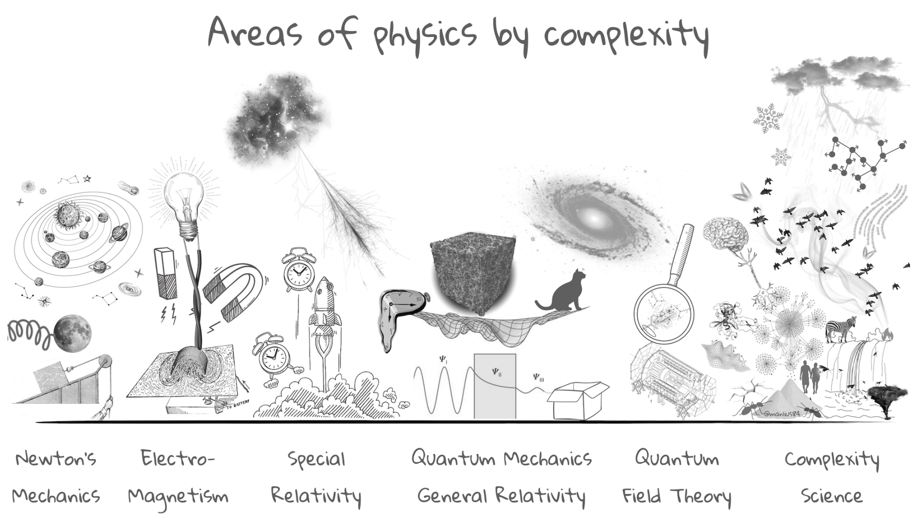
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



Project # 1, # 11, # 39

Mezquita, Tomás

Last update: September 16, 2024

Contents

1 | Ising model on arbitrary topologies

Task leader(s): *Mezquita, Tomás*

1.1 | Theoretical background

The Ising model is widely used in statistical mechanics. It allows us to model how state systems interact. The conventional Ising model is described by a lattice in which each node is described by a state or spin (normally +1 or -1) and the interactions are given by the following Hamiltonian [3]:

$$H = -\sum_{<ij>} J_{ij} s_i s_j - \sum_i H_i s_i \tag{1.1}$$

The first term gives us the interaction of a node with its neighbors of the lattice. The second one describes the interaction with an external field.

Onsager solved The two-dimensional Ising model proving a phase transition between an ordered and disordered phase.

We can study certain magnitudes such as magnetization and examine how it evolves with the temperature to see at what critic temperature the phase transition takes place.

1.2 | Ising model on complex networks

We can implement complex networks on the Ising model if instead of using a lattice we use one of the network topologies that we have learned during the course [2]. Then, a spin or state is assigned to each node of the network.

The topology of the network will have an effect on the behaviour that emerges from the Ising model. Important parameters of the model can change depending on properties of the initial network, such as it's degree distribution.

The simulation will have the following schema [1]:

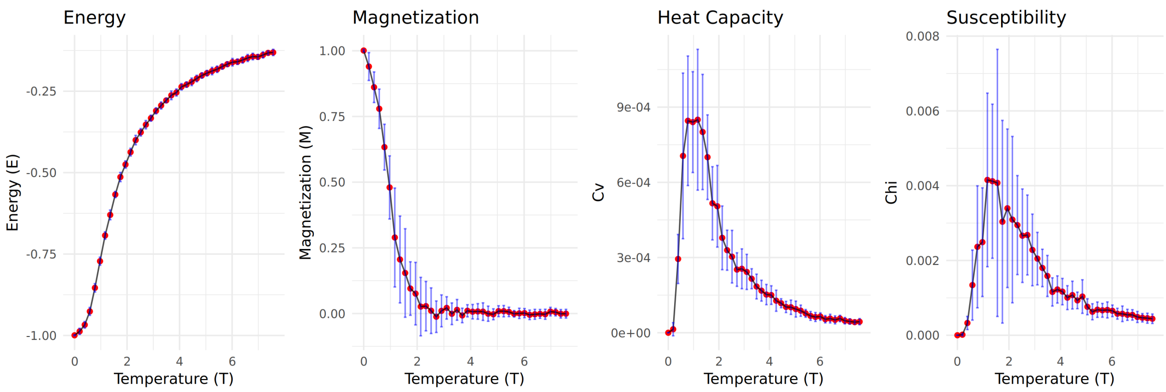
1. Initialize the network and assign initial spins to each node (i.e. all ones).
2. Define complementary functions such as one to compute the energy (using the Hamiltonian) and the magnetization and one metropolis algorithm to update the spins for a given temperature

3. Update the spins for different temperatures and compute the following magnitudes for each temperature: energy, magnetization, specific heat, and magnetic susceptibility.

