IMT School for Advanced Studies, Lucca
Lucca, Italy

Scale-Invariant Random Graphs: a multiscale approach to network modeling

PhD Program in Systems Science

XXXIV Cycle

By

Margherita Lalli

2024

The dissertation of Margherita Lalli is approved.

PhD Program Coordinator: Prof. Alberto Bemporad, IMT School for Advanced Studies Lucca

Advisor: Prof. Diego Garlaschelli, IMT School for Advanced Studies Lucca

The dissertation of Margherita Lalli has been reviewed by:

Prof. M. Ángeles Serrano, Universitat de Barcelona

Prof. Andrea Gabrielli, Università degli Studi "Roma Tre"

IMT School for Advanced Studies Lucca
2024

Contents

List of Figures

Acknowledgements

Each chapter in this thesis is based on a work by the author and collaborators, either published [1], or under submission [2, 3]. At the beginning of each chapter, the corresponding paper is referenced. However, I would like to explicitly address here the scientists who have taken an active part in this work, which has derived uncountable benefits from their contribution. A particular thank goes to Diego Garlaschelli, for the enlightening black-board sessions and the gentle nudges he has offered in the last years. I would also like to thank Luca Avena and Rajat Hazra for their mathematically rigorous and physically warm support, as well as Elena Garuccio for her hospitality.

I am also grateful to Mirco Tribastone, Rossana Mastrandrea, Tiziano Squartini, Fabio Saracco, Pierfrancesco Dionigi, Leonardo Ialongo, Alessio Catanzaro, and Riccardo Milocco for the inspiring discussions. Thanks to Flavio Bombacigno for his remarkable computing skills and his willingness to share them.

My gratitude also goes to Michela Tassi, Daniela Giorgetti, Barbara Iacobino, Martina Lucchesi, Federica Pierotti, and Emilia Spinetti for their always kind support.

Finally, I want to thank Fosca Giannotti, for her encouragement and patience in the final phase of this thesis.

Vita

December 3, 1989	Born in Rome (Italy)
2014	BSc in Physics Universitá di Roma La Sapienza, Rome (Italy)
2018	MSc in Theoretical Physics cum laude Final Thesis: Authorship Attribution via Sequence Compression Methods Supervisors: Dott. Francesca Tria and Prof. Vittorio Loreto Universitá di Roma La Sapienza, Rome (Italy)
2021	Guest Researcher Leiden University, Leiden (The Netherlands)
2022	Research Assistant IMT School for Advanced Studies, Lucca (Italy)
2023-2024	Research Assistant Scuola Normale Superiore, Pisa (Italy)

Publications

1. Lalli, M., Tria, F., and Loreto, V., *Data-compression approach to authorship attribution*, Elena Ferrante: A Virtual Author, 61-83 (2018).
2. Garuccio, E., Lalli, M., and Garlaschelli, D., *Multiscale network renormalization: scale-invariance without geometry*, Physical Review Research, 5(4), 043101 (2023).
3. Avena, L., Garlaschelli, D., Hazra, R. S., and Lalli, M., *Inhomogeneous random graphs with infinite-mean fitness variables*, arXiv preprint arXiv:2212.08462 (2022).
4. Raffaelli, G. T., Lalli, M., and Tria, F., *Innovation processes for inference*, arXiv preprint arXiv:2306.05186 (2023).
5. Lalli M. and Garlaschelli D., *Geometry-free renormalization of directed networks: scale-invariance and reciprocity*, arXiv preprint arXiv:2403.00235 (2024).

Presentations

1. M. Lalli, "Spectral coarse-graining in complex-networks", Workshop on "Network coarse-graining", November 2019, IMT Lucca.
2. M. Lalli, "Introduction of a Scale-Invariant random graph", NetSci 2020, Rome (online).
3. M. Lalli, "Scale-Invariance without geometry", Networks 2021, Washington (online).
4. M. Lalli, "Multiscale networks coarse-graining", Complex Networks 2021, Madrid (online).

