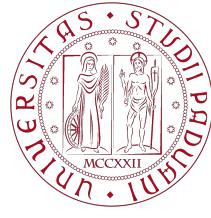


# Final Report

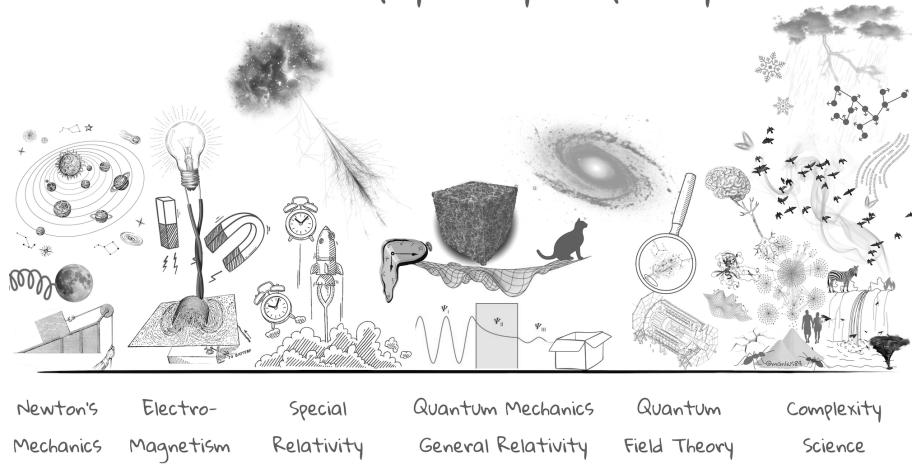
Physics of Complex Networks: Structure and Dynamics

Last update: September 11, 2024



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Areas of physics by complexity



## PoCN: Projects

Cerni, Paolo Lapo

# Contents

---

<b>1</b>	<b>SOC model</b>	<b>1</b>
1.1	Dynamical model for self-organization . . . . .	1
1.2	Numerical simulations . . . . .	2
<b>2</b>	<b>Dissemination of culture</b>	<b>3</b>
2.1	Social influence and consensus formation . . . . .	3
2.2	Numerical simulations . . . . .	4
<b>3</b>	<b>Social Connectedness Index</b>	<b>5</b>
3.1	Introduction . . . . .	5
3.2	Methods and analysis . . . . .	5
3.3	Comparisons . . . . .	6
<b>4</b>	<b>Bibliography</b>	<b>7</b>
<b>A</b>	<b>Supplementary material</b>	<b>9</b>
A.1	SOC: other figures . . . . .	9
A.2	SOC: interdependent networks . . . . .	10
A.3	Axelrod's model: other figures . . . . .	11

# 1 | SOC model

---

**Task 15:** *sandpile model for self-organized criticality*

## 1.1 | Dynamical model for self-organization

---

The Bak–Tang–Wiesenfeld sandpile model is a simple example of a complex system leading to the so-called *self-organizing criticality*, where large systems with many interacting components evolve toward a critical state in which even minor events can trigger significant changes, like an avalanche [1].

This model is operationally defined on top of a network (originally a lattice) where each node  $i$  is assigned to a critical load  $t_i$ . This threshold can be uniform, drawn by a particular distribution, or set to be equal to the degree of each node ( $t_i = k_i$ ). The height  $z_i$  of each site is initialized to zero. At each time step, the load by a uniformly random chosen node is increased by a unit ( $z_i \rightarrow z_i + 1$ ), changing the state of the system. A state is said to be *unstable* if at least one node has height  $z_i$  strictly greater than  $t_i$ . When this happens, all the grains at the node topple to the neighbors ( $z_i \rightarrow 0$  and  $z_j \rightarrow z_j + 1 \quad \forall j \in \mathcal{N}_i$ ), possibly causing a chain of toppling events, i.e. an *avalanche*<sup>1</sup>. This model requires some *boundary conditions* that resemble the open boundaries of a regular lattice. Thus, either  $N_b$  nodes are selected as boundaries and dissipate the load, or a fraction  $f$  of the grains is lost at each iteration.

After a transient period, one can measure the following quantities [2]: (a) the number  $A$  of distinct sites participating in the toppling event, i.e. the avalanche area, (b) the total number of toppling event  $S$ , (c) the number of toppled grains  $G$ , and (d) the duration  $T$  of a given avalanche. In large enough networks,  $A$  is equal to  $S$  due to the lack of loops.

The main feature of this model is the emergence of a power law with an exponential cutoff in the avalanche size distribution:  $p_a(s) \sim s^{-\tau} \exp(-s/s_c)$ , where  $s$  is the avalanche size and  $s_c$  the characteristic size. The exponent  $\tau$  depends on the type of the network. It has been shown that  $\tau \approx 1.5$  for ER network [3], consistently with the mean-field solution in Euclidian space [4]. Since they are everywhere in nature, it is also interesting to study what happens in SF networks, i.e. networks with power-law degree distribution  $p_d(k) \sim k^{-\gamma}$ . In this case the expected exponent  $\tau$  is related to  $\gamma$  as  $\tau = \gamma/(\gamma - 1)$  for  $2 < \gamma < 3$  [2]. For  $\gamma > 3$  one recovers the mean-field solution.

Another interesting case study is interdependent networks, e.g.  $z_a - z_b$  regular graphs coupled by Bernoulli-distributed coupling studied by Brummitt et al. [5]. In the case of two independent networks  $a$  and  $b$ , one can investigate the sizes of avalanches in  $a$  caused by an initial topple in  $a$  or  $b$ , namely  $S_{aa}$  and  $S_{ba}$ . It has been shown that increasing interconnectivity  $p$  suppresses large local cascades for small  $p$  and amplifies them

---

<sup>1</sup>In practice, there are different ways to model the sequential update of the states during an avalanche. In my simulations, I used a first-in-first-out queue.

for large  $p$ . Also, tuning the interdependence to suppress cascades of a certain range leads to increased events in other ranges. This research has important consequences in many real-world networks that have a modular structure. It would be interesting to study other types of interconnectivity, for instance, in the SBM ensemble. The result of some simulations can be found in the supplementary material A.2.