Nesting loops within other loops involves high computational power and lead to long execution times. To adress this, I have implemented parallelization across the cores of my local machine, which significantly improves performance and reduces running times.

Results for a scale free network

Firstly, I am going to see the results for a scale free network of a given size, to see how energy, magnetization, specific heat and magnetic susceptibility evolves over the temperatures. The results are shown in the figure below:

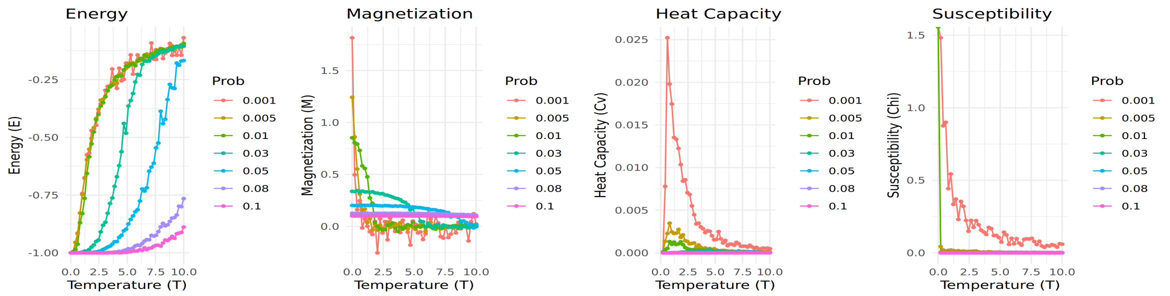


We have got some expected results [3]. As a start configuration we choose all spins up. This will traduce into an ordered phase for low temperatures (ferromagnetic). Due to the configuration choice we start from a minimal energy point and a maximum magnetization. These magnitudes will evolve as the temperature where a transition phase is made where we enter into a disordered phase (paramagnetic) where the magnetization decreases and the energy increase to 0 (approximately there will be same up and down states.

This transition is produced in the critic temperature, where the magnetization and energy change abruptly. Magnetic susceptibility and heat capacity are magnitudes that describe how the magnetization and energy change respectively, and the biggest change is produced in the critical temperature. Consequently, there will be a peak at the critic temperature. We can see that for this network, the value of the critic temperature is around T=1,0-1,5.

Ising model for Erdos Renyi networks

Erdos Renyi networks are simple networks to see how the Ising model behaves under different topologies. We will obtain results for different connection probabilities. This figure contains the obtained results:



As we increase the connection probability and, consequently, the mean degree of the network, the critical temperature rises. This indicates that the system requires a higher temperature to transition into the disordered phase. In the energy plot, for higher probabilities, the phase transition is pushed beyond the temperature range considered, reflecting this shift in critical temperature.

In the magnetization plot, we observe that the initial magnetization starts at lower values as the connection probability increases. This is unexpected, as magnetization should begin at 1 for T = 0.0 (corresponding to an initial all-up spin configuration). Setting this anomaly aside, we see that magnetization consistently drops to zero after passing the critical temperature, which is higher for networks with greater connectivity.

The heat capacity and susceptibility exhibit sharper peaks at lower probabilities because phase transitions occur more abruptly in less connected networks. For higher probabilities, both quantities show lower and broader peaks, indicating a smoother transition into the disordered phase.

In summary, less connected networks require a lower temperature to enter the disordered phase. This is because the Ising model dynamics propagate more easily in simpler networks with fewer connections. In contrast, more connected networks are more resistant to disorder, as each node is influenced by a larger number of neighbors, requiring higher temperatures to disrupt the ordered state.

