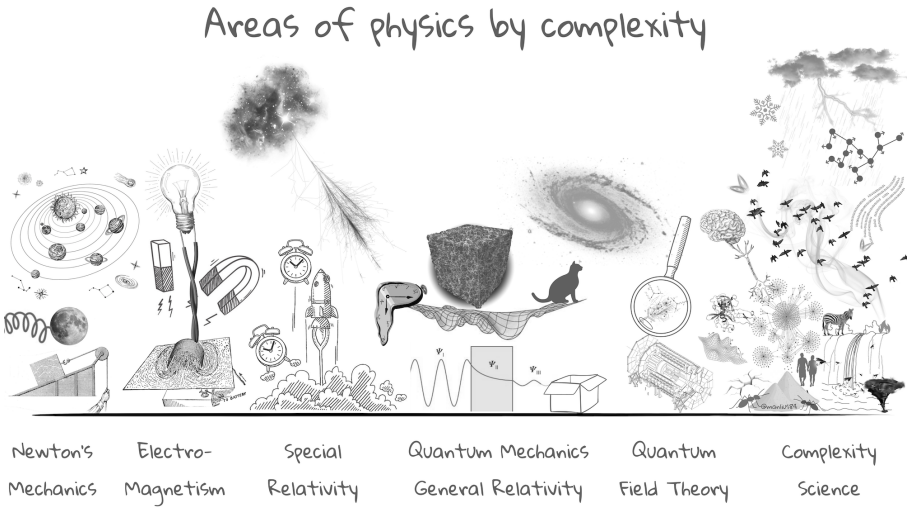
Final Report

Physics of Complex Networks: Structure and Dynamics

Università degli Studi di Padova



PoCN Projects

Tramarin, Chiara

Last update: February 15, 2026

Contents

1 | Task #15 – Sandpile Model for Self-Organized Criticality on Complex Networks

1.1 | Model and Theoretical Background

Self-organized criticality (SOC) refers to the spontaneous emergence of scale-invariant dynamics in slowly driven, dissipative systems [3]. The Bak–Tang–Wiesenfeld (BTW) sandpile model provides a paradigmatic example: under slow driving, the system self-organizes into a marginally stable state where small perturbations may trigger cascades (avalanches) spanning a broad range of sizes.

We consider the BTW dynamics on a network $G(V, E)$ with N nodes. Each node i carries an integer load h_i. At each driving step, a grain is added to a randomly selected node,

$$h_i \to h_i + 1. \tag{1.1}$$

Whenever the load exceeds a local threshold z_i, the node topples and redistributes one unit of load to each neighbor,

$$h_i \to h_i - k_i, \qquad h_j \to h_j + 1 \quad \forall j \in \mathcal{N}(i), \tag{1.2}$$

where k_i denotes the degree of node i. In order to ensure stationarity and prevent indefinite mass accumulation, a small dissipation probability is introduced during grain transfer.

A single grain addition may trigger an avalanche characterized by its size S (total number of toppling events), area A (number of distinct toppled nodes), and duration T. On locally tree-like random networks one typically finds $A \sim S$ [4], so that the avalanche size distribution takes the form

$$P(S) \sim S^{-\tau} e^{-S/S_c}, \tag{1.3}$$

where τ is the avalanche exponent and S_c is a cutoff scale induced by finite size and dissipation.

The critical behavior can be understood through a mapping to a branching process [4]. For networks with finite degree variance, such as Erdős–Rényi graphs, the mean-field exponent $\tau = \frac{3}{2}$ is recovered. For scale-free networks with degree distribution $P_d(k) \sim k^{-\gamma}$, the behavior depends on γ. When $\gamma > 3$, the degree variance

remains finite and the mean-field exponent persists. For $2 < \gamma < 3$, the divergence of the degree variance modifies the effective branching statistics, yielding [4] $\tau = \frac{\gamma}{\gamma-1}$. At $\gamma = 3$, logarithmic corrections to scaling are expected. This topology-dependent behavior arises only when the threshold is degree-dependent ($z_i = k_i$).

