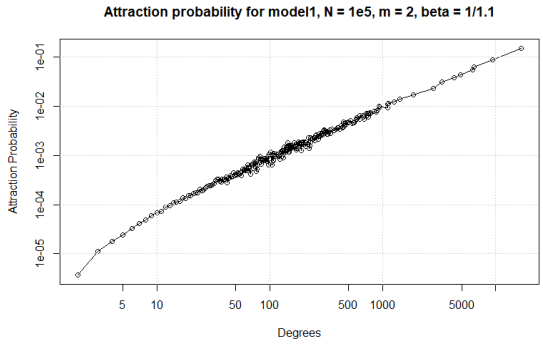


Figure 1.1: Attraction probability vs degree of nodes in a popularity × similarity network, grown up to N = 10^5, with $m = 2$ and $\beta = 1/1.1$.

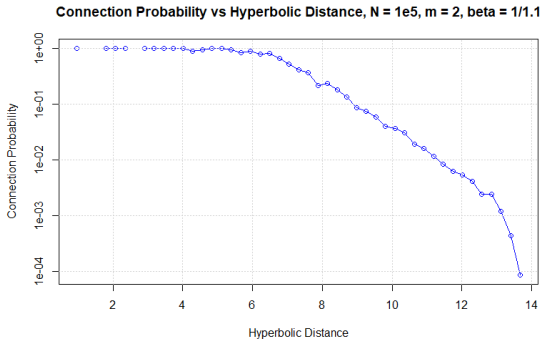


Figure 1.2: Connection probability for different intervals of the hyperbolic distance, of the same popularity × similarity network.

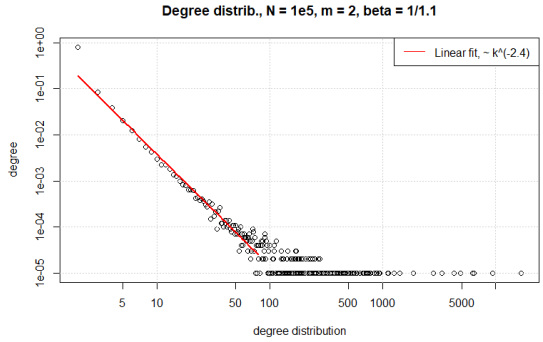


Figure 1.3: Degree distribution in a popularity x similarity network; the expected power law, from theory, is scaling with $k^{-2.1}$.

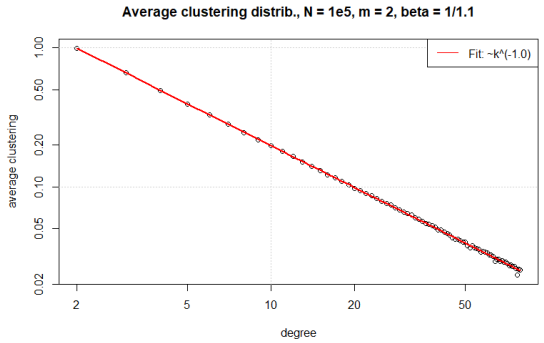


Figure 1.4: Average clustering of k-degree nodes; the $1/k$ behavior is caught in the experimental results.

collect the degrees of the nodes to whom they connect, using the stochastic formula provided above (see figure 1.5). The degree distribution is also computed (figure 1.6). Now that the connection between new nodes and already existing ones is stochastic, depending on a free parameter T, the average clustering for varying temperatures has been investigated (figure 1.7). As these plots show, when $T = 0$ the two models (1 and 2) coincide. In fact, it can be proven that when $T = 0$, the average clustering is the maximum possible for a network with that degree distribution. This result is highlighted in the Supplementary Notes and here in figure 1.8. The same metrics are shown for the networks grown according to Model 2' in figures 1.9, 1.10 and 1.11.

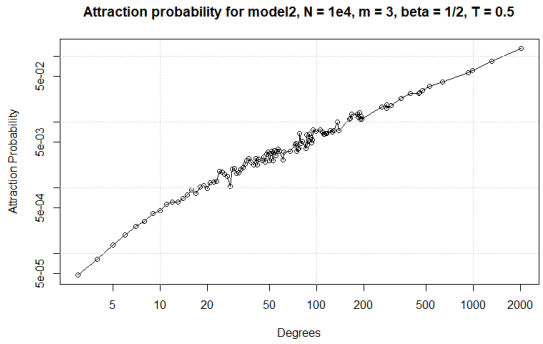


Figure 1.5: Attraction probability in a popularity × similarity network with Model 2.

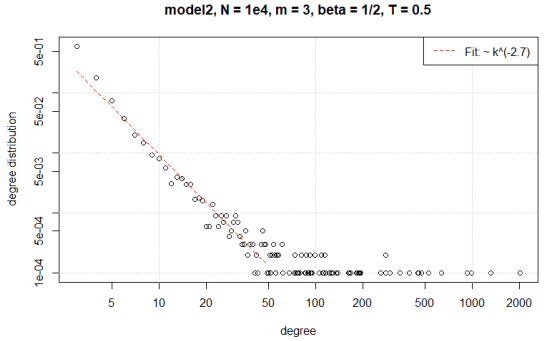


Figure 1.6: Degree distribution of the same network.