## 1.2 | Numerical simulations

The following numerical simulation compares the sandpile dynamics on top of 8 networks with  $10^4$  nodes each. The networks are generated with the `NetworkX` Python package [6]. In particular, two of them are Erdős–Rényi networks ( $p = 7 \times 10^{-4}$  and  $p = 2 \times 10^{-3}$ ), and two are Barabasi-Albert networks ( $m = 1$  and  $m = 5$ ). Four of them are undirected scale-free networks obtained varying the  $(\alpha, \beta, \gamma)$  parameters in the `scale_free_graph()` function to obtain different values of the power-law exponent  $\gamma$ . The simulations last for  $10^7$  iterations with a dissipation rate of  $f = 1/N$ . This work's main limitations are the network size and the number of simulations.

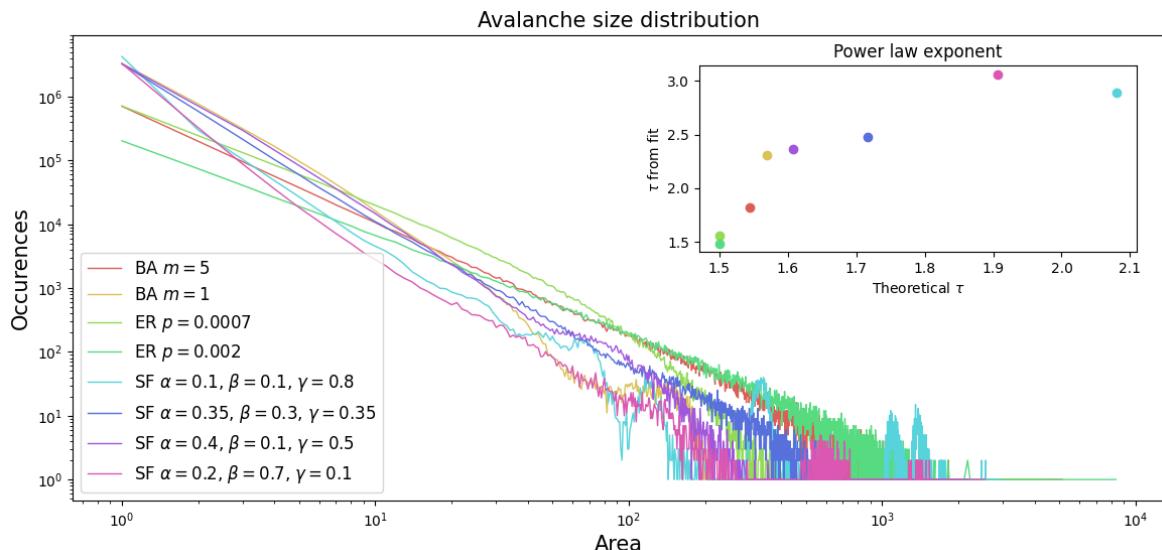


Figure 1.1: (main) Power-law distributed avalanche sizes for eight different networks of  $N = 10^4$  nodes. Simulations of  $10^7$  iterations with a dissipation rate  $f = 1/N$ . (inset) Comparison between the exponent  $\tau$  expected from the properties of the network and the one obtained by fitting the curves.

To find the expected  $\tau$ , one must first estimate the degree exponent  $\gamma$  for the scale-free networks. This already comes with statistical implications for the reliability of this analysis. For instance, a BA network with  $m = 1$  is expected to have  $\gamma = 3$  but was estimated at 2.76. Then, one can fit the avalanche size distribution and compute the correspondent  $\tau$  from the simulations. To limit the noise of the distribution tail, the fit is done with the first 10 points only.

In scale-free networks, the power-law exponent from the simulations is typically greater than the expected one. This can be related to the finite-size effects of this analysis, leading to curves with higher slopes and faster decay. To improve the agreement, one should run longer simulations in larger networks.

Other figures can be found in the supplementary material A.1 and A.2.

## 2

# Dissemination of culture

**Task 30:** Axelrod's model for dissemination of culture

## 2.1 | Social influence and consensus formation

The *Axelrod's model* aims at understanding the formation of cultural domains and the effects of a convergent social influence, describing how local convergence can generate global polarization. Here, the term *culture* refers to the set of individual attributes shaped by social dynamics. The following mechanism is based on three principles: an agent-based approach, the lack of a central authority, and adaptive - rather than rational - behavior [7].

The model is defined on top of a network where each node  $i$  has a set of  $F$  integer variables  $\sigma_{i,F}$  representing the cultural "features". Each feature  $f$  is initially drawn randomly from a Poisson distribution of parameter  $q$ . Thus,  $q$  measures the initial cultural variability. At each time step, a pair of nearest neighbors  $(i, j)$  and a feature  $f$  are randomly selected. If  $\sigma_{i,f} \neq \sigma_{j,f}$  nothing happens. If instead  $\sigma_{i,f} = \sigma_{j,f}$  then nodes  $i$  and  $j$  interact and another feature  $g$  such that  $\sigma_{i,g} \neq \sigma_{j,g}$  is chosen and set equal:  $\sigma_{i,g} \rightarrow \sigma_{i,g} = \sigma_{j,g}$ .

Note that a link is *active* if the number of shared features is strictly greater than zero and lower than  $F$  and the total diversity (namely the number of different values of  $f$  in the system) always decreases. The dynamics converge to one of the many *absorbing states* such that, on each link, either all features are equal or they are all different [8, 9].

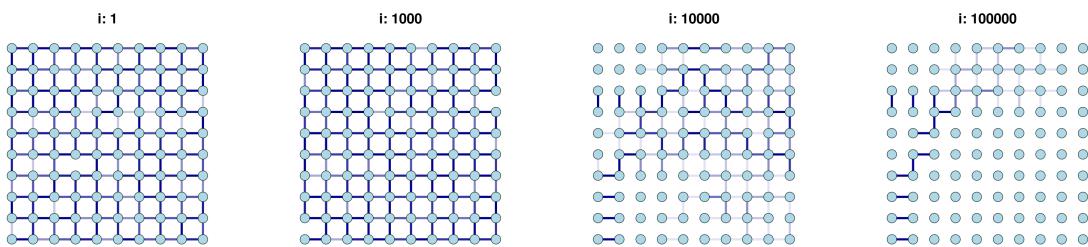


Figure 2.1: Example of consensus formation: the darker the blue at the edge, the fewer features the nodes share. Four snapshots of the dynamics simulated on top of a 2d regular lattice of size  $L = 10$ . Here,  $F = 10$  and  $q = 4$ .

An interesting particularity of this model is the existence of a phase transition between a "culturally polarized" and "culturally fragmented" final state. This can be studied by analyzing the size of the largest cultural domain  $s_{\max}$  (the order parameter) as a function of the initial variability parameter  $q$  [8, 9]. The transition becomes sharp and well-defined for large systems [9], thus I was not able to properly characterize it in this

report due to the lack of proper computational resources <sup>1</sup>.

