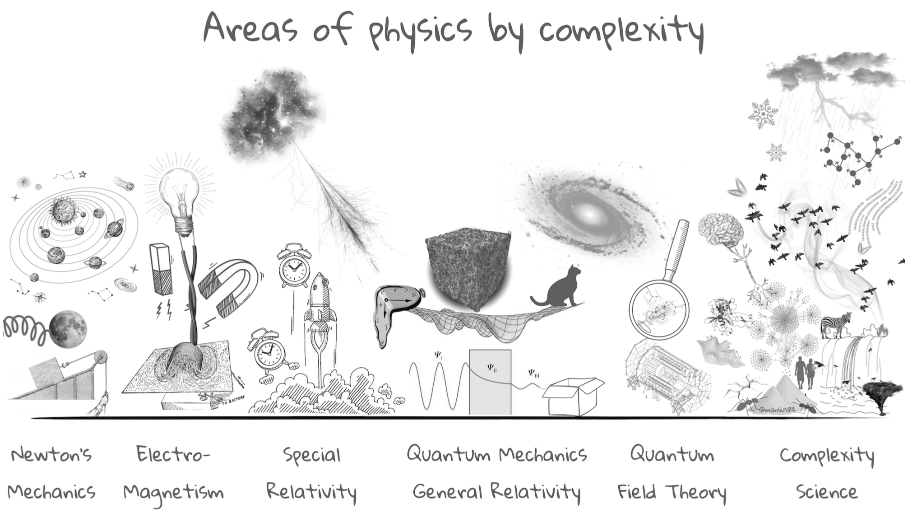
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



Final Report

Carotenuto, Jacopo

Last update: September 10, 2024

Contents

1 | Task # 15 - Cascade Failures: SOC Sandpile Model

Task leader(s): *Jacopo Carotenuto*

1.1 | Introduction

In this task we will study the "Sandpile Model" on different types of networks. This model is a type of dynamic process that can be simulated on a network that exhibits self-organized criticality (SOC). The model has been studied in different types of networks and some interesting relations have emerged, especially about the distribution of the avalanche size and duration. In this task we will simulate the model on different networks and calculate the corresponding distributions.

1.2 | Model Description

The "Sandpile Model" is a type of dynamic process that can be simulated on a network that exhibits self-organized criticality (SOC). The mechanism of this model is rather simple:

1. The network is initialized with every node having zero height h and having threshold $z_i = k_i$ where k_i is the degree of the node
2. At every time step one randomly chosen node i height is incremented by 1.
3. If the height exceeds the threshold, the node shed some of its height to his neighbor in this manner: $h_i \to h_i - k_i$ and $h_j \to h_j + 1$ where j are the neighbors of node i
4. A small fraction f of the height of node i is lost. This act as a sink t prevent the overloading of the system. f is called the "shedding probability".

Numerous works have studied this type of dynamic and some interesting relations have emerged, especially about the distribution of the avalanche size and duration. Both quantities seems to have a power-law distribution of this type:

$$p(x) \sim x^{-\tau_x} \exp\left(-x/c_x\right)$$

Where x is either the avalanche size or the avalanche duration, τ_x is the associated exponent and c_x is the characteristic value of the quantity. [9]

This values vary based on the type of network and parameters used for the simulations, but the underlying power-law distribution seems to hold for all type of networks.

In [9] it's found that, for the avalanche size, $\tau = \frac{\gamma}{\gamma - 1}$ in scale-free networks with power-law degree distribution with exponent γ.

Other works [7] analyzed different type of network with different degree distributions and found similar power-law distributions. In this task I simulated the sand-pile model on a variety of different networks and calculated the corresponding parameters.

1.3 Simulations

The Sandpile model was implemented in the Julia programming language [6]. Each network was generated with $100k$ nodes and 2 million grains were simulated. The shedding probability was 10^{-4} for all networks. The Cascade Size is defined as the number of nodes (with the possibility of repeating nodes) that the cascade affect. The Cascade Duration is the number of topple events until relaxation. The following networks where used:

- Gaussian Degree Distribution ($\mu = 20, \sigma = 6$)
- Uniform Degree Distribution ($[1, 40]$)
- Static Scale Free Network ($\gamma = 2, 2.5, 3, 4$) [8]
- Erdos-Renyi Network ($p = 2 \cdot 10^{-4}$)
- Barabasi-Albert Network ($k = 10$)

Here we can see some examples of distributions, where in blue is the simulation data and in red is the fitted power-law. For the size-duration relation a simple exponential relation was hypothesized according to [9], where the exponent z is called "Dynamic Exponent". In both distributions, the tail was cut to exclude noise from finite-size effects.