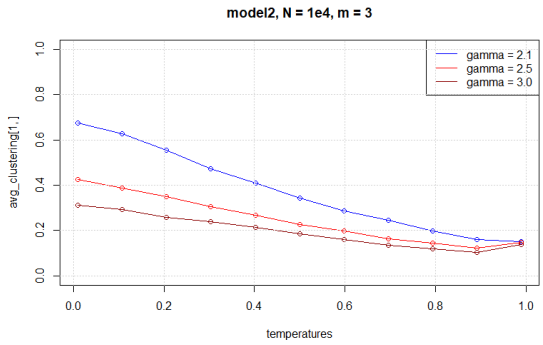


Figure 1.7: Average clustering at $t = 10^3$ as a function of the temperature; notice the decreasing behavior with a maximum reached for $T = 0$.

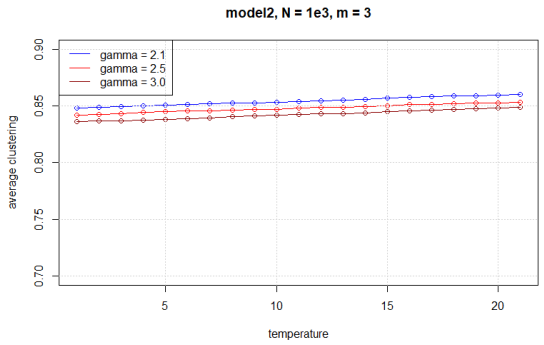


Figure 1.8: Average clustering as a function of the number of accepted clustering-increasing rewirings in networks grown according to Model 1.

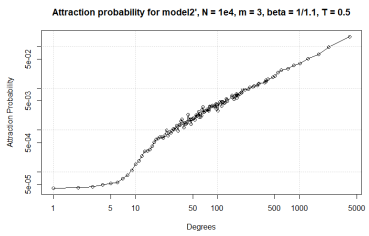


Figure 1.9: Attraction probability vs degree for network grown up to N = 10^4, with $m = 3$, $\beta = 1/1.1$ according to Model 2'.

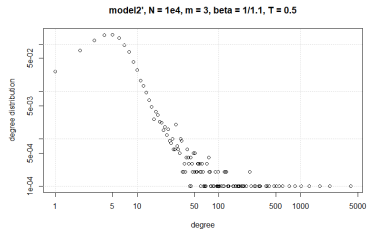


Figure 1.10: Degree distribution of the network grown according to Model 2'. The shape of the distribution is supported form theory in Supplementary Notes.

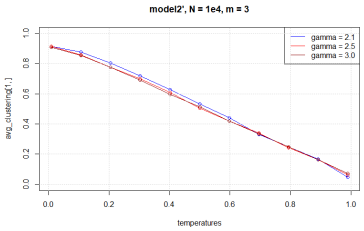


Figure 1.11: Average clustering as a function of the temperature for a network grown using Model 2'.

2 | Project 27: Robustness of Noisy Quantum Networks

Task leader(s): *Roccatello Mattia*

2.1 | Simulating bond percolation in quantum networks

This report presents the replication of the results of the article *'Robustness of Noisy Quantum Networks'* [1], focusing on the mathematical and computational framework underlying the study of bond percolation in quantum networks. Quantum networks enable the transmission of quantum information by linking nodes via entangled pairs of qubits. These networks are modeled using classical graph structures, where links represent quantum entanglement between nodes. The study investigates the robustness of these networks under bond percolation.

A key mathematical feature of the article is the assumption that the number of entangled pairs per link follows an exponential distribution:

$$g(n) = \frac{1}{\langle n \rangle} e^{-n/\langle n \rangle},$$

where $\langle n \rangle = (15 \ln N)^{\alpha}$ is the mean number of entangled pairs, N is the number of nodes, and α is a scaling parameter that controls the relationship between N and $\langle n \rangle$.

The operational probability of a link, denoted as $p^{op}(l^{op})$, is the probability that a link remains functional considering quantum constraints: $n > (l^{op})^{\alpha}$, to have the link functioning, where l^{op} is the operational distance. It is given by:

$$p^{op}(l^{op}) = \int_{n^{op}}^{\infty} g(n)\, dn,$$

where $n^{op} = (l^{op})^{\alpha}$ is the threshold number of entangled pairs required for the link to remain operational. Substituting $g(n)$, and rearranging for l^{op}:

$$p^{op}(l^{op}) = \exp\left(-\frac{(l^{op})^{\alpha}}{\langle n \rangle}\right) \rightarrow l^{op}(p^{op}) = \Big(-\langle n \rangle \ln p^{op}\Big)^{\frac{1}{\alpha}}$$

To account for both quantum constraints and random failures, the survival probability of a link, p, is defined as: $p = p^{op} \cdot p^{ext}$, where p^{ext} is the external probability of a link surviving random failures, independent of the quantum properties of the network. The effective diameter of the network, $D(p)$, is defined as the longest shortest path in the largest connected component, normalized by $\ln N$. It serves as a measure of the

network's scale for a given survival probability p. To determine the critical external probability p^{ext}, the equation to solve is:

$$D(p) = l\left(\frac{p}{p^{ext}}\right),$$

where $l(p/p^{ext})$ represents the operational distance associated with the operational probability $p^{op} = p/p^{ext}$.