1.2 | Simulations

The primary objective is to reproduce the main theoretical result of the model, namely the dependence of the avalanche size exponent τ on the degree exponent γ. We test the recovery of the mean-field value $\tau = 3/2$ for $\gamma > 3$ and its predicted modification in the heterogeneous regime $2 < \gamma < 3$.

Networks were generated using the static model with tunable exponent γ. We considered $\gamma = \infty$, 4.0, 3.0, 2.6, and 2.2, together with an Erdős–Rényi network as an independent mean-field reference. Simulations were performed with $N = 20000$ nodes and average degree $\langle k \rangle \simeq 4$. After discarding 10^5 transient steps, 2×10^5 avalanches were recorded. The relatively small system size enhances finite-size effects and restricts the observable scaling window.

Dissipation was implemented with probability f during grain transfer. Since f controls the cutoff scale S_c, preliminary tests were performed for $f = 10^{-3}$, 10^{-4}, and 10^{-5} (Appendix). Decreasing f shifts the cutoff toward larger S without significantly altering the initial scaling region, consistent with $P(S) \sim S^{-\tau} e^{-S/S_c}$. For the main analysis we adopt $f = 10^{-4}$ as a compromise between an extended intermediate regime and computational feasibility: smaller f increases S_c but substantially slows down the dynamics, while larger f suppresses the scaling range.

The primary observable is the avalanche size S. Exponent estimation is performed using the complementary cumulative distribution function (CCDF), which is more robust than direct PDF fits in the presence of noise and cutoff effects. If $P(S) \sim S^{-\tau}$, then the CCDF scales as $S^{-(\tau-1)}$, allowing extraction of τ from the log–log slope. Fits are restricted to intermediate windows excluding both the small-S discreteness regime and the large-S cutoff region.

To verify that deviations from theoretical values are not artifacts of dissipation, bulk-only avalanches (no grain loss) were also analyzed, as done in the original article (Appendix). Although bulk cascades better approximate the conservative limit, they drastically reduce tail statistics at finite N, shortening the effective scaling window. In our simulations, bulk-based fits yield significantly larger exponent estimates but are highly unstable due to the limited number of large events. For this reason, exponent measurements in the main text use the full avalanche ensemble.

Figure 1.1 shows the CCDF for different γ. All curves display the expected finite-size structure: an initial curved region, an intermediate quasi-linear regime, and a dissipation-induced cutoff. For $\gamma = \infty$ and $\gamma = 4.0$, fits in $S \in [20, 200]$ yield effective exponents $\tau_{\text{eff}} \simeq 1.30$, below the asymptotic mean-field value $\tau = 3/2$. The Erdős–Rényi baseline gives a comparable estimate, indicating that the discrepancy arises from finite-size and cutoff effects.

For $\gamma = 3.0$, $\tau_{\text{eff}} \simeq 1.27$ is obtained in $S \in [20, 180]$. For $\gamma = 2.6$ and $\gamma = 2.2$,

increasing heterogeneity visibly alters the distribution, but no sufficiently extended linear regime emerges before the cutoff. Exponent estimates in this regime are therefore window-dependent and should be interpreted as effective finite-size values rather than asymptotic exponents. At $N = 20000$, the theoretical prediction $\tau = \gamma/(\gamma - 1)$ for $2 < \gamma < 3$ cannot be quantitatively verified due to the limited scaling range.

Network	Fit window	τ_{eff}	R^2
Static ($\gamma = \infty$)	[20, 200]	1.303	0.989
Static ($\gamma = 4.0$)	[20, 200]	1.299	0.985
Static ($\gamma = 3.0$)	[20, 180]	1.273	0.973
Erdős–Rényi	[20, 200]	1.313	0.989

Table 1.1: Effective avalanche exponents obtained from CCDF fits in log–log scale. For $\gamma \leq 2.6$, no stable scaling window is observed at this system size.

Figure 1.2 (Appendix) reports the corresponding log-binned probability density functions, which emphasize the crossover structure and the dissipation-induced cutoff limiting the observable scaling region.