Abstract

In the last decades, consistent efforts have been spent to capture specific shades of real systems through the development of *random graph models*, which have been studied extensively either for their practical value as statistical benchmarks and their theoretical appeal, as abstract tools capable of generating synthetic graphs with realistic properties.
In particular, establishing a robust representation of a graph at multiple scales of observation would enable considerable progress in the description, modeling, and control of real-world complex systems. Here, by building on the principles of renormalization group theory from statistical mechanics, we derive a random graph model precisely conceived to provide a statistically consistent description of a network for different resolutions of its units and *in an exact manner*.
We explore two interesting facets of the proposed model, which interlace with different branches of network science. On the one hand, it allows complying with empirical networks to provide up-scaled and down-scaled reconstructions according to a chosen hierarchy of partitions of the original nodes. In this sense, the model constitutes solid support for harboring a coarse-graining scheme of real systems without relying on any arbitrary introduction of a metric space. Secondly, this scale-invariant random graph itself turns out to generate networks with topological properties that are widespread among real-world systems and thus its mathematical sifting has its own theoretical interest.

Chapter 1

Introduction

In their simplest formulation, networks are mathematical objects composed of a collection of units connected in pairs by lines. Despite their simplicity, in the past decades they acquired a central role in complex systems modeling for their capability of providing profound insights about a given system by only focusing on the pattern of interactions among its fundamental units.
Reducing a system to such an abstract representation implies a net loss of information, which is nevertheless repaid by the opportunity to apply an extensive set of tools and techniques for extracting relevant information from very different systems. In particular, the advent of the computer age has incited an increasing interest in the fundamental properties of complex networks. Indeed, the surge of computational power allowed to store and investigate large datasets from the natural, technological, and social realms, and led to the acknowledgment that the emergent architectures, neither purely regular nor purely random, show profound similarities that suggested a common formation mechanism and blended the interest in network science of researchers from many different fields [4].

One of the most evident common features of real networks is that they can be *large*, and in some cases analyzing them in detail or providing a de-

terministic model for how they came to be is effectively impossible. Akin to the study of physical systems in statistical mechanics, this problem has been worked around by resorting to probability theory to consider *random graphs* as network models, which provide a probabilistic description of the local rules by which nodes are connected to one another [5].
In general, a random graph model is a network model in which the specific structure of the network is considered as randomly drawn, conditionally on a given set of parameters, which are taken as deterministic. In the oldest and most studied random graph, introduced by Paul Erdős and Alfréd Rényi in a series of seminal papers in the 1950s [6–8], the number of vertices n and the number of edges m are fixed, while any realization of the network is generated by progressively placing each of the m edges between pairs of nodes, uniformly chosen at random [1]. Notwithstanding the remarkable advantage of the Erdős - Rényi random graph of being exactly solvable for many of its average properties [9], it is not completely satisfactory for most of the real-world networks, since it differs from the latter in at least two of the universal aforementioned features. The first one, pointed out by Watts and Strogatz [10], deals with the high propensity of nodes from several real systems to cluster together, i.e. to form connected triads even in very sparse regimes. The tendency of nodes to be more likely connected if they share a connection with a third node finds a very natural reformulation in the context of social sciences since our common experience is that, usually, our social sphere grows by progressively including adjacent possible contacts (i.e. the friends of our friends). Moreover, a strong presence of short loops has been also found in technological and biological networks [11]. In the Erdős-Rényi random graph, the probabilities of vertex pairs to be connected are by definition independent, so that they are not affected by the existence of a mutual neighbour and thus the density of connected triads equals the density of links, thus converging to zero as the latter does.
A second widespread feature of real networks that defies the classical

[1] More precisely, a random graph model is defined as an *ensemble* of networks, i.e. a probability distribution over all possible networks. In Erdős - Rényi random graph, each realization is drawn by choosing uniformly at random among the set of all simple graphs with exactly n vertices and m edges.

random graph model concerns the degree distribution, and it has been particularly emphasized in the work of Albert, Barabási, and collaborators [12, 13]: the expected degrees of Erdős-Rényi random graphs exhibit a Poisson distribution where their empirical counterparts appear to have power-law tailed degree distributions, possibly with exponential cut-off ([12–16]). The presence of a long-tailed degree distribution, generally referred to as *scale free* property (since it implies the lack of a typical scale for the nodes' degree), induces a highly heterogeneous structure in the network, in which a small fraction of nodes (called *hubs*) hold the vast majority of connections and thus strongly affects the behavior of the whole system. This hierarchical organization has been found to have tremendous impacts on the dynamics of processes taking place on networks (among all the epidemic processes [17, 18]) and a large variety of random graphs models have been conceived to reproduce such heterogeneity in the successive years [13, 19–21].
Although exhibiting unrealistic clustering and degree distribution, the Erdős-Rényi random graph has proved extraordinarily useful as a source of insight into the structure of networks, among other reasons, for its capability of providing a grounded, well-understood benchmark. Indeed, although network modeling aims, in general, to mimic patterns of connections of real networks to aid the understanding of their formation mechanisms, there is no model capable of capturing just all about the features of an empirical system (if not without consistent overfitting), and every random graph focus on different aspects, depending on the specific nuances of the system under study and on the specific inquiry one would like to address. A paradigmatic class of random graphs with this regard is represented by the family of *Exponential Random Graphs* [22–24], originally introduced in social science and then reframed in the context of statistical mechanics by Park and Newman, that derived them from the principle of maximum entropy [23]. The ERGs provide the technical framework for sampling networks that exhibit, in average, certain local features of real-world systems but are otherwise maximally random [24].