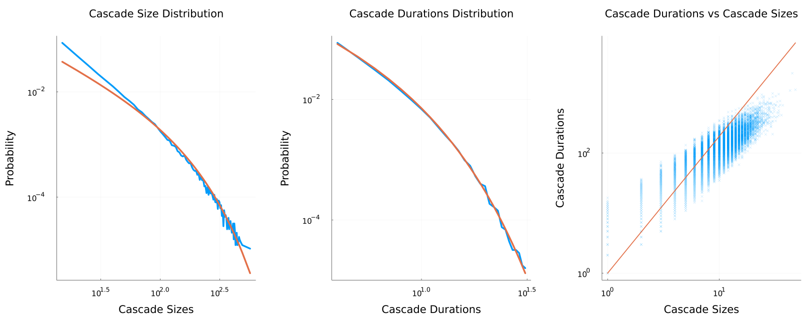


Figure 1.1: Static Scale-Free Network, $\gamma = 2.5$

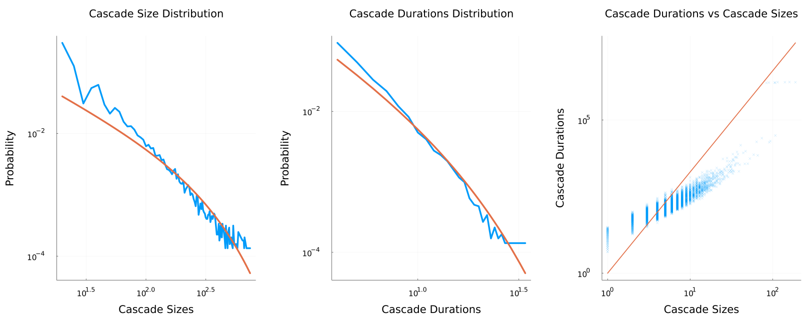


Figure 1.2: Erdos Renyi Network, $p = 2 * 10^{-4}$

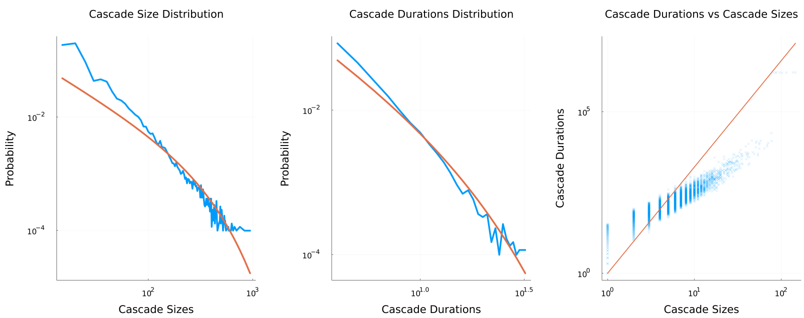


Figure 1.3: Gaussian Degree Distribution, $\mu = 20, \sigma = 6$

For the scale-free networks, the τ_{size} were not exactly in accordance with the value predicted by [9] but it was near enough to hypothesize that the difference was caused by the low number of simulations performed here.

We can the compare all the size distribution by plotting them all together:

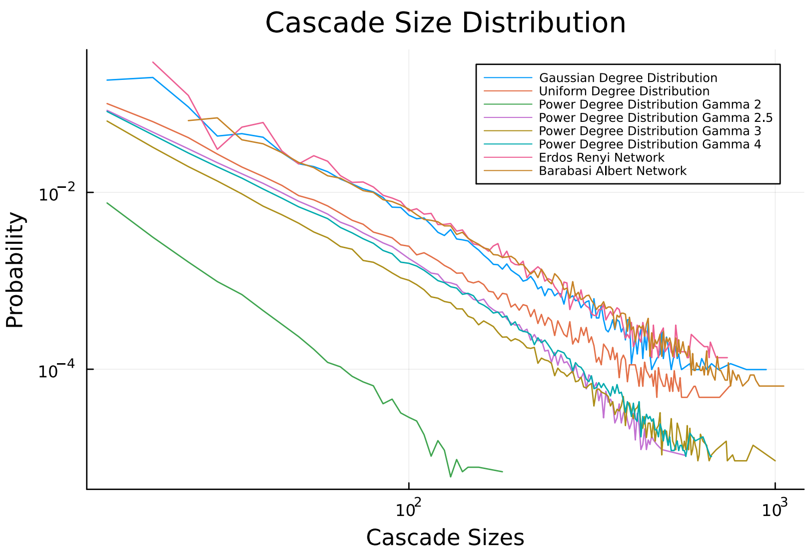


Figure 1.4: All Networks Size Distributions

2 | Task # 34 - Sociophysics: Game Theory on networks

Task leader(s): *Jacopo Carotenuto*

2.1 | Introduction

In this task we will study the evolution of strategies according to different "survival" rules. We will consider the "Ultimatum Game" on different networks with different types of players and strategies.