## 2.2 | Numerical simulations

The model has been tested in five different networks, recording both the size of the largest cultural domain  $s_{\max}$  and the density of active links  $n_{\text{active}}$  varying  $q$  and  $F$ . The networks are obtained via Erdős–Rényi ( $p = 2 \times 10^{-3}, 7 \times 10^{-3}, 1.2 \times 10^{-2}$ ) and Barabasi-Albert ( $m = 1, 5$ ) instances. Each network has  $N = 2 \times 10^3$  nodes and the results refer to 4 independent simulations of  $5 \times 10^7$  iterations each. Moreover, 3 massive simulations have been done in  $50 \times 50$  regular lattices for  $10^{10}$  time steps.

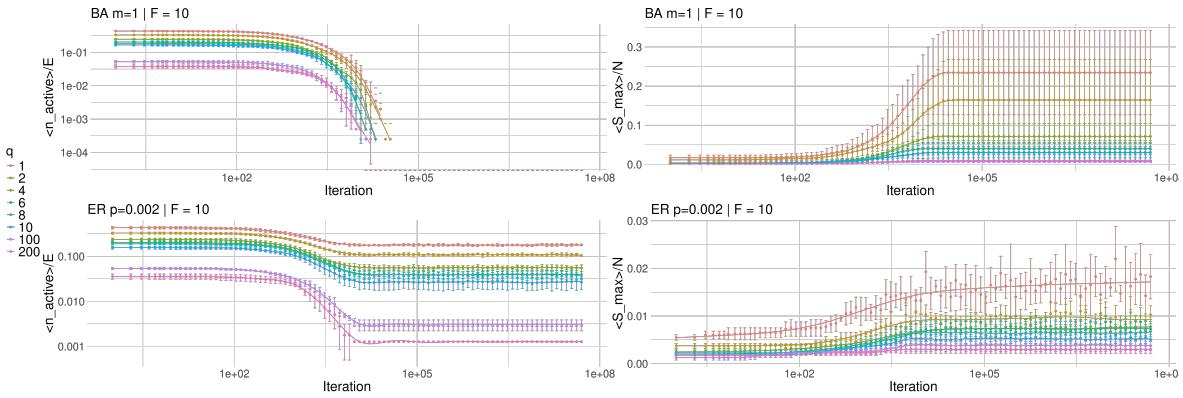


Figure 2.2: Evolution of the  $n_{\text{active}}$  (left) and  $s_{\max}$  (right) in BA  $m = 1$  (top) and ER  $p = 0.002$  (bottom) networks.  $F = 10$ .

Figure 2.2 and 2.3 make evident the role of the network topology in the final state, even for compatible values of the average degree. The dynamics on top of BA networks converge faster to a frozen state both in the case of global consensus (low values of  $q$ ) and cultural fragmentation. Other figures can be found in the supplementary material A.3. The natural continuation of this analysis is to increase the number of nodes and the number of iterations.

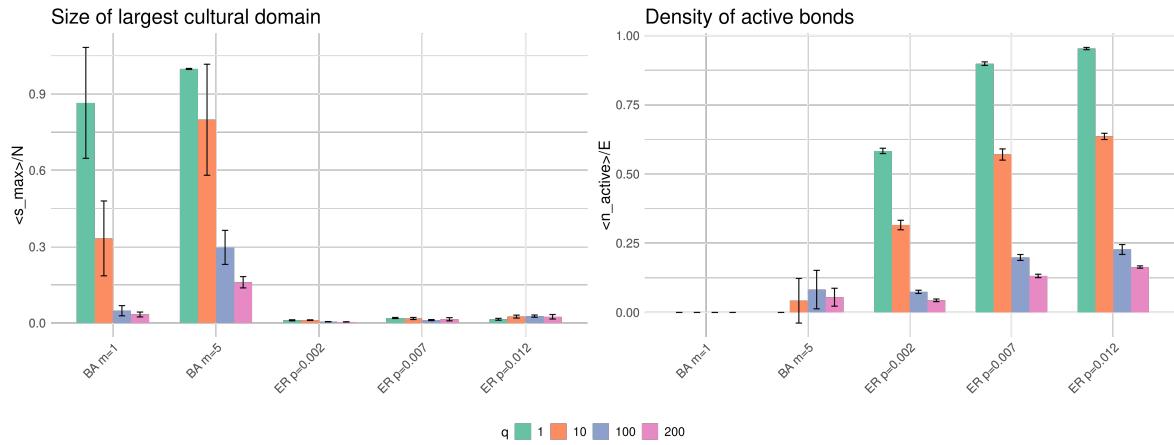


Figure 2.3:  $s_{\max}$  and  $n_{\text{active}}$  after  $5 \times 10^7$  time steps.  $N = 10^3$ ,  $F = 10$ .

<sup>1</sup>An attempt can be found in the supplementary material A.3

## 3 | Social Connectedness Index

---

**Task 44:** *Social Connectedness Index from Facebook*

### 3.1 | Introduction

---

The *Social Connectedness Index* (SCI) measures the strength of the connection between pairs of geographical regions [10]. Firstly introduced in 2018 to analyze connectedness across US counties [11], it is based on an anonymized snapshot of active Facebook users and their friendship networks.

Specifically, the SCI measures the relative probability that two individuals across two locations are friends on Facebook. The SCI between two locations  $i$  and  $j$  can be formalized as:

$$SCI(i, j) = \frac{FB - connections(i, j)}{FB - users(i) * FB - users(j)} , \quad (3.1)$$

where  $FB - users(i)$  and  $FB - users(j)$  are the number of Facebook users in location  $i$  and  $j$  according to their profile, and  $FB - connections(i, j)$  represents the number of Facebook friendship connections between those two regions. The denominator of (3.1) is the total number of possible connections, i.e. the possible pairs of users<sup>1</sup>.

The data have been collected and processed following the methodology described in [10]. The data analyzed in this report refers to a snapshot of the active Facebook users on October 13, 2021. This dataset should not be compared with older ones, since the SCI methodology has changed over time.

The geographic locations are obtained using the *Database of Global Administrative Areas* (GADM, version 2.8) and the *European Nomenclature of Territorial Units for Statistics* (NUTS 2016). European countries are divided into their NUTS3 regions (e.g. 110 regions in Italy), while most other countries are usually divided into their GADM level 1 regions<sup>2</sup>. The United States, Canada, and some countries in South Asia are divided into their GADM level 2 regions (e.g., US counties) due to their sizes.