Besides topological properties, one could be interested in intersecting further facets of real-world systems that can be modelled by complex

networks. While a univocal and formal definition of complexity is still under debate, one of the settled aspects, which constitutes the motivation for this thesis, is that complex systems usually exhibit structural features extending over multiple characteristic scales [25, 26]. In network science, methods developed for a multiple resolution screening of empirical networks allowed to detect a widespread modular structure, in which nodes divide naturally into communities (i.e. groups of nodes that are more densely connected), that in their turn are organized hierarchically across progressively higher-level structures [27–29]. This is resulting in an overall arrangement into distinct levels (not necessarily straightforwardly identified) that define a variety of scales of observation and can in some cases be ordered according to these scales. A straightforward example is provided by socio-economic networks, which are built hierarchically across several levels, from single individuals up to groups, countries, and whole geographical regions. The information concerning interactions among these entities may, at times, be accessible only at aggregate levels and may exhibit disparities across different units. For instance, the statistical agency of one country might report comprehensive and reliable data on all domestic and international transactions at the firm level, whereas another may provide such information only at the sector or even at the country level. This gives rise to a complex structure in which elements on different mesoscales come together to interact, in addition to the horizontal interactions that occur among elements at the same scale. Such hierarchical organization calls for robust multi-scale representations and models capable of capturing the properties of a system at arbitrary scales of resolution, for both practical and theoretical reasons, including the reconstruction of a network topology from partial information (due for example to confidentiality constraints) or the analysis of the propagation of information from lower to higher levels of aggregation.

Already in the second half of the last century, the analysis of physical systems at different scales successfully assisted the understanding of their critical behavior, leading to the development of the Renormalization Group (RG) theory in statistical mechanics [30–33]. The *real space* RGT can be loosely introduced as a formal apparatus that allows for a systematic

investigation of a system through a repeated coarse-graining scheme, whereby its elementary units are progressively merged according to a minimum-distance principle. The iteration of these coarse-graining steps, each one followed by a proper rescaling of the parameters in the Hamiltonian, induces the so-called renormalization flow, which turned out to be deeply related to the phase diagram of the system under study [32, 33]. Clearly, RG techniques are based on the availability of a valid source of geometric scaling, i.e. a metric distance defining 'how close' the units being coarse-grained should be and identifying possible characteristic lengths and scaling properties in the system. For this reason, the application of RG methods has been strongly limited in those cases where the system is not naturally embedded in a metric space, as is the case for many real-world systems modelled by complex networks. Indeed, the lack of an explicit geometric embedding implies that vertices do not necessarily have well-defined coordinates, and the renormalization procedure cannot be defined in a way similar to that employed for regular geometric lattices.

Nevertheless, consistent efforts have been devoted to addressing this limitation in recent decades. One of the earliest attempts proposes a box-covering procedure borrowed from fractal analysis and based on the identification of distance with the shortest path lengths between nodes [34, 35]. Grouping the nodes according to their topological distances has important limitations from practical and theoretical points of view: on the one hand, finding the optimal partition of nodes (that is, the minimum number of boxes that cover the system in such a way that the elements in each block are separated by a given distance) is a NP-hard problem, for which only approximate solutions can be found [36]; on the other, the shortest path length is a poor source of variability and scaling due to the *small-world property* exhibited by the majority of the real-world networks, by which most pairs of nodes are connected by short paths (implying that the renormalization flow leads to the limit of complete connectedness after only a few iterations) [10, 14, 37].