2.2 | Playing the Ultimatum Game on Networks

We will follow the setup of [10]. The Ultimatum Game is a two-player game where one player, the proposer, is given a sum of money (here 1) and proposes how to divide it between himself and the other player. The other player, the responder, can either accept or reject the proposal. If the responder accepts, the money is divided as proposed. If the responder rejects, both players receive nothing. The game is played twice, with the roles of proposer and responder reversed in the second round.

Different network types will be considered, where each node represents a player and each edge represents a connection between players. The players will play only with their neighbors. Each player will have a two numbers associated with it (from 0 to 1): p, the division offered when playing as the proposer and q the threshold for accepting the division when playing as the responder. There will be 3 types of players:

(A) **Fair Player**: The player will have $p = q$.

(B) **Pragmatic Player**: The player will have $p = 1 - q$. This player wants the same reward both as proposer and responder.

(C) **Random Player**: The player will have independent p and q.

At each cycle, each player will play the game with each of its neighbors and the score will be updated (the score being the sum of the money received). Two updating rules will be considered:

1. **Natural Selection**: In this framework, originally introduced in [37, 38], each player i in the network selects at random one neighbor j and compares its payoff Π_i with that of j, Π_j. If $\Pi_j > \Pi_i$, player i adopts the strategy of $j, (p_j, q_j)$, for the next round of the UG with a probability proportional to the payoff difference:

$$P_{ij} = \frac{\Pi_j - \Pi_i}{2 \max \{k_i, k_j\}}$$

where k_i and k_j are the degrees of i and j respectively. However, if $\Pi_i \leq \Pi_j, i$ keeps its strategy for the following round.

2. **Social Penalty**: The player with lowest payoff in the whole population and its neighbors, no matter how wealthy they are, are removed. These agents are replaced in their nodes by new players with random strategies (so that they only inherit their contacts).

2.3 | Simulations

Three types of networks were considered: Erdos-Renyi, Barabasi-Albert and Scale-Free. The networks were generated with $10k$ nodes and average degree of 4. On all three networks each type of player was simulated with each type of updating rule. The distribution of p and q among the player was recorded at predetermined intervals. The following figures show the distribution of p for the different types of players and networks.

All combinations were averaged over ten different simulation run for $2 \cdot 10^4$ cycles (Natural Selection) or for 10^5 cycles (Social Penalty). The Social Penalty updating rules was simulated more because it was slower to reach a stable configuration.

The following figures show the distribution of p for the different types of players and networks.