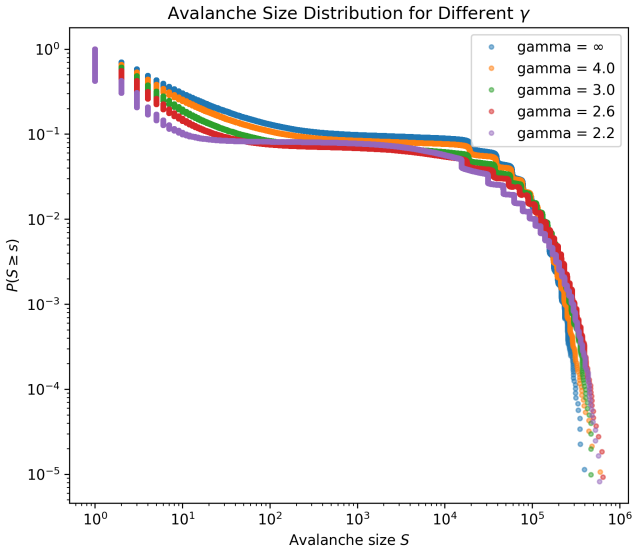


Figure 1.1: CCDF of avalanche size S for static-model networks with different γ. An intermediate quasi-linear regime is visible for $\gamma > 3$, while increasing heterogeneity reduces the apparent scaling window and enhances cutoff effects.

1.3 | Appendix: Additional Numerical Diagnostics

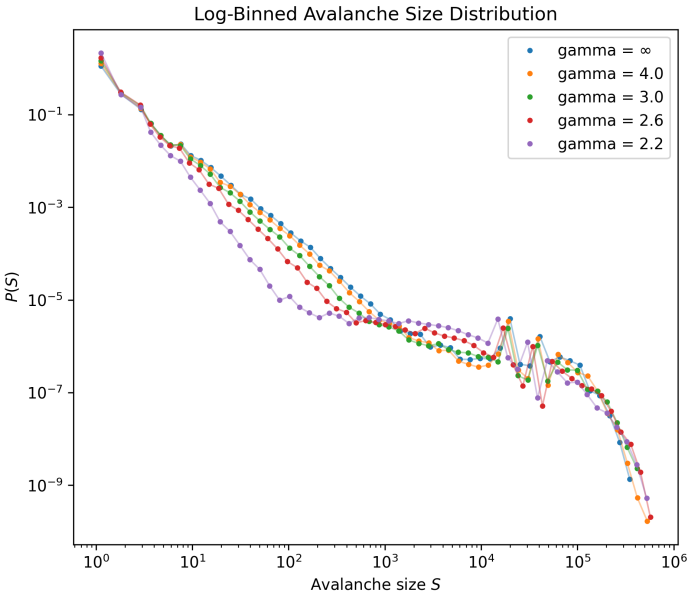


Figure 1.2: Log-binned avalanche size distribution $P(S)$ for the same simulations. The limited extent of the apparent linear region and the exponential cutoff explain the systematic underestimation of the asymptotic exponent at finite N.