The task is to build a - possibly weighted - network for each country. The output consists of two files: one for nodes, including latitude and longitude for each location, and one for edges list.

### 3.2 | Methods and analysis

---

The analysis leverages three R libraries: `geodata` [12] (for GADM regions), `eurostat` [13] (for NUTS regions), and `tigris` [14] (for USA counties). Briefly, the pipeline consists of associating each region to the corresponding country and level (e.g. GADM1,

<sup>1</sup>If  $i$  and  $j$  are the same, the denominator is adjusted as  $FB - users(i) * (FB - users(j) - 1)$  since users cannot be friends with themselves.

<sup>2</sup>Countries with a population less than 1 million are not divided.

GADM2, ...), extracting the metadata of each node (latitude, longitude, name, ID), and creating the node table and the edge list.

The major limitation of this analysis regards the GADM version: the SCI data uses version 2.8 (released in 2015), while the oldest online available version is 3.6 (released in 2020). Version 2.8 should be available through the official GADM website [15], but no shapefile can currently be downloaded. Since the official log is under development, estimating what biases we are introducing using a different version is not trivial. The only observable consequence is that two countries (United Arab Emirates and Indonesia) were expected to have one more region each.

### 3.3 | Comparisons

All the obtained graphs are *complete* (at least one friendship relation for each pair of regions of the same country) and *undirected*. Thus, many of the classical network metrics are trivial. However, one can consider the *weighted* version of the networks and base the comparisons out of that.

In particular, figure 3.1 (left) shows four analytics implemented in the `igraph` R package and how they relate with each other and with the number of network nodes. The mean betweenness, clustering coefficient, and weighted distance are correlated with the number of nodes. The feature that exhibits the largest variability is the mean entropy of the networks, which measures the global diversity between the vertices based on the Shannon entropy of their incident edges.

Figure 3.1 (right) analyses the full world network of social connectivity and investigates the relation between the haversine distance and the strength of such connectivity. As expected, there is a negative correlation between these two quantities: closer areas tend to have stronger connections.

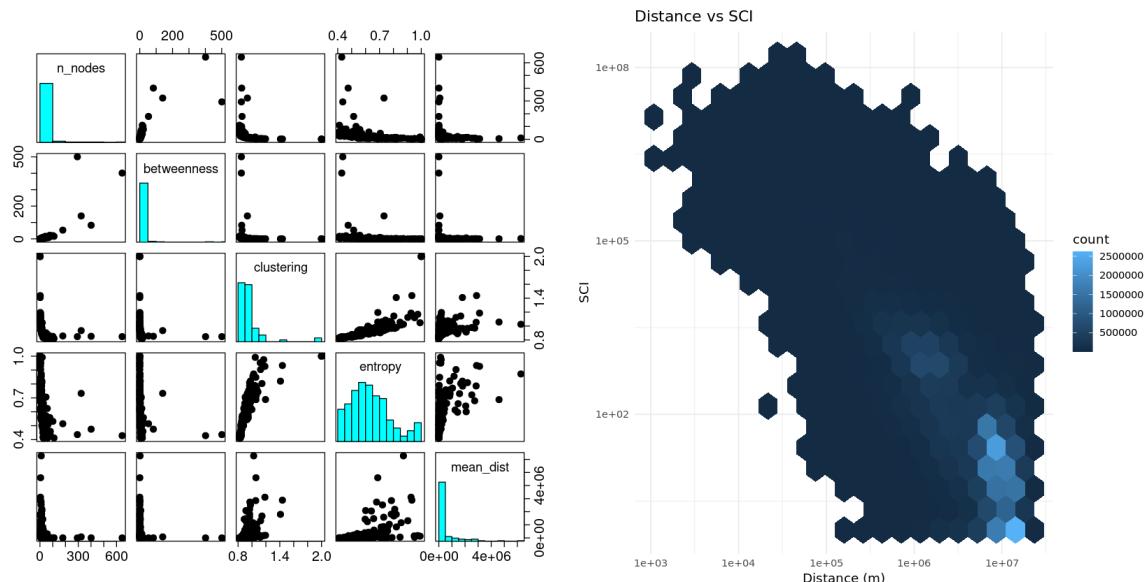


Figure 3.1: (left) The pairwise relations between different metrics for the weighted networks. Each point is a country. (right) 2D histogram of the distribution of the SCI against the haversine distance.

## 4 | Bibliography

---

- [1] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, 59(4):381, 1987.
- [2] K-I Goh, D-S Lee, B Kahng, and D Kim. Sandpile on scale-free networks. *Physical review letters*, 91(14):148701, 2003.
- [3] Eric Bonabeau. Sandpile dynamics on random graphs. *Journal of the Physical Society of Japan*, 64(1):327–328, 1995.
- [4] Preben Alstrøm. Mean-field exponents for self-organized critical phenomena. *Physical Review A*, 38(9):4905, 1988.
- [5] Charles D Brummitt, Raissa M D’Souza, and Elizabeth A Leicht. Suppressing cascades of load in interdependent networks. *Proceedings of the national academy of sciences*, 109(12):E680–E689, 2012.
- [6] Package networkx. <https://networkx.org/documentation/stable/index.html>. [Accessed 4-Sept-2024].
- [7] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.
- [8] Claudio Castellano, Matteo Marsili, and Alessandro Vespignani. Nonequilibrium phase transition in a model for social influence. *Physical Review Letters*, 85(16):3536, 2000.
- [9] Maxi San Miguel, Victor M Eguiluz, Raul Toral, and Konstantin Klemm. Binary and multivariate stochastic models of consensus formation. *Computing in Science & Engineering*, 7(6):67–73, 2005.
- [10] Social connectedness index - dataforgood.facebook.com. <https://dataforgood.facebook.com/dkg/tools/social-connectedness-index>. [Accessed 13-Aug-2024].
- [11] Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–280, 2018.
- [12] Package geodata - cran. <https://cran.r-project.org/web/packages/geodata/index.html>. [Accessed 13-Aug-2024].

- [13] Package eurostat - cran. <https://cran.r-project.org/web/packages/eurostat/index.html>. [Accessed 13-Aug-2024].
- [14] Package tigris - cran. <https://cran.r-project.org/web/packages/tigris/>. [Accessed 13-Aug-2024].
- [15] Gadm maps and data. <https://gadm.org/index.html>. [Accessed 13-Aug-2024].