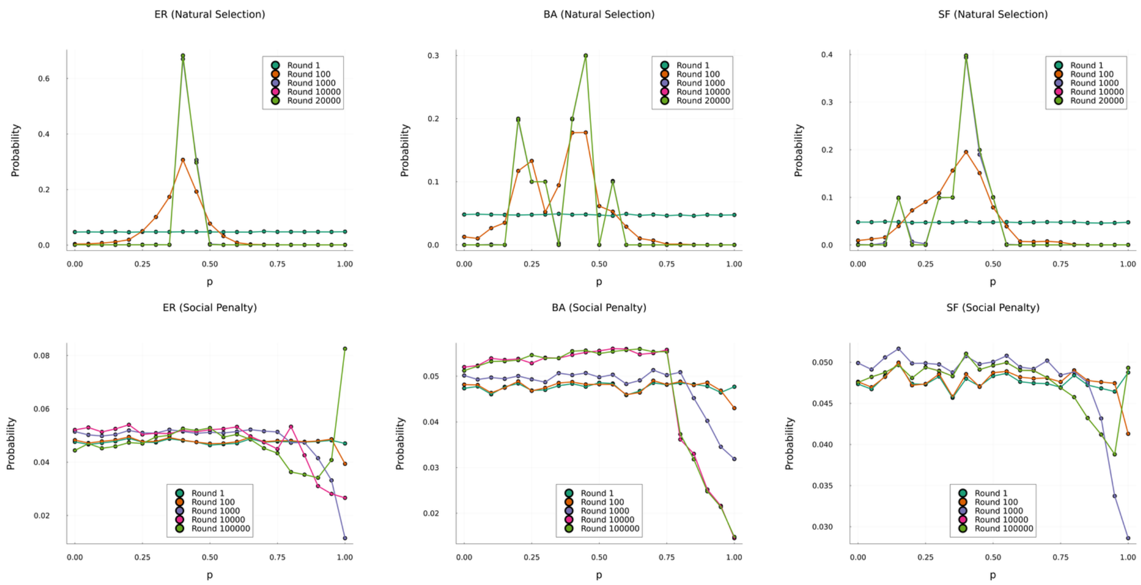


Figure 2.1: Networks with only type A players. First row is with Natural Selection, second row is with Social Penalty.

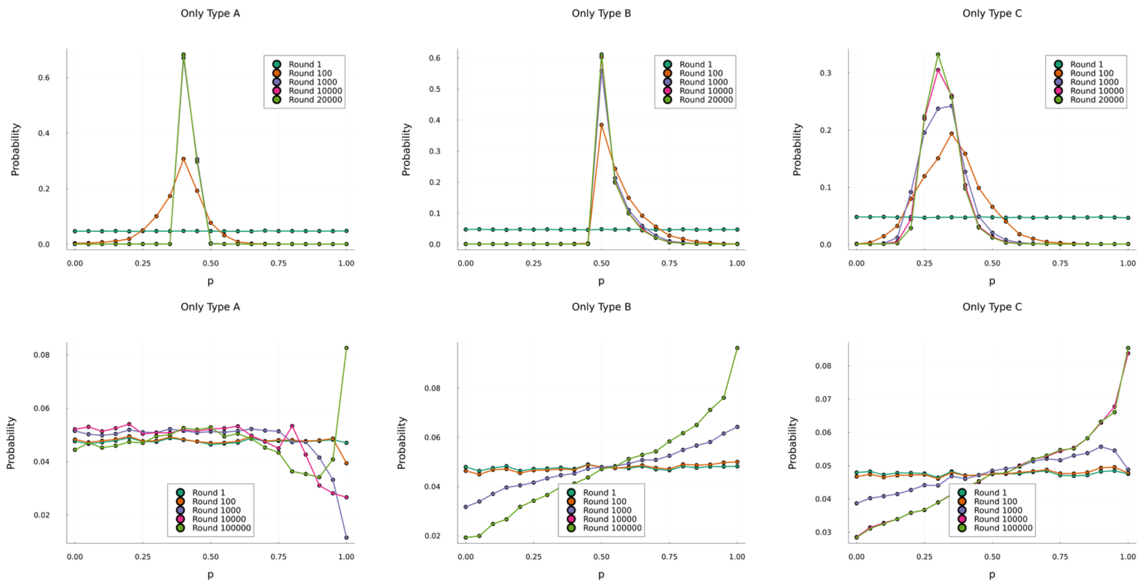


Figure 2.2: Networks with all types of players on Erdos-Renyi networks. First row is with Natural Selection, second row is with Social Penalty.

As stated in [10] the change in topology does not massively change the distribution of strategies, but the updating rule does. This results corroborate this conclusion. The individual distribution are different mainly because in this task a very low number of realization was utilized.

3 | Task # 44 - Social Connectedness Index from Facebook

Task leader(s): *Jacopo Carotenuto*

3.1 | Introduction

In this task the goal was to extract a network for each country present in the Facebook Social Connectedness Index data [1]. The data was collected by Facebook and it is based on the number of friendships between people in different countries. From the documentation:

> The Social Connectedness Index uses an anonymized snapshot of active Facebook users and their friendship networks to measure the intensity of social connectedness between locations. Users are assigned to locations based on their information and activity on Facebook, including the stated city on their Facebook profile, and device and connection information.

3.2 | Network Extraction

3.2.1 Raw Data

The goal was to build two files:

1. `node_list.csv`: a file containing the node ID, longitude, latitude and label for each node.
2. `edge_list.csv`: a file containing the source node ID, target node ID, weight (SCI), country and country ISO code for each edge.