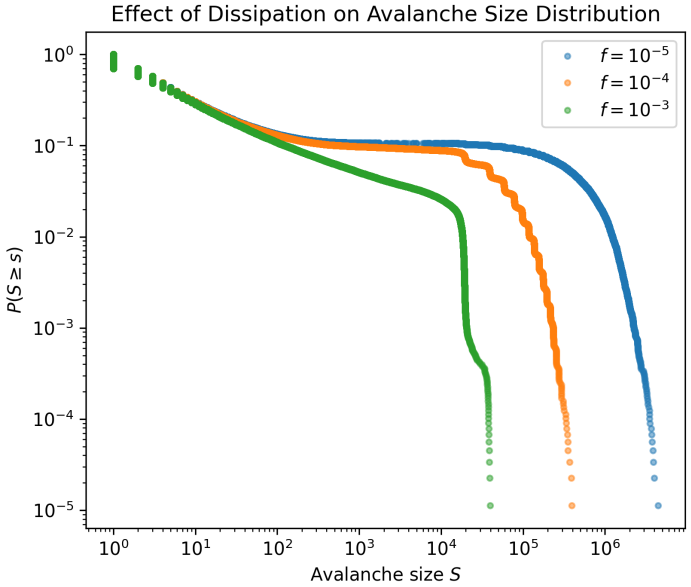


Figure 1.3: Effect of dissipation on the avalanche size CCDF for $\gamma = \infty$. Decreasing f shifts the cutoff toward larger avalanche sizes, confirming the expected dependence of the characteristic scale on dissipation.

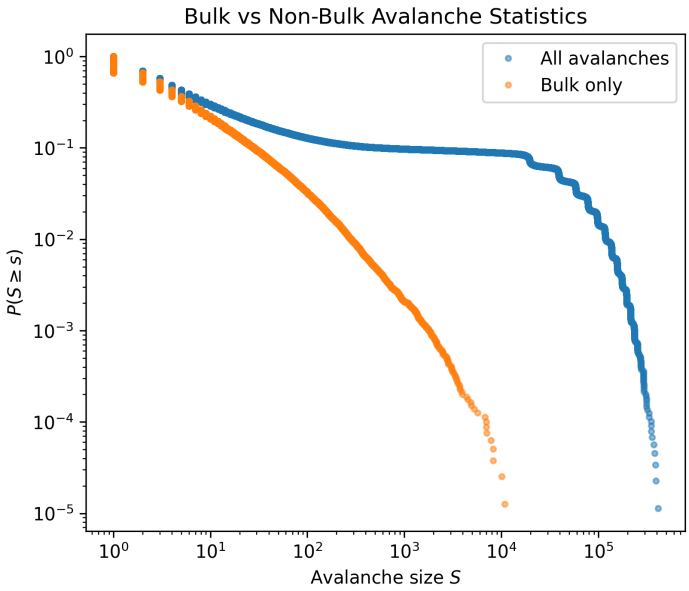


Figure 1.4: Comparison between the full avalanche ensemble and bulk-only cascades. Bulk filtering significantly suppresses large events at finite system size, reducing the effective scaling window and destabilizing exponent estimates.

2 | Task # 44 – Social Connectedness Index II from Facebook

2.1 | Social Connectedness Index

The Social Connectedness Index (SCI) quantifies the intensity of social ties between pairs of geographic regions based on the prevalence of mutual friendship links on Facebook. The index was first introduced by Bailey et al. [1] to study social connectedness across U.S. counties and has since been extended to subnational and international contexts using anonymized snapshots of active Facebook users and their friendship networks.

For two distinct locations i and j, the SCI is formally defined as

$$\mathrm{SCI}_{i,j} = \frac{\mathrm{FB\_Connections}_{i,j}}{\mathrm{FB\_Users}_i \cdot \mathrm{FB\_Users}_j}, \tag{2.1}$$

where $\mathrm{FB\_Connections}_{i,j}$ denotes the number of friendship links between users in locations i and j, and $\mathrm{FB\_Users}_i$ and $\mathrm{FB\_Users}_j$ are the numbers of active Facebook users in the respective locations. For self-pairs ($i = j$), the denominator is adjusted to account for the absence of self-friendship links.

In the publicly released datasets, SCI values are rescaled within each dataset to lie between 1 and 1,000,000,000. To preserve user privacy, a small amount of random noise is added and values are rounded to the nearest integer before release [1].

The SCI dataset can therefore be interpreted as an undirected, weighted graph in which nodes represent geographic areas and edge weights encode the relative strength of social connectedness between regions.

2.2 | Dataset and Network Construction

The construction of country-level social networks from the Facebook Social Connectedness Index (SCI) data proceeded through multiple stages of data integration, spatial mapping, and network building. The principal input consists of two files from the October 2021 release [6, 5]: `gadm1_nuts3_counties_levels.csv`, a reference table where each geographic identifier (`key`) is annotated by its administrative level (e.g., `country`, `GADM1`, `GADM2`, `NUTS3`); and `gadm1_nuts3_counties-gadm1_nuts3_counties.csv`, containing scaled SCI values for pairs of regions (`user_loc`, `fr_loc`, `scaled_sci`).