A

# Supplementary material

## A.1 | SOC: other figures

The following two figures report some analysis of the duration of each avalanche for the sandpile model described in 1. The duration is here operationally defined as the number of times the queue of unstable nodes has been updated, i.e. the number of times a toppling event leads to other sites being unstable.

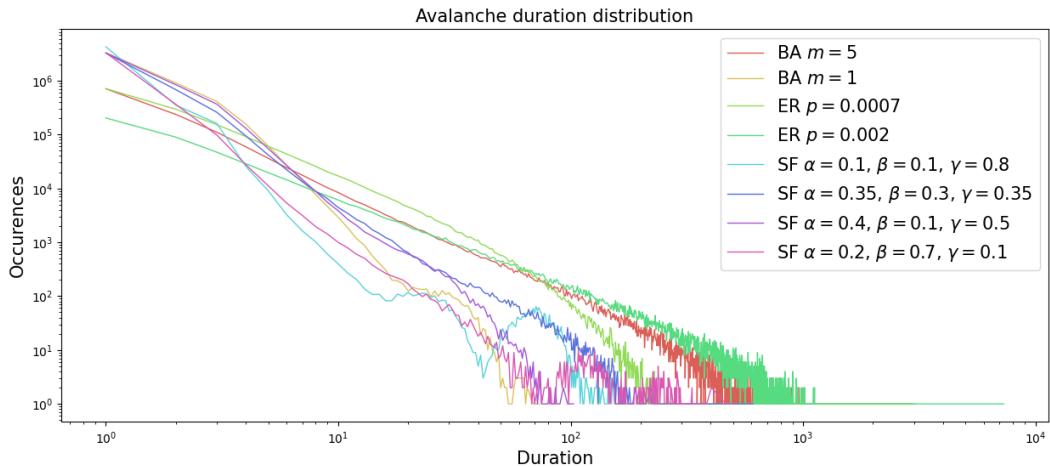


Figure A.1: Distribution of the duration of each avalanche.

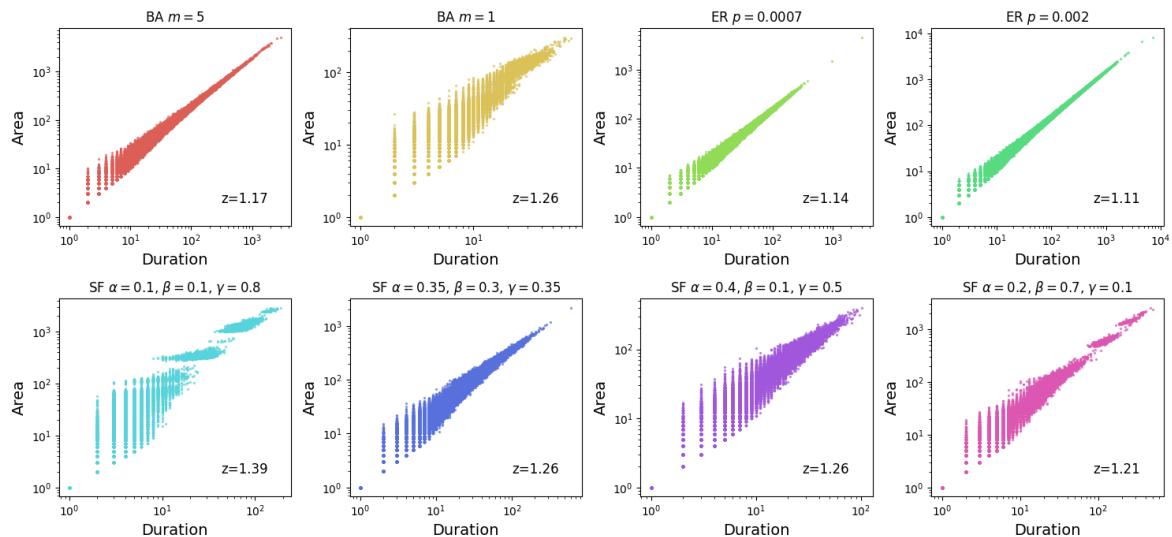


Figure A.2: Relation between the avalanche area and the duration of the cascade.

## A.2 | SOC: interdependent networks

I simulated the sandpile dynamics over a coupled  $R(3)$ - $B(p)$ - $R(3)$  networks, as described in [5]. Each network is a  $z = 3$  regular graph with  $N = 2 \times 10^3$  nodes. The coupling is achieved by selecting uniformly at random  $l$  nodes from each network with probability  $p$  and connecting the correspondent edge stubs. For each value of  $p$ , eight independent realizations have been analyzed.

The simulations last  $5 \times 10^7$  iterations with a dissipation rate  $f = 0.01$ .

Even with this simplified analysis, one can appreciate the minimum in the number of large avalanches in network  $a$ , obtained by tuning the coupling probability around  $p = 0.07$ .

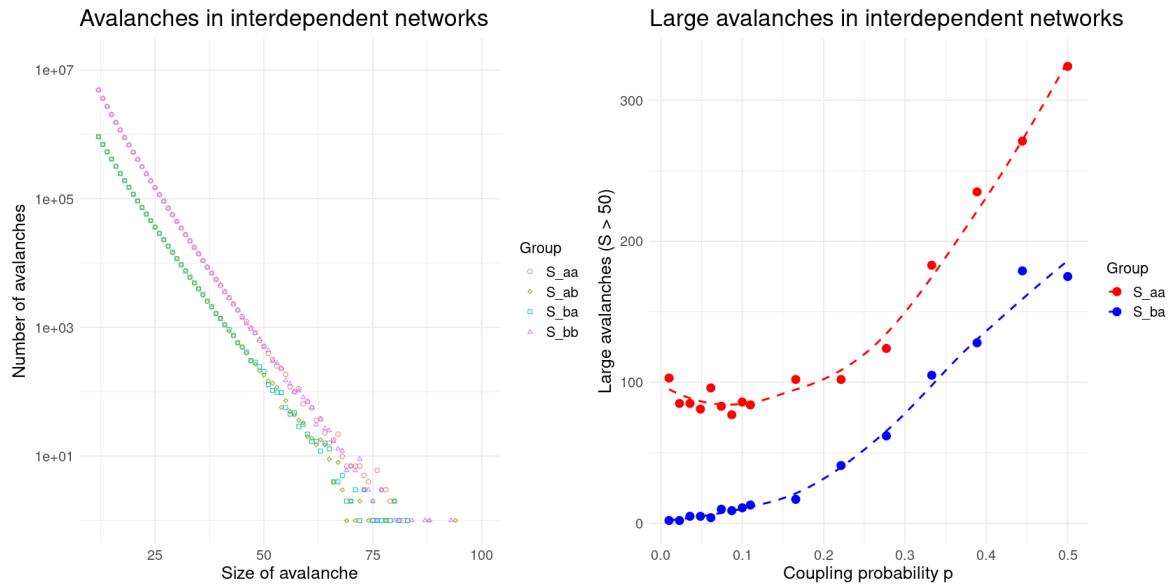


Figure A.3: (left) The avalanche size distribution in interdependent networks. Only values for  $S > 10$  are shown. (right) Number of large avalanches ( $S > 50$ ) in network  $a$  as a function of the coupling probability  $p$ .  $S_{aa}$  and  $S_{ba}$  refer to avalanches started in network  $a$  and  $b$  respectively.