As the goal was to build a network for each country, edges between different countries were not considered. The data provided was in a ".tsv" file with the following columns: "`user_loc`", "`fr_loc`" and "`scaled_sci`" representing the User Location, Friend Location and Scaled Social Connectedness Index respectively. As written in the documentation, the SSCI was built to be between 1 and 10^9. All data manipulation and analysis were done in Julia[6].

3.2.2 Edge List

The first step was to extract all the unique locations present in the data and assign a unique ID to each of them. As the location were divided in different types of denominations (GADM[2], NUTS[3] and US Counties) the file "`gadm1_nuts3_counties_levels.csv`" was used to determine the type of denomination for each location. Then each location was assigned the ISO3 code of the country it belonged to using the "Countries" package in Julia[4]. To each entry of the original file (so to each edge) the User ISO3 code and the Friend ISO3 code were added and all the edges between different countries were removed. The final edge list was saved in a ".csv" file, saving only the source node ID, target node ID, weight, country and country ISO code.

3.2.3 Node List

The node list was more of a challenge as the longitude and latitude of each location were not provided. To assign to each location its coordinates, the geospatial data for each subdivision was found:

- For the GADM subdivisions, the data was found in the GADM website[2].
- For the NUTS subdivisions, the data was found in the Eurostat website[3].

The provided data was in the form of a ".shp" or ".geojson" files, so for the coordinates extraction the "GeoStats.jl" package was used[5]. For all the administrative subdivision a polygon was provided, so the centroid of each polygon was used as the coordinates for the location. As the provided Facebook data contained different levels of administrative subdivisions, each level was analyzed separately. Then, as the Facebook area code used was different from all the other denominations, the area codes of each location were transformed in the Facebook format. Then, for each unique node (excluding the ones in the USA) we associated the coordinates and the label (also found in the official data). The final node list was saved in a ".csv" file, saving the node ID, longitude, latitude and label.

3.3 | Network Comparison

The resulting countries network were all analyzed using Julia and some important network properties were extracted. Some of them are contained in the following plots.

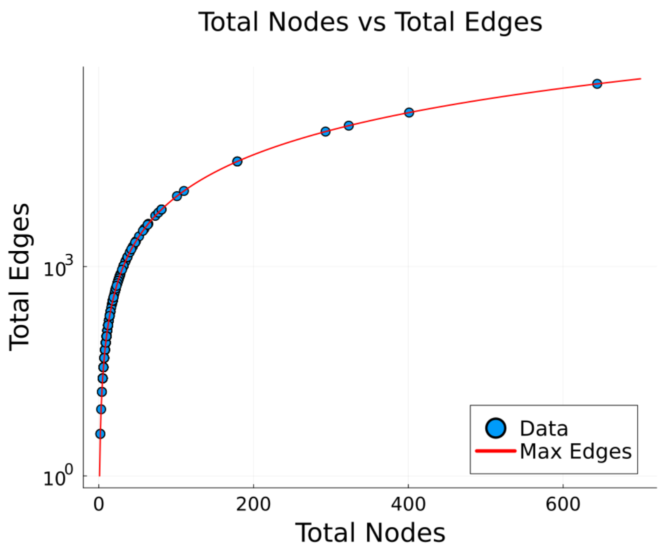


Figure 3.1:

From plot 3.1 we can see that the number of edges is proportional to the number of nodes (it's exactly the maximum number of connections, allowing for self-loops), as expected as all the networks are complete networks: every location has at least some people connected to every other location in the same country.

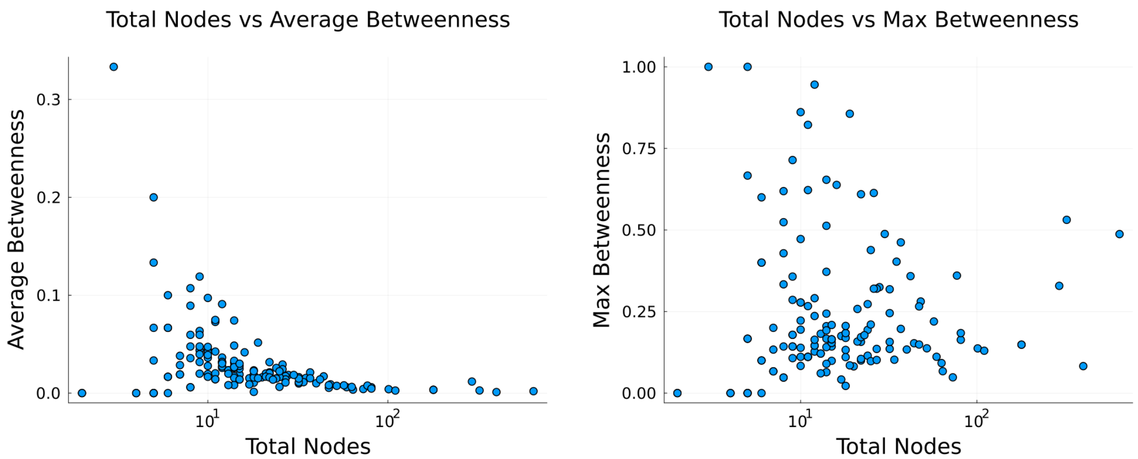


Figure 3.2:

From plot 3.2 we can see that the average betweenness is inversely proportional to the number of nodes in a exponential way but the maximum betweenness is less correlated with the number of nodes.

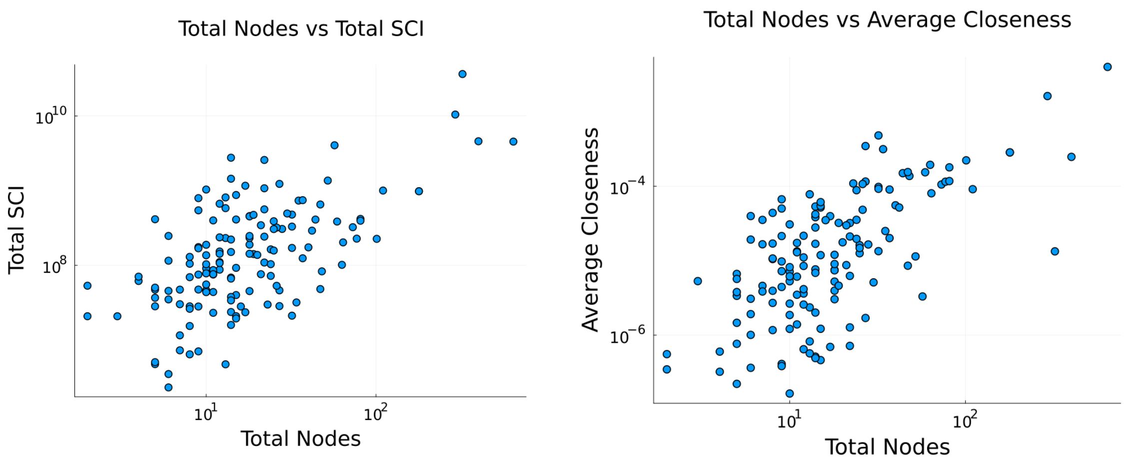


Figure 3.3:

From plot 3.3 we can see that there seem to be correlation between the total number of nodes and the average closeness.

From plot 3.3 we can see that the total SCI is proportional to the number of nodes in a exponential way.

4 | Bibliography

[1] Facebook social connectedness index. https://data.humdata.org/dataset/social-connectedness-index. [Accessed 25 August 2024].

[2] Gadm. https://gadm.org/. [Accessed 25 August 2024].

[3] Nuts. https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/territorial-units-statistics. [Accessed 25 August 2024].

[4] Countries julia package. https://github.com/cjdoris/Countries.jl. [Accessed 25 August 2024].

[5] Geostats julia package. https://github.com/JuliaEarth/GeoStats.jl. [Accessed 25 August 2024].

[6] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017. URL https://doi.org/10.1137/141000671.

[7] Eric Bonabeau. Sandpile dynamics on random graphs. *Journal of the Physical Society of Japan*, 64(1):327–328, 1995. doi: 10.1143/JPSJ.64.327. URL https://doi.org/10.1143/JPSJ.64.327.

[8] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.*, 87:278701, Dec 2001. doi: 10.1103/PhysRevLett.87.278701. URL https://link.aps.org/doi/10.1103/PhysRevLett.87.278701.

[9] K.-I. Goh, D.-S. Lee, B. Kahng, and D. Kim. Sandpile on scale-free networks. *Phys. Rev. Lett.*, 91:148701, Oct 2003. doi: 10.1103/PhysRevLett.91.148701. URL https://link.aps.org/doi/10.1103/PhysRevLett.91.148701.

[10] R Sinatra, J Iranzo, J Gómez-Gardeñes, L M Floría, V Latora, and Y Moreno. The ultimatum game in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(09):P09012, sep 2009. doi: 10.1088/1742-5468/2009/09/P09012. URL https://dx.doi.org/10.1088/1742-5468/2009/09/P09012.