Another notable, more recent approach relies on the optimal embedding of nodes onto a 'latent' hyperbolic space that can then be used to guide

the renormalization scheme [38]. The groundwork is thus the adoption of a hidden space random graph model, where the connection probabilities among nodes are determined by their coordinates in a hyperbolic space whose geometry shapes the structure and induces the topology of the observed system [21, 39–41]. Driving the network into a metric space grants to elude the lack of geometry by providing a metric that allows for a clear definition of *scale of observation* and a natural implementation of a zooming-in and zooming-out procedure.
A latter strategy involves the Laplacian operator for graphs and leverages its spectral properties to introduce a diffusion-based renormalization scheme [42]. Provided that the considered network exhibits a non-trivial Laplacian susceptibility, this method is capable of successfully identifying a spatiotemporal hierarchical organization in the network and uncovering its mesoscale structures.
All the aforementioned approaches focus on the possibility of identifying hierarchical partitions of nodes within a given real network, while either neglecting the problem of defining a graph model consistent with the identified partitions or addressing it in a manner that is contingent on the specific methodology used to establish the coarse-graining scheme itself. However, it is in general desirable to resort to a multiscale network model that is independent of the specific scheme used to extract the hierarchy of partitions, for instance, because the notion of coexistence of multiple scales in networks may be more general than the one defining the coarse-graining method itself.

In this work we want to contribute to the discussion around multiscale modeling by addressing the following question: is it possible to define a random graph model only from scale-invariance requirements attained on a statistical description of the network ? And is a parsimonious choice of the parameters compatible with the complex structures exhibited at multiple scales? Under these terms, however, these are ill-posed issues, since, as pointed out, networks in general are not embedded in any metric space and therefore lack a clear definition of 'scale'. We will approach these issues in the three chapters that interweave in this thesis, which is

organized as follows.

In chapter 2, we first derive the random graph model that is the focus of this contribution (the Scale-Invariant Model) for undirected and binary networks. This is done by first defining a coarse-graining scheme that progressively reduces the size of a network and thus identifying each step in this coarse-graining flow with a different *scale of observation*. Then, inspired by Kadanoff's construction of the block-spin model [30], we argue that a consistent multi-scale graph probability should preserve its symmetries across these scales, therefore we find the unique solution that fulfills the requirement upon realistic ansätze. In this sense, we interlace the two distinct topics of finding a random graph model capable of retaining the hierarchical structure of complex networks and the one of devising a self-consistent way to apply a coarse-graining scheme on metric-less systems. We then follow up by discussing the different roles played by the parameters of the model and expound on its relation with other random graphs. Different settings of these parameters lead to different scenarios. We identify and discuss two viable applications: the first one fits in a *quenched scenario*, where the parameters are deterministic and the model can comply with an empirical network to guide its coarsening or making predictions at any level of aggregation. The possibility to merge the nodes according to non-homogeneous partitions of the original nodes allows us to account for elements at different levels of the hierarchy within the same snapshot, thereby harbouring possible multi-scale interactions. Secondly, we approach an *annealed scenario*, where the parameters are represented by random variables and the model becomes an abstract tool for generating synthetic networks that turn out to exhibit realistic properties.

In chapter 3, we revisit the initial derivation of the SIM to encompass a broader class of networks in this modelization task, specifically, directed networks. Indeed, a vast majority of real interacting systems include directionalities, and thus a generalization of the SIM capable of capturing the richer topological properties of directed networks was

required. In particular, we focus on the possibility of accounting for non-trivial patterns of network reciprocity, which refers to the tendency of vertex pairs to form mutual connections and plays a central role in the characterization of directed graphs. The chapter is organized analogously to chapter 2, it therefore includes the detailed derivation of the Directed Scale-Invariant model (in section 3.2), an application of the DSIM aimed at producing down-scaled replicas of a real network (section 3.4) and an analysis of synthetic networks drawn from our ensemble (in section 3.5).

In chapter 4, the topological properties of the SIM in its undirected formulation are sifted by adopting a mathematically rigorous perspective. Along the chapter, the annealed scenario introduced in section 2.4 is considered, and the analytical results previously obtained in a limited setting are extended to more general claims. In particular, precise statements around the expected degrees and their asymptotics are made in section 4.3.1, while in section 4.3.2 the local geometry of the graph is explored by counting the expected number of triangles and wedges in the graph ensemble.

Chapter 2

Definition of the SIM: geometry-free network renormalization

This chapter is based on the work [1] by E. Garuccio, M. Lalli, and D. Garlaschelli. We introduce the random graph model that is the focus of this thesis and which we name Scale-Invariant Model (or SIM). Its explicit derivation is reported together with two applications that illustrate how it can be used either for reproducing multi-scale snapshots of a real network (specifically the International Trade Network) or for generating scale-free networks without fine-tuning and without geometry.

2.1 Introduction

Several societal challenges, including the development of more resilient economies, the containment of infectious diseases