### A.3 | Axelrod's model: other figures

The following section contains the figures regarding Axelrod's dynamics on top of all the networks analyzed in this report. The general analysis is reported in the main text (see chapter 2). Each network has  $2 \times 10^3$  nodes and the results refer to 4 independent simulations of  $5 \times 10^7$  iterations each. Moreover, the dynamics have been tested also on top of a  $L \times L$  lattice ( $L = 50$ ). The simulations last for  $10^{10}$  iterations, repeated 3 independent times. This numerical analysis should be compatible with the ones in the literature.

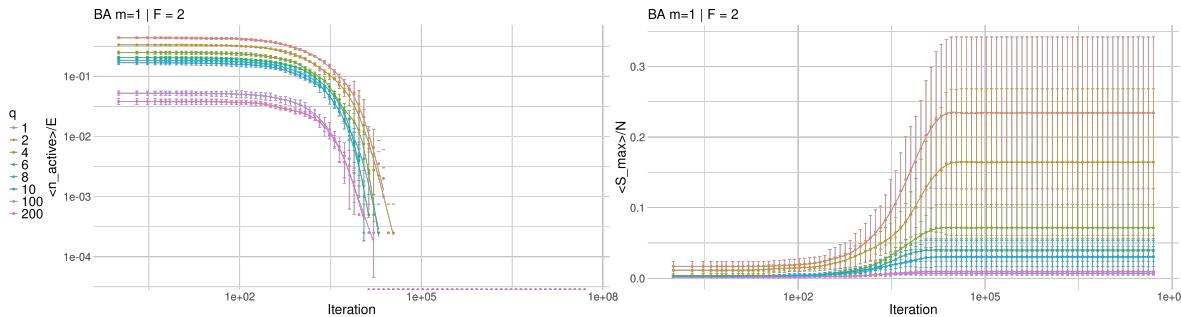


Figure A.4: Barabasi-Albert with  $m = 1$ .  $F = 2$ .

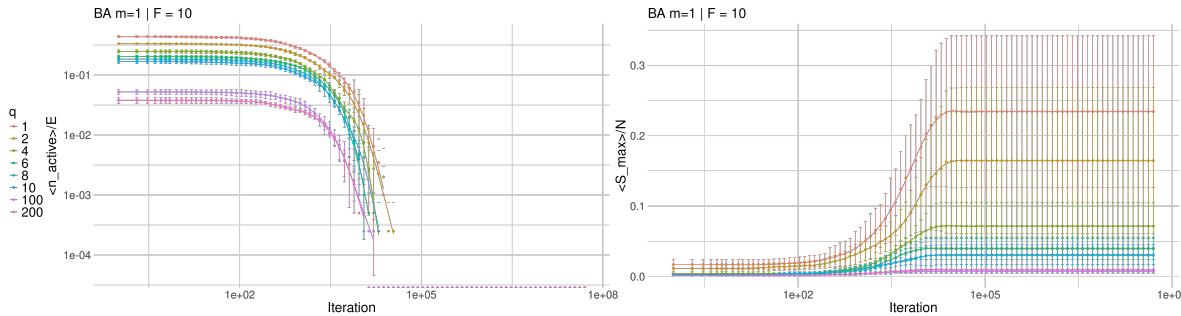


Figure A.5: Barabasi-Albert with  $m = 1$ .  $F = 10$ .

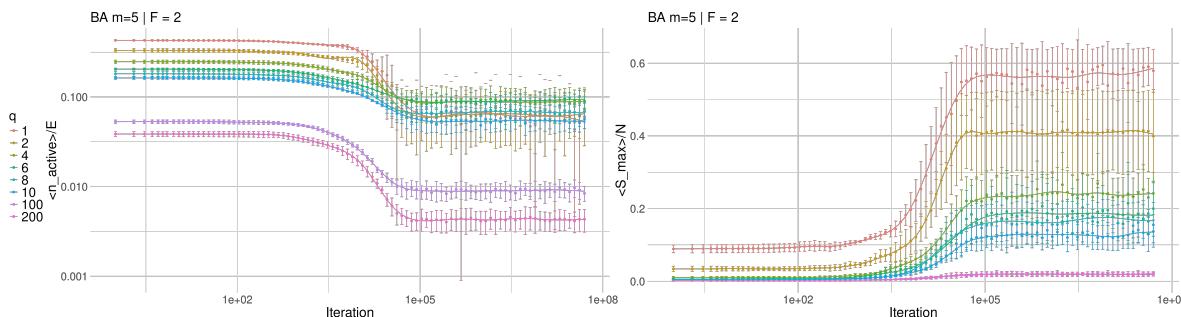


Figure A.6: Barabasi-Albert with  $m = 5$ .  $F = 2$ .

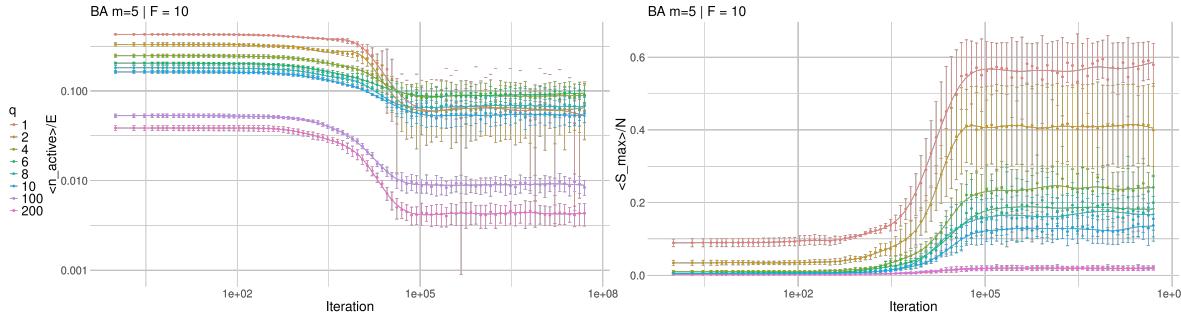


Figure A.7: Barabasi-Albert with  $m = 5$ .  $F = 10$ .