2 | Growth shrink models

Task leader(s): *Mezquita, Tomás*

2.1 | Theoretical background

Growth models are very crucial in the study of complex networks. These models start from an initial network and incrementally add nodes and edges, leading to changes in the network topology and connectivity. One of the most important growth model it's the Barabasi-Albert preferential attachment model. In this model, new nodes tend to attach to well-connected nodes, resulting in a "rich-get-richer effect" that amplifies the connectivity of high-degree nodes while completing the network with new less connected nodes. This mechanism leads to interesting properties like a scale-free topology, which can be used to study real-life networks exhibiting such characteristics.

In contrast, shrink models involve removing nodes and their connections. These removals will affect the network by lowering its connectivity. The two principal shrink models involve the random removal of nodes and its connections or the targeted removal of nodes, for instance, removing the higher degree nodes.

The objective of this project is to examine how growth and shrink models affect the network topology, connectivity and the emergence of scale-free properties while varying the parameters of the model.

2.2 | Results

In order to achieve our objectives and obtain meaningful results, we will implement the following computational structure:

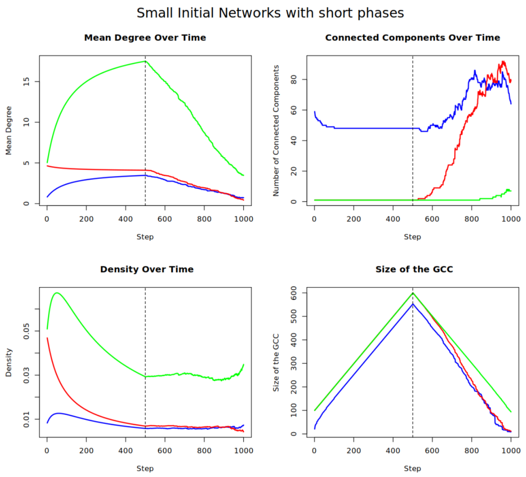
1. **Initialize the network**: We start with a basic graph. For simplicity, we use an Erdos-Renyi network with a specified number of initial nodes and a connection probability.
2. **Growth Phase**: We then add nodes to the network. Each new node connects to existing nodes based on their degree, following the Barabasi-Albert preferential attachment model.
3. **Shrink Phase**: Finally, we perform the shrink phase by randomly removing nodes and their connections from the network.

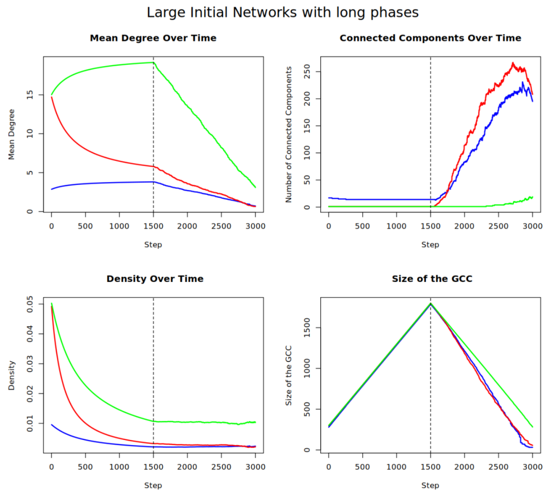
Throughout these phases, representative magnitudes of the network have been computed in each step, to see the evolution of these magnitudes as the network changes.

The set of parameters that we are going to work with are: number of initial nodes, connection probability, number of grow and shrink steps, and edges added per step in the grow phase (m).

As the space is reduced we have picked six representative cases or set of parameters to visualize the evolution of the network. Starting from small and large networks we have:

- Sparse initial network with low edge addition (blue in the graph)
- Dense initial network with low edge addition (red in the graph).
- Dense initial network with high edge addition (green in the graph).