Finally, the last metric to measure networks' robustness is the quantum backbone size (QBS), defined as the fraction of nodes in the largest connected component of the network that remains functional after percolation.

2.2 | Results

2.2.1 Data Generation

In this analysis, bond percolation was applied to both Erdős–Rényi and Barabási–Albert networks to systematically study the effects of varying network size N, percolation probability p, and average degree $c = 6,\ 8$ for E.R. networks, and $c = 6$ for B.A. networks. For each value of p, the diameter $D(p)$ and the QBS were computed as dynamically updated averages over 15 iterations. This averaging reduces noise in the data, especially for larger networks and higher percolation probabilities. However, the process significantly increases execution time. For small networks ($N < 10^4$) and Barabási–Albert networks, the QBS metric exhibits higher noise due to the variability of the largest connected component. The network sizes N were logarithmically spaced, ranging from $10^{2.5}$ to $10^{4.5}$ for Erdős–Rényi networks. The percolation probabilities were divided into two regions for sampling: fine-grained values from $p = 0$ to $p = 0.4$ and coarser values from $p = 0.42$ to $p = 1$. This approach ensured a balance between resolution and computational efficiency. To account for computational constraints, the results were saved to external files since the execution time for networks with $N > 10^4$ was prohibitive, often taking several hours to compute the required metrics.

2.2.2 Critical probabilities

In this analysis, the robustness of Erdős–Rényi (ER) and Barabási–Albert (BA) networks under bond percolation was studied by numerically solving the equation $D(p) = l(p/p^{ext})$ for p^{ext}. $D(p)$ represents the network diameter, and $l(p/p^{ext})$ is the operational length of the network, which depends on the combined survival probability of links p/p^{ext}. The value of p^{ext} corresponds to the external probability at which the network's diameter matches its operational length, encapsulating the interplay between quantum constraints and percolation dynamics. To solve this equation efficiently, the bisection method was employed. This method is particularly suited for finding roots of nonlinear equations within a bounded interval. Starting with an initial range $[a, b]$ for p^{ext}, the midpoint $c = (a + b)/2$ was iteratively updated. At each step, the sign of the function $f(p^{ext}) = D(p) - l(p/p^{ext})$ determined whether p^{ext} lay in the left or right subinterval. This process ensured convergence to the solution with a predefined tolerance. The iterative procedure was repeated for all values of p, yielding the dataset

$p^{ext}(p)$, which is central to characterizing the structural transitions of the network under percolation. Critical probabilities p_c^1 and p_c^2 were identified using the computed $p^{ext}(p)$ and the Quantum Backbone Size (QBS). The algorithm first searched for local maxima and discontinuities in QBS as a function of p^{ext}: by definition, they represent the boundary of the interval in which the transition probability occurs. By iterating over the range of p^{ext}, the differences between consecutive values of $p^{ext}(p)$ were analyzed to detect abrupt changes, which were then used to assign p_c^1 and p_c^2.