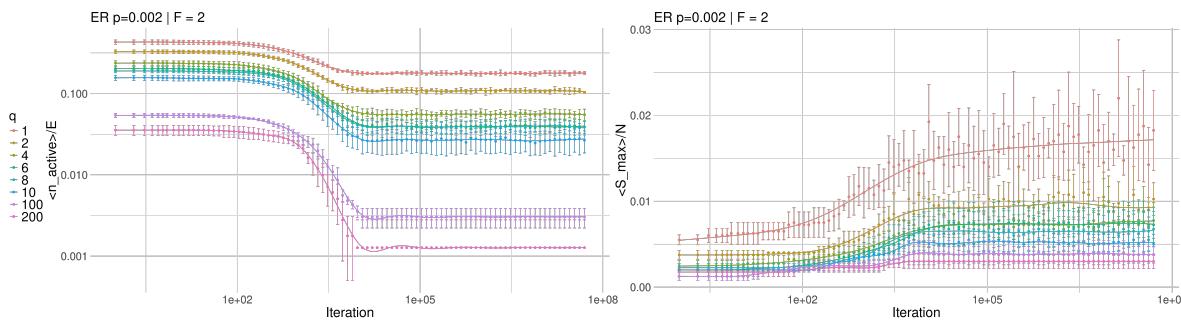


Figure A.8: Erdős–Rényi with  $p = 2 \times 10^{-3}$ .  $F = 2$ .

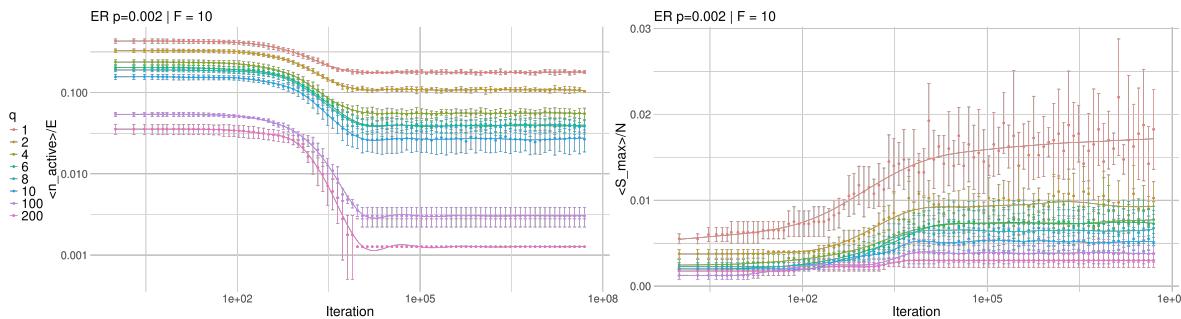


Figure A.9: Erdős–Rényi with  $p = 2 \times 10^{-3}$ .  $F = 10$ .