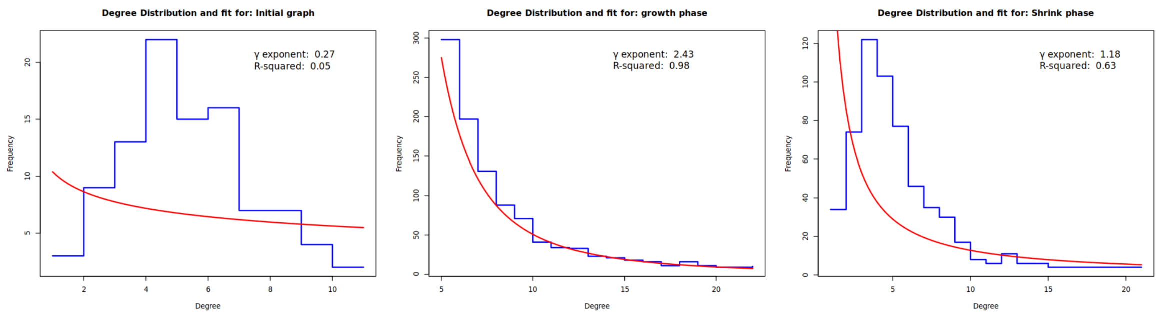
- **Mean degree**: during the growth phase, the mean degree increases if the network is sparse and the edges added per step (m) are low. For denser networks, if m is insufficient, the mean degree may decrease as it already starts from a high mean degree. The mean degree tends to converge to $\langle k \rangle = 2m$ as predicted. In the shrink phase, the mean degree decreases linearly, with randomness causing either abrupt or gradual decreases depending on node connectivity.
- **Connected components and size of GCC**: In the growth phase, if m is high relative to the initial network density, the density initially increases, reaching a peak before decreasing as the network size grows. For larger networks, density directly decreases. During the shrink phase, density remains roughly constant as edges are removed along with nodes.
- **Density** : In the growth phase, if m is high relative to the initial network density, the density initially increases, reaching a peak before decreasing as the network size grows. For larger networks, density directly decreases. During the shrink phase, density remains roughly constant as edges are removed along with nodes.

Other magnitudes, such as clustering coefficient and average path length, were computed, but are not showed as they show less variability. The clustering coefficient remains low, while the average path length tends to increase during the shrink phase..

The study of the degree distribution in the preferential attachment model its vital because it results in a power law distribution which help us understand the scale free networks. We are going to see also how the shrink phase affect the degree distribution.

Degree distribution

If the preferential attachment grow phase, degree distribution often follows a power law, that have the following structure $P(k) = k^{-\gamma}$. Using R programming I can obtain the degree distribution and visualize it, and also a power law fit which will give us the coefficient γ and the R-squared to measure the quality of the fit. I have chosen an arbitrary network of 100 initial nodes and ER probability 0.05, 1000 grow steps, 500 shrink steps and 5 edges per growth step, and computed its distribution and fit at the initial graph, after the growth phase, and after the shrink phase. This are the results for an arbitrary network of:



For the initial Erdos Renyi graph, where edges are placed randomly should follow a Poisson distribution centered around $< k >= Np = 5$, as observed in the results. Therefore, the distribution doesn't fit to a power law.

As commented previously, during preferential attachment growth phase the power law/scale free nature starts to emerge. As shown in the middle graph, the degree distribution fits a power law, the value of R-squared close to 1 gives us proof that the fit is good. Also notice that all the nodes have degree larger than 5, which is consistent with the number of edges added per step. The γ coefficient is 2.43, for pure Barabasi-Albert the expected coefficient is 3, but in general for scale free networks is $2 < \gamma < 3$ [2], which aligns with our results.

After the shrink phase, the degree distribution deviates from the power law as nodes are removed randomly causing the emergence of a Poisson distribution for lower degree nodes.

Due to the limited space I couldn't explore new directions like trying for different initial networks, and different grow and shrink models, but I think they would be interesting to study in a more requiring project.

The study of the phenomena of grow shrink networks can help us understand some of the most important real life networks as networks are rarely static and tend to grow or shrink over time. Some examples of these networks are the world wide web, social networks, biological networks on species...