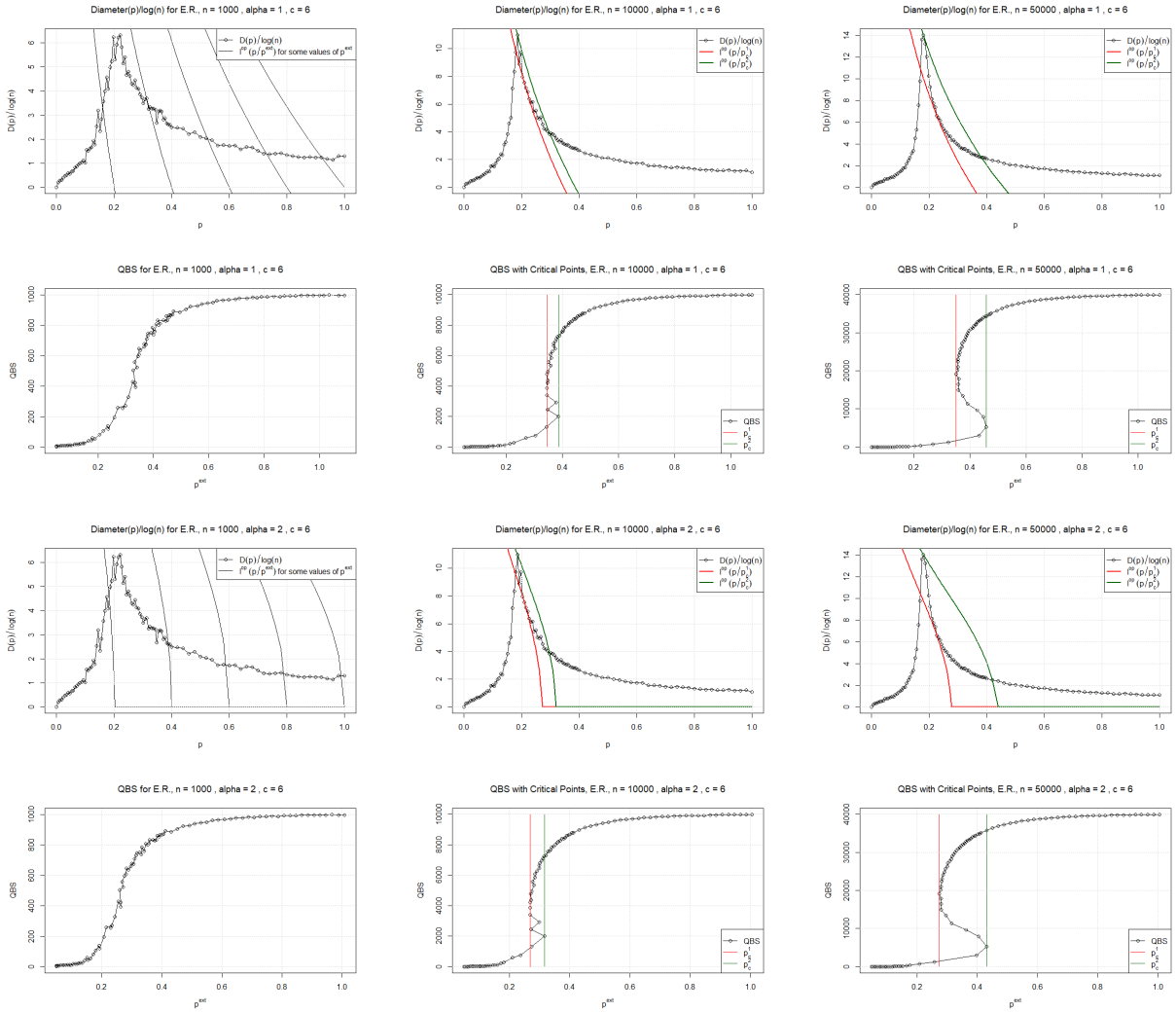


Table 2.1: Quantum Backbone Size vs percolation probability and diameter vs p^{ext} for Erdős–Rényi (ER) networks with average degree $c = 6$ and number of nodes $N \in (10^3, 10^4, 5 \cdot 10^4)$, $\alpha \in (1, 2)$.

2.2.3 Phase diagrams

To analyze the dependence of the critical probabilities on the average number of entangled qubits in the network, a phase diagram was computed. The phase diagram represents the variation of p_c^1 and p_c^2 as functions of the average number of entangled pairs per node, normalized by $\log(n)$. For each average entangled pair value n_{mean}, the external probabilities $p^{ext}(p)$ were computed by solving $D(p) = l(p/p^{ext})$ using the

bisection method. Subsequently, p_c^1 and p_c^2 were identified by analyzing the Quantum Backbone Size (QBS) for abrupt changes and transitions in the network structure. This process was repeated across a range of n_{mean} values, generating datasets of p_c^1 and p_c^2. The phase diagram shows how the critical probabilities evolve with increasing quantum resources, providing insight into the network's robustness and structural transitions under varying entanglement conditions. Some noisiness in the data led to an abrupt separation in the phase diagrams, while the distance between p_c^1 and p_c^2 should be slighter.

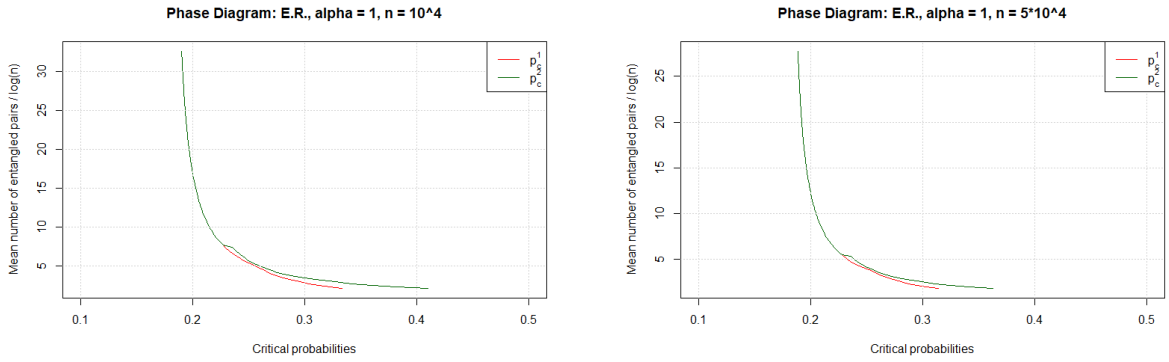


Table 2.2: Phase Diagrams obtained from varying the value of $< n >$ in the $l^{op}(p)$ definition, for the E.R. networks with $N = 10^4$ and $N = 5 \cdot 10^4$, $\alpha = 1$.