To associate region keys in the SCI file with standardized geographic units, identifiers were matched to the reference table. Spatial geometries for these units were then obtained using R packages such as `geodata` and `eurostat`, which download official shapefiles. For each polygon, a representative point was computed to assign latitude and longitude coordinates to nodes. Regions lacking valid geometry or representative points were excluded to maintain spatial consistency.

European regions were matched to NUTS3 shapefiles from the official NUTS 2016 classification, as done in the original SCI documentation [2]. For non-European regions, we used GADM version 3.6 administrative units in place of the original GADM 2.8 geometries referenced in the SCI data, as the latter are no longer readily obtainable. Because the SCI data rely on GADM 2.8 for non-European region definitions, exact spatial matching with GADM 3.6 geometries is not always possible, and regions without valid spatial geometries were excluded from the analysis.

Only links between regions within the same country were retained, so that each network reflects internal social connectedness; links involving regions from different countries were excluded. To ensure sufficient data coverage and robust network statistics, the analysis was restricted to the top 100 countries ranked by the number of available subnational regions. Region identifiers were normalized across differing coding schemes to ensure consistent spatial matching, and hierarchical priorities were adopted when multiple administrative resolution levels existed (NUTS3 > GADM2 > GADM1). Self-loops (edges where a region is paired with itself) were removed, and symmetric duplicates (i, j) and (j, i) in the SCI file were merged by averaging their scaled SCI values to produce undirected edge weights. No thresholding was applied to the SCI weights, as additional filtering would introduce arbitrary assumptions and alter the network structure.

For each country, the network construction procedure produced two CSV output files with the following columns:

- `nodes.csv`: `nodeID`, `nodeLabel`, `latitude`, `longitude`;
- `edges.csv`: `nodeID_from`, `nodeID_to`, `country_name`, `country_ISO3`.

These files serve as the basis for subsequent network analysis and visualization.

2.3 | Simplified Network Analysis

A simplified network analysis was performed to compare the intra-country SCI networks.As a preliminary sanity check, we first inspected the basic structural and weighted properties of the intra-country SCI networks. Figure 2.1 shows that the number of edges grows almost quadratically with the number of nodes, indicating that the networks are highly dense. As a consequence, purely topological metrics, such as degree, clustering, or centrality measures, are weakly discriminative across countries and provide limited insight into structural differences.

The distribution of SCI edge weights, also shown in Figure 2.1, spans several orders of magnitude, revealing a strongly heterogeneous structure. This suggests that the informative content of the SCI networks lies primarily in the weights of the connections