3 | Subways I

Task leader(s): *Tomás Mezquita*

3.1 | Making the node and edge files

We are working with data from different city subways: Moscow, New York (new and old), Tokyo, Paris, Osaka, Seoul, and Shanghai. Each city has a folder which contains different txt files. In this folders we can find:

- **Topologies folder**: containing the networks of all the years since the subway is active. I did not use these files.
- **Line files**: Containing two columns with the name of the line's start and end station. One of the files also contains all the lines and the year that started. These files are important to create the edge file.
- **Station Position Years (SPY) file**: contains the name, latitude and longitude, and year of inauguration of each station. This file is important to make the node list.

There are also some of the cities folders containing old versions of its subway data.

To process this data I am using R language using read.table to read the files and different functions to do the processing.

For the node file, it was pretty straightforward, using the SPY file, I just had to add the nodeID to the columns. The only problem that I encountered is that in some files there were more than one year. My guess, is that the first one is the date of the inauguration and the following ones could be renovations of the station, so I took the first four columns of the file as they contain the information that I wanted to save. I obtained the node file for each city selecting the paths of each city using list.files to select the pattern of the paths to the SPY file (which can be different), and grep to discard some of the files that don't contain relevant information.

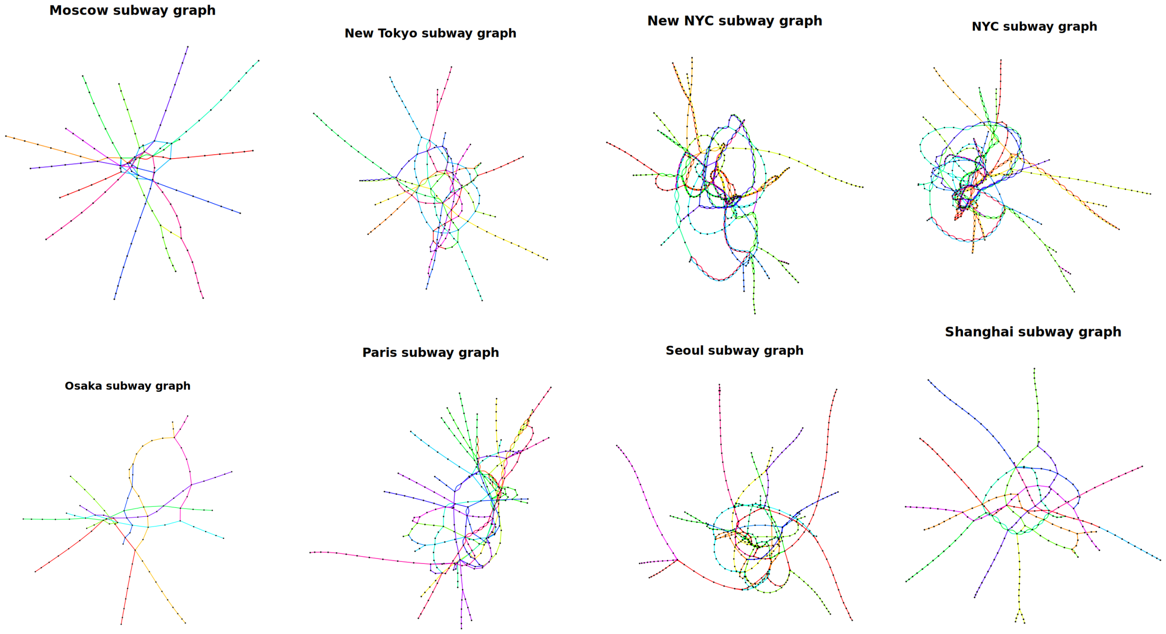
Creating the edge file was more challenging, I used dplyr library to take advantage of relational databases. To ensure that the nodeID is the same as in the node file I created a lookup table where the nodeID and nodeLabel (name of the station) are stored. Moreover, I have to process the line files and change them from the names of the start and end stations to the nodeID of each station using the lookup table. I also have to process the lines file to obtain the line name and year.

Having these support functions, we can create the principal function which creates an edge file for a given city, we will need primarily, as inputs the node file, line files, and the lines file where the years are stored. The line paths have approximately the following structure: "data/City/city-lineN.txt" and using list.files we want to extract the line name ("lineN") then we run our secondary functions to obtain the different data and for each subway line path we obtain the year comparing the line name obtained from the path and the lines file, later we put the start and final nodeID, line and year of each edge into the edge file.