2.2.4 Maximum Diameter

Subsequently, some key metrics, such as the maximum network diameter, were evaluated for networks of varying sizes (N) and average degrees (c). For each network, results were saved and reused to avoid redundant computations, especially for larger networks where execution times were prohibitive. For Erdős–Rényi networks, cases with $c = 6$ and $c = 8$ were considered, while for Barabási–Albert networks, the average degree was fixed at $c = 6$. The scaling behavior of critical probabilities and maximum diameters with respect to the network size was also examined, providing insights into the structural transitions of these networks under bond percolation.

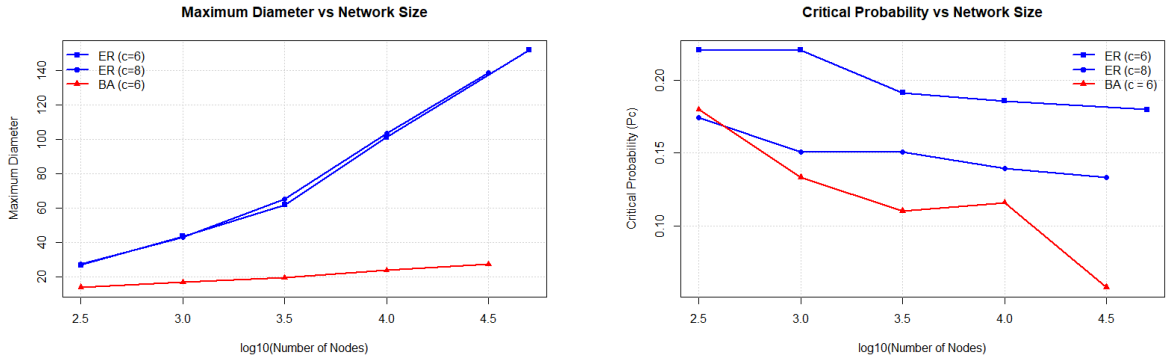


Table 2.3: Maximum Diameter and critical probability vs network sizes for E.R. and Barabasi-Albert (B.A.) networks.

3 | Project 39: Subways I

Task leader(s): *Roccatello Mattia*

3.1 | The Subways I dataset

In the dataset, the topologies of subway systems from eight major cities around the world are collected: Moscow, Tokyo, New York, Osaka, Paris, Seoul, and Shanghai have been analyzed. These datasets span multiple years, capturing the structural evolution of these transportation networks. Each pair of cities (here, nodes) is associated with the name of the line it belongs to. Some of this information, especially for older topologies, has been found to be missing. A placeholder for these missing labels has been put for each Subway's topology. The dataset contains raw topological information for each city. For New York, only the data from the "NewNYC" folder were used, as some historical data from the "NYC" folder are incomplete. The years covered for each city are summarized in Table 3.1

City	Moscow	Tokyo	New York	Osaka	Paris	Seoul	Shanghai
First Year	1936	1928	1880	1934	1901	1975	1996
Last Year	2009	2009	2009	2010	2009	2009	2010

Table 3.1: Years of available topology information for subway networks of selected cities.

3.2 | Results

3.2.1 Networks reconstruction

Subway network data are processed by systematically reading and structuring raw files for nodes, lines, and edges across multiple cities and years. The workflow involves three main steps. First, node information is extracted using the `read_nodes` function, which retrieves node labels, geographical coordinates, and years from city-specific files, compiling them into data frames that structure information for each city. Second, line information is processed using the `read_lines` function, which maps node pairs to specific subway lines, creating a connection network for each city and associating each line with its constituent nodes. Finally, edge topologies are constructed with the `read_topologies` function, which generates yearly adjacency data by integrating node IDs and line associations, capturing the structural evolution of the network over time. A sample of the reconstructed networks is shown in 3.1, with and without node labels.

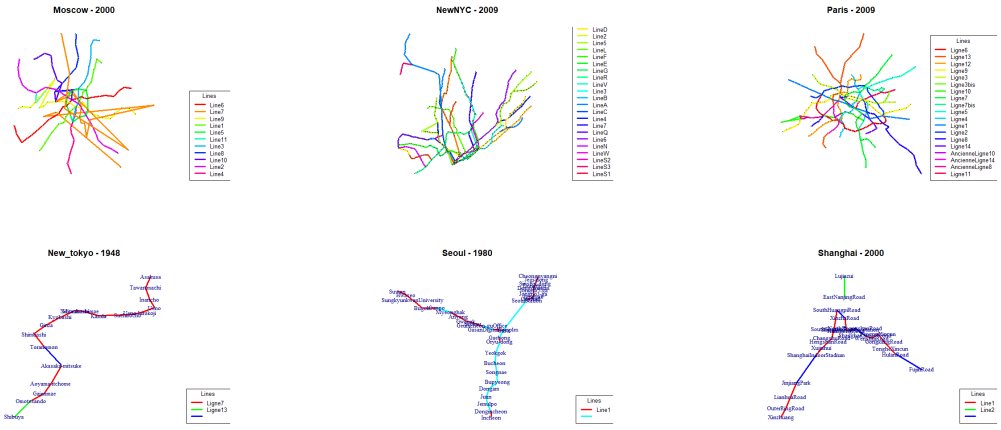


Figure 3.1: Subways of some cities ad different years. Legend gathers each edge membership to its own line. Missing labels in the legend stand for couple of cities not present in the edge files to assess the right membership.