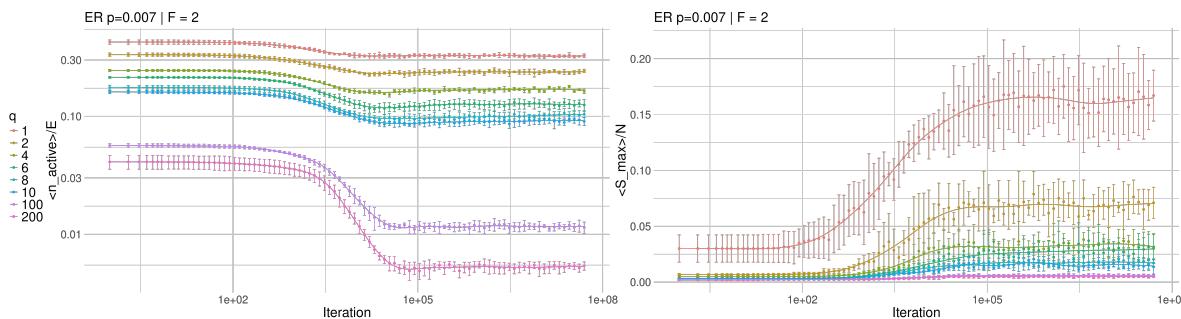
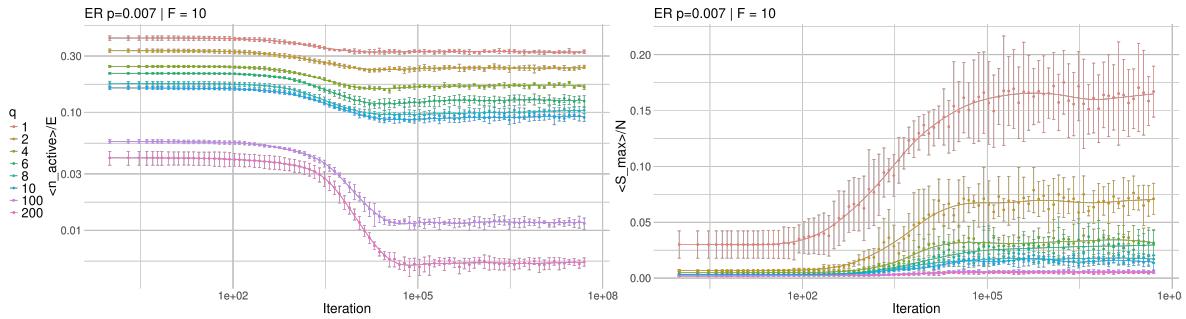
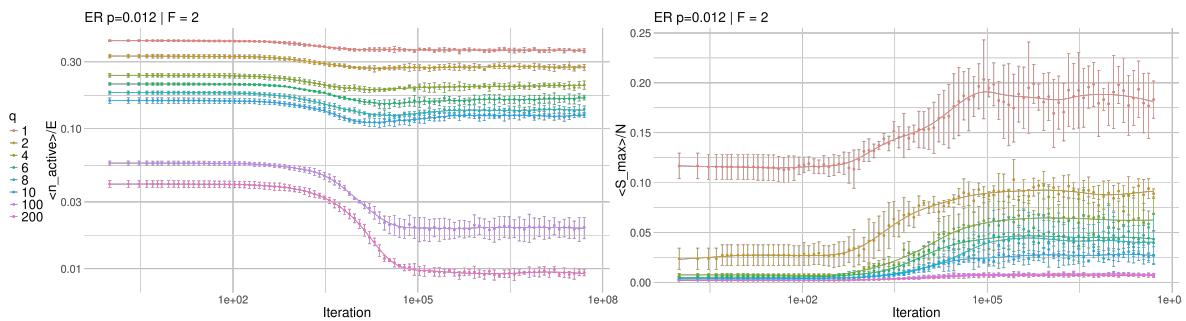
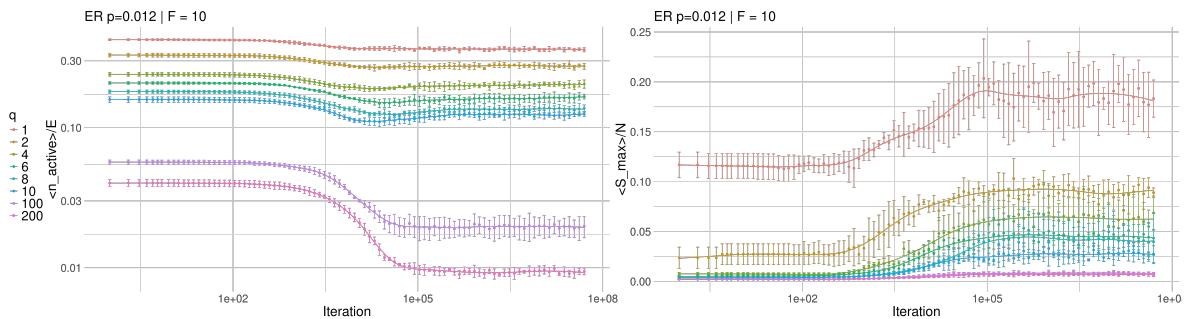
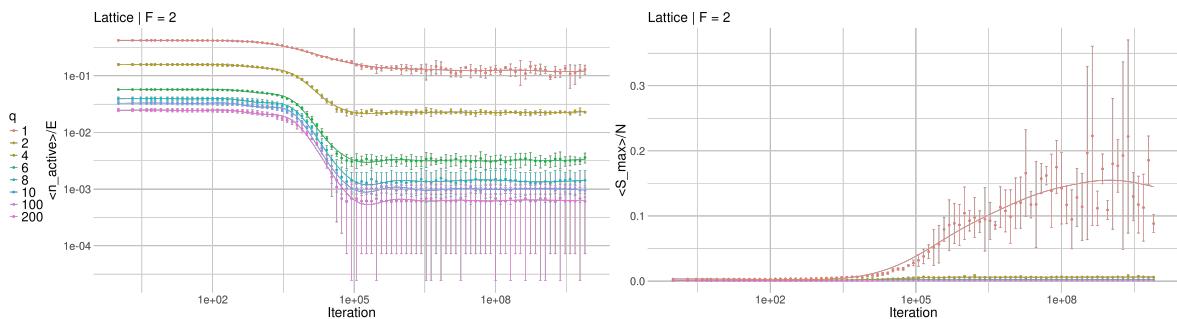


Figure A.10: Erdős–Rényi with  $p = 7 \times 10^{-3}$ .  $F = 2$ .

Figure A.11: Erdős–Rényi with  $p = 7 \times 10^{-3}$ .  $F = 10$ .Figure A.12: Erdős–Rényi with  $p = 1.2 \times 10^{-2}$ .  $F = 2$ .Figure A.13: Erdős–Rényi with  $p = 1.2 \times 10^{-2}$ .  $F = 10$ .Figure A.14: Regular lattice  $L = 50$ .  $F = 2$ .

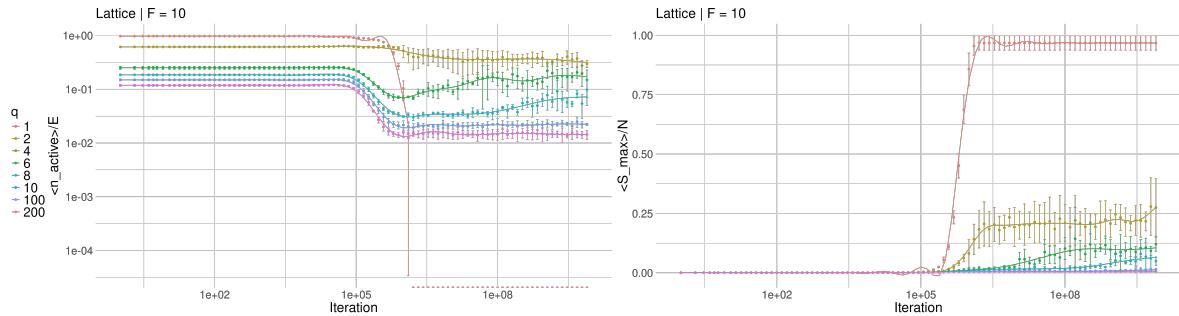


Figure A.15: Regular lattice  $L = 50$ .  $F = 10$ .

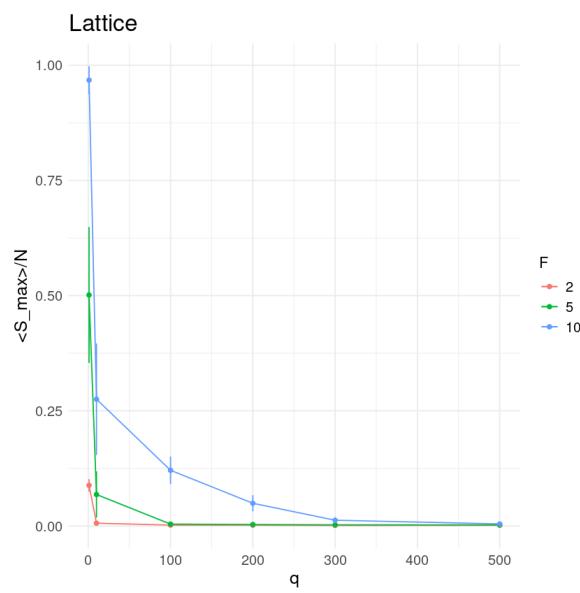


Figure A.16: Size of the largest cultural domain after  $10^{10}$  iterations as a function of the initial disorder  $q$ .