The second challenge of this task is to select the file paths for each city and ensure that the line name obtained from the path is the same as in the lines file. The problem is that some city path structures are different, so I had to add some controls for some cities to extract the correct line name. After addressing this issues, I successfully created the edge files as required.

3.2 | Analysis of the networks

With the nodes and edges files in hand, we can easily reconstruct the graph using the "graph_from_data_frame" function from the "igraph" library. Here is the visualization of all the 8 cities subway graphs. For clarity, I have set the lines of the subways in different colors. Here are the results:



As observed, all the cities exhibit a similar pattern: a central area that is more connected, representing the city center, where most of the lines pass through and nodes have a higher degree, or in other words, where the stations are more likely to connect with other lines. Otherwise, the outer regions of the graph display lines that are less connected with other lines. We observe a representative case of this phenomenon in the Moscow graph, where the external lines are rarely connected. This is due to these lines extending into the outskirts of the city, moving away from the city center and becoming more isolated from other lines.

This pattern makes sense, as city centers have higher population densities and more activity so the transportation requirements are higher. Contrarily, the outskirts of a city experience less activity and the requirements are lower, with more routes converging towards the city center. It would be interesting to study a multilayer network that combines the population density with vatious transportation networks for a better understanding of these dynamics.

Next, we will analyze the graphs quantitatively by computing some quantities that will help us with a better understanding of the nature of the graphs and its characteristics. These metrics include mean degree, cluster coefficient, average length path, size of the connected component, etc. I have summarized these magnitudes in the table below:

City	Mean Degree	Diameter	Avg. Path Length	Cluster Coef.	Assortativity
Moscow	2.343284	24	9.440579	0.019608	0.465932
New Tokyo	2.562212	32	10.398916	0.025428	0.253371
New NYC	4.322581	59	20.989421	0.018070	0.371283
NYC	3.958525	55	16.114228	0.029114	0.319052
Osaka	2.300000	23	8.454545	0.000000	0.209897
Paris	2.664452	33	11.793289	0.022405	0.200149
Seoul	3.141667	66	19.912804	0.008336	0.760314
Shanghai	2.292887	41	14.577160	0.001943	0.175337

From these magnitudes, we can extract significant information on the network structures. For instance, both the New York (both new and old systems) and Seoul subways show high mean degrees (3-4), large diameter of the GCC (40-60), high average path length (15-20), and high assortativity (0,3-0,7).

This metrics suggest dense, large, and more connected graphs. In practical terms, this means they have a more developed structure with more interconnected stations making them more navigable. High assortativity shows that stations with similar connectivity levels are likely to be connected, indicating a hierarchical structure where major stations are more likely to connect with other hubs.

In contrast, Moscow, Osaka, and Shanghai have a simpler structur with lower magnitudes compared to the previous ones: low mean degree (around 2,3), diameter (20-25), and average path length (8-10). Seoul seems to have a bigger subway structure as it has a higher diameter (41) and average length path (14,58). This type of networks traduce into a simpler subway structure with a more connected small central zone, with most lines extending into the periphery without connecting with other ones. The subways of Paris, Tokyo, and Seoul represent an intermediate case, offering average-size networks with average connectivity.

There are also some common patterns across cities, like the low clustering coefficient suggesting that subway networks are designed to minimize loops and clusters, reducing congestion and allowing a better flow.

To sum up, analyzing the structure of the different subways through complex network techniques provides valuable insights into their structure and functionality. The results aligned with a visual representation of the graphs often reveals the optimization of these structure to maximize connectivity and mobility.

4 | Bibliography

[1] Mahmoud Jafari Ashkan Shekaari. Theory and simulation of the ising model. *arXiv preprint arXiv:2105.00841*, 2021.

[2] Manlio De Domenico. *Physics of Complex Networks: Structure and Dynamics*. University of Padova, last update: March 2024.

[3] Richard Fitzpatrick. Thermodynamic limit — farside.ph.utexas.edu. `https://farside.ph.utexas.edu/teaching/329/lectures/node110.html`, 2016. [Accessed 29-Mar-2023].