3.2.2 Structural and robustness analysis

In order to investigate the structural evolution and characteristics of subway networks over time, a comprehensive analysis of various network metrics was conducted. The goal is to capture insights into the growth, connectivity, and robustness of these systems. Metrics such as degree distribution, centrality measures, density, clustering coefficients, and path lengths were computed to understand the underlying structure and dynamics of the networks. Additionally, mixing patterns and degree assortativity were analyzed to assess node interaction tendencies. Community detection algorithms were applied to uncover modular structures and evaluate network organization, while robustness was tested through simulated percolation processes. All the results are shown in 3.2, 3.3, 3.4.

3.2.3 What if... simulation of an epidemic

In a hypothetical post-apocalyptic world, humanity is forced to live underground in subway stations. Each station functions as a self-contained community, with movement between them restricted to subway lines. This study simulates the spread of a contagious disease in such a network, accounting for reinfection scenarios and limited mobility.

The simulation is based on a time-dependent SIR (Susceptible-Infected-Recovered) model, which incorporates mobility dynamics using a mobility flow matrix, Φ, derived from the subway's adjacency structure. The model captures the evolution of disease states over time across the network. The state variables in the model are: $S_i(t)$ (susceptible individuals in patch i at time t), $I_i(t)$ (infected individuals in patch i at time t) and $R_i(t)$ (recovered individuals in patch i at time t). The dynamics of the disease

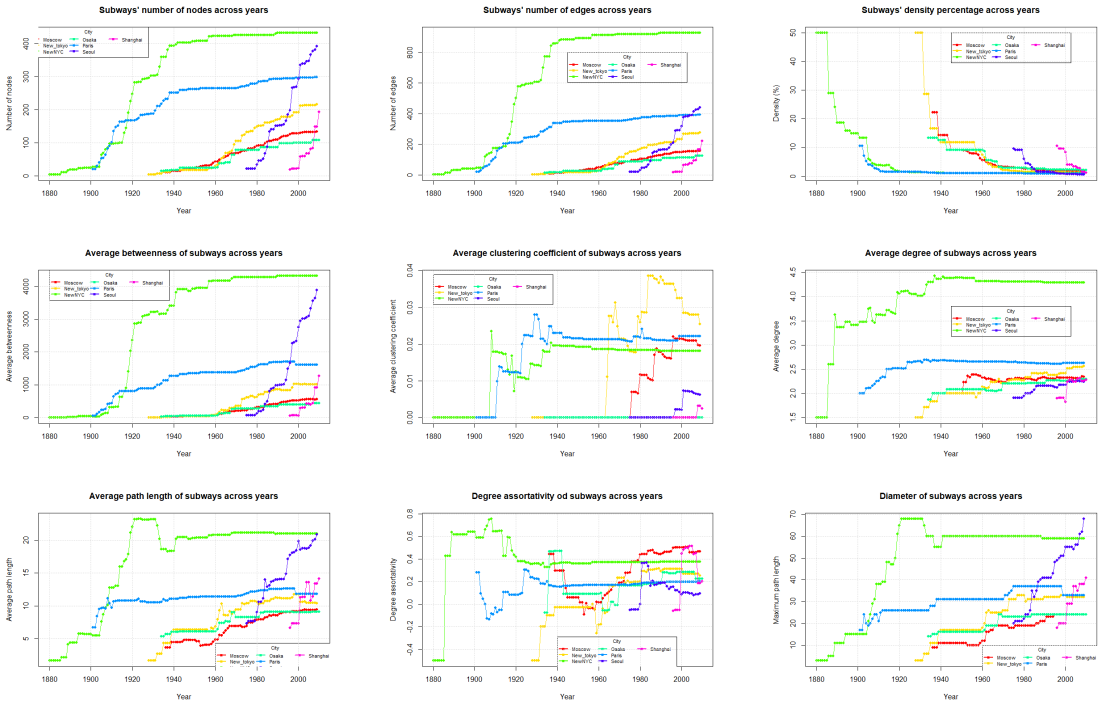


Figure 3.2: Collection of some structural/topological measures for subway of all cities across years.

spread are governed by the following differential equations for each station i:

$$\frac{dS_i}{dt} = -\beta \frac{S_i I_i}{N_i} + \rho R_i + \frac{1}{\epsilon}\left(\sum_j \Phi_{ji} S_j - \sum_j \Phi_{ij} S_i\right),$$

$$\frac{dI_i}{dt} = \beta \frac{S_i I_i}{N_i} - \gamma I_i + \frac{1}{\epsilon}\left(\sum_j \Phi_{ji} I_j - \sum_j \Phi_{ij} I_i\right),$$

$$\frac{dR_i}{dt} = \gamma I_i - \rho R_i + \frac{1}{\epsilon}\left(\sum_j \Phi_{ji} R_j - \sum_j \Phi_{ij} R_i\right),$$

The parameters of the model are defined as follows: β is the transmission rate, γ the recovery rate, ρ the re-susceptibility rate (the rate at which recovered individuals become susceptible again), and ϵ the mobility scaling factor, which determines the influence of movement between stations. The mobility flow matrix entry Φ_{ij} represents the fraction of movement from station i to station j, is derived from the subway graph's adjacency matrix. It is symmetric (ensuring bidirectional flow), positive only where edges exist (reflecting movement along subway connections), and doubly stochastic (normalized so that row and column sums equal 1, ensuring conservation of the total population). Finally, $N_i = S_i + I_i + R_i$ represents the total population of station i. The populations' evolution is plotted in 3.5, while an animated GIF is available in the main folder.