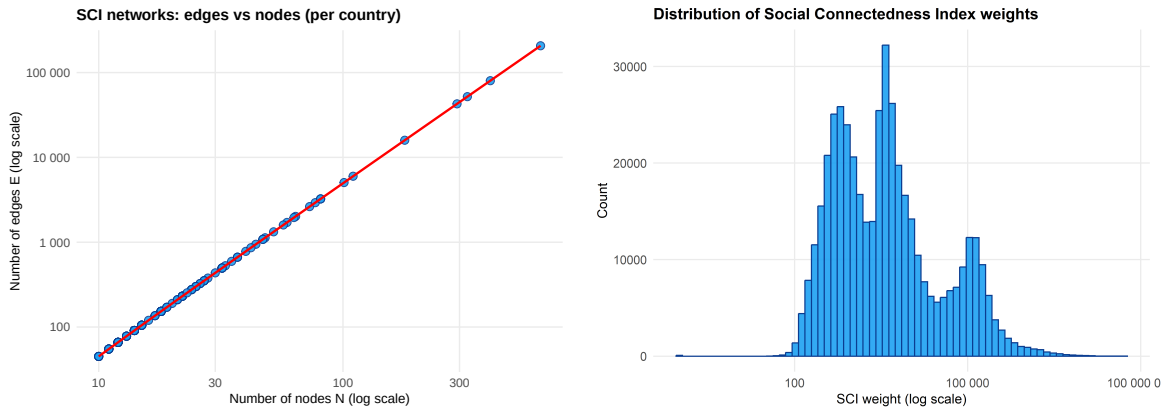


Figure 2.1: **Left:** Number of edges E versus number of nodes N for the intra-country SCI networks. The red line indicates the complete-graph reference $E = N(N-1)/2$. **Right:** Distribution of Social Connectedness Index (SCI) edge weights across all countries on a logarithmic scale.

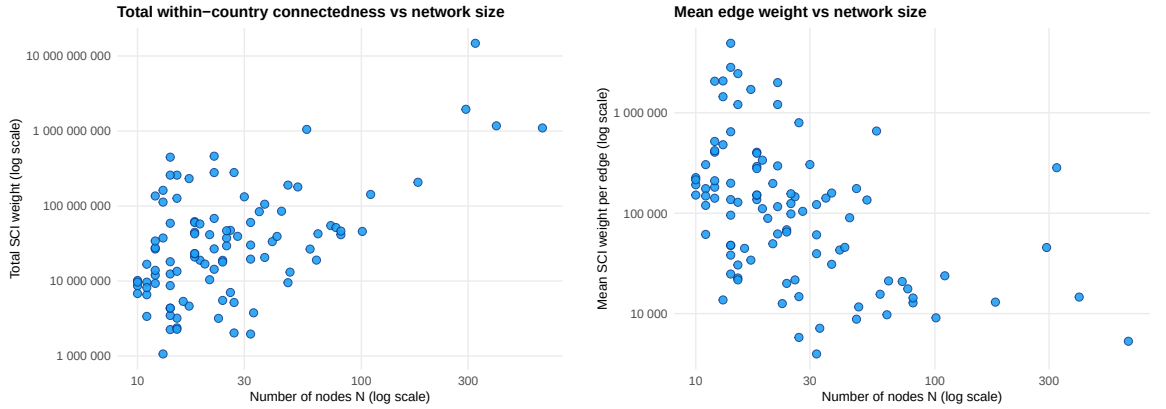


Figure 2.2: **Left:** Total within-country social connectedness, measured as the sum of SCI weights, as a function of network size. **Right:** Mean SCI edge weight as a function of network size.

rather than in their mere presence or absence. For this reason, no thresholding was applied to the SCI weights, as the index does not exhibit a natural cutoff and filtering would introduce arbitrary assumptions.

Figure 2.2 reports the total within-country social connectedness as a function of network size. While the total weight increases with the number of nodes, a substantial dispersion is observed among countries with comparable sizes. To remove the trivial dependence on network size, the same figure shows the mean SCI edge weight as a function of the number of nodes, highlighting non-trivial differences in the average intensity of social ties across countries. Overall, these results indicate that meaningful cross-country comparisons in SCI networks are driven by weighted properties rather than by topological features.

3 | Bibliography

[1] 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.

[2] Michael Bailey, Drew Johnston, Theresa Kuchler, Dominic Russel, Bogdan State, and Johannes Stroebel. The determinants of social connectedness in europe. In *Social Informatics*. Springer, 2020.

[3] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, 59(4):381–384, 1987. doi: 10.1103/PhysRevLett.59.381.

[4] Kwang-Il Goh, Do-Soon Lee, Byungnam Kahng, and Doochul Kim. Sandpile on scale-free networks. *Physical Review Letters*, 91(14):148701, 2003. doi: 10.1103/PhysRevLett.91.148701.

[5] Humanitarian Data Exchange. Social connectedness index dataset. https://data.humdata.org/dataset/social-connectedness-index, 2021. Accessed: 2026-02.

[6] Meta Platforms, Inc. Social connectedness index. https://ai.meta.com/ai-for-good/datasets/social-connectedness-index/, 2021. Accessed: 2026-02.