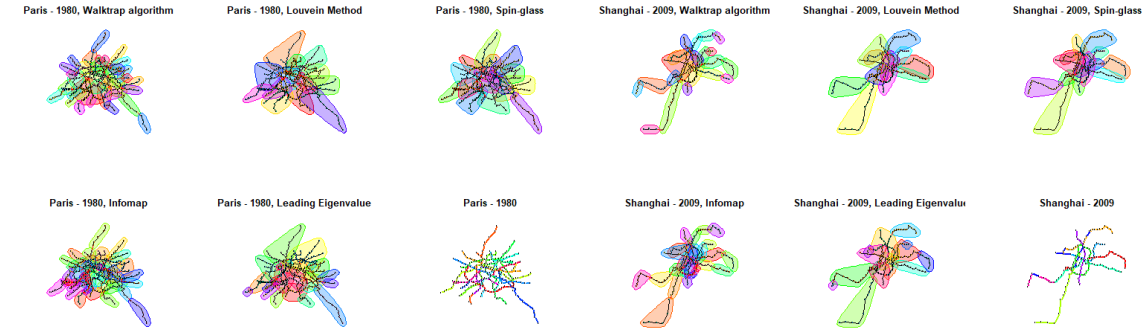


Figure 3.3: Community detection of some networks using different algorithms, with the purpose of detecting districts within the cities.

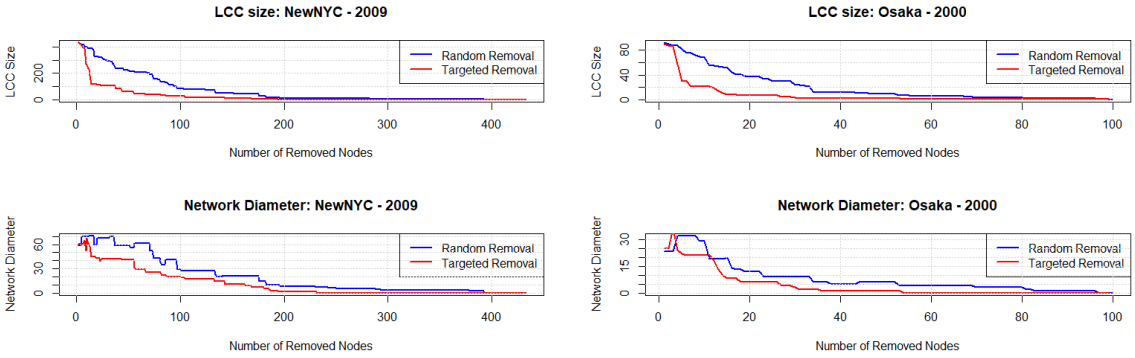


Figure 3.4: LCC size and Diameter vs number of deleted nodes, following a percolation protocol or targeted attack (on higher degree nodes), for Subways of New York City (2009) and Osaka (2000)

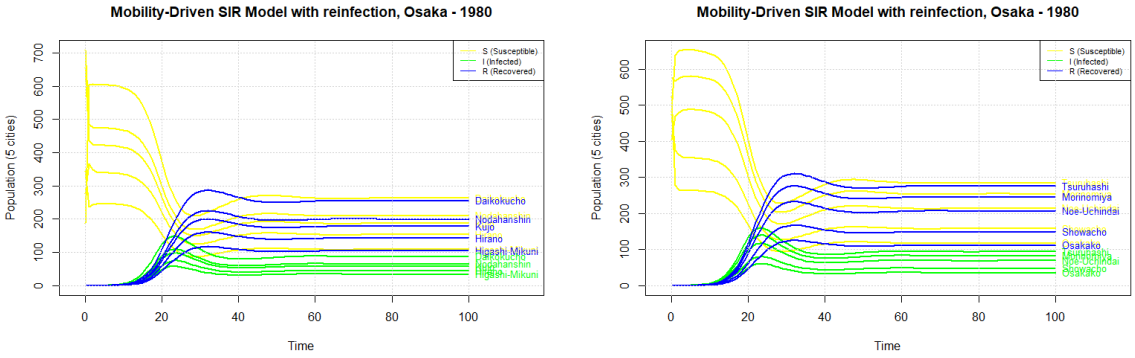


Figure 3.5: Evolution of the three populations for a simulated epidemic in a post-apocalyptical Osaka across some stations of the full subway for better reading of results. Year 1980 has been chosen to speed up the computation, because of the lower number of nodes in the network. Initial conditions: Population N_i in each station is sampled from a normal distribution with a mean of 500 and a standard deviation of 200. A small number of infected individuals is introduced in one randomly chosen station. **Parameters:** $\beta = 0.5$, $\gamma = 0.3$, $\rho = 0.1$, and $\epsilon = 0.01$. The system is simulated over 100 time steps.

4 | Appendix

.1 | Project 13

.1.1 Special Cases for R_t in `model2` and `model2'`

The computation of R_t depends on the values of T (temperature) and β (growth exponent). Four cases need to be considered, that can be found solving the limits of R_t for $T \to 0$ and $\beta \to 1$:

- When $T = 0$ and $\beta = 1$:

$$R_t = \log\left(\frac{2r_t}{\pi m}\right)$$

- When $T = 0$ and $\beta \neq 1$:

$$R_t = \log\left(\frac{2(1 - e^{-(1-\beta)r_t})}{\pi m(1-\beta)}\right)$$

- When $T \neq 0$ and $\beta = 1$:

$$R_t = \log\left(\frac{2Tr_t}{\sin(T\pi)m}\right)$$

- When $T \neq 0$ and $\beta \neq 1$:

$$R_t = \log\left(\frac{2T(1 - e^{-(1-\beta)r_t})}{\sin(T\pi)m(1-\beta)}\right)$$

.1.2 Special Cases for Wiring Probabilities in `model2` and `model2'`

The wiring probability, $P_{s,t}$, depends on the hyperbolic distance $x_{s,t}$, R_t, and T. Four cases need to be considered:

- When $T > 0$:

$$P_{s,t} = \frac{1}{1 + e^{\frac{x_{s,t} - R_t}{T}}}$$

- When $T = 0$ and $x_{s,t} - R_t < 0$:

$$P_{s,t} = 1$$

- When $T = 0$ and $x_{s,t} - R_t = 0$:

$$P_{s,t} = 0.5$$

- When $T = 0$ and $x_{s,t} - R_t > 0$:

$$P_{s,t} = 0$$

.1.3 Hyperbolic distance

Given the radial and angular coordinates of two nodes created at times s and t, that are (r_s, θ_s) and (r_t, θ_t) respectively, the exact hyperbolic distance is computed as:

$$x_{s,t} = \frac{1}{2}\text{settcosh}(\cosh 2r_s \cosh 2r_t - \sinh 2r_s \sinh 2r_t \cos\theta_{s,t})$$

where $\theta_{s,t} = \pi - |\pi - |\theta_s - \theta_t||$ is the angular distance between the two points.

.1.4 Key Differences from the Main Article

A comparison of Figures 1.1 and 1.2 with their counterparts in the main article highlights a noticeable difference in the level of noise present in the plotted data. This discrepancy is likely attributable to the absence of averaging across multiple instances of popularity × similarity networks with identical parameters. While performing an average over additional networks would have required significantly more computational resources, the key trends—namely, the proportionality to k observed in Figure 1.1 and the decreasing behavior starting from a probability equal to 1 in Figure 1.2—remain evident even without averaging.

.2 | Project 37

.2.1 Key Differences from the Main Article

To perform bond percolation and compute the largest connected component (LCC) size and the network diameter, the implementation requires computationally intensive calculations. To optimize runtime and manage computational resources, several adjustments were made compared to the methodology outlined in the main article. Firstly, the averages for graph percolation were computed over 15 independent realizations of random link deletion, instead of 100 as in the original study. The maximum network sizes analyzed were limited to 4.8×10^4 nodes for an Erdős–Rényi network with an average degree of $c = 6$, and 4×10^4 nodes for both the Erdős–Rényi network with $c = 8$ and the Barabási–Albert network. Additionally, the set of percolation probabilities p was discretized unevenly to balance resolution and efficiency: a fine-grained sampling was applied for $p \in [0, 0.4]$, and a coarser sampling for $p \in [0.42, 1]$. Furthermore, the analysis of critical probability versus network size and maximum diameter versus network size (as shown in Figure 2.3) includes fewer instances from different network topologies than the main article. This choice was made to reduce computational overhead while retaining sufficient data to validate the trends observed. Finally, in Figure 2.1, individual plots are displayed sequentially rather than being overlaid

within the same figure, as was done in the main article. This decision ensures clarity and avoids visual clutter, particularly given that the data in this study exhibit higher noise levels due to the reduced number of samples used for averaging compared to the original work.

A | Bibliography

[1] Bruno Coelho Coutinho, William John Munro, Kae Nemoto, and Yasser Omar. Robustness of noisy quantum networks. *Communications Physics*, 5(1):105, 2022. doi: 10.1038/s42005-022-00866-7. URL https://doi.org/10.1038/s42005-022-00866-7.

[2] Fragkiskos Papadopoulos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov. Popularity versus similarity in growing networks. *Nature*, 489(7417):537–540, 2012. doi: 10.1038/nature11459. URL https://doi.org/10.1038/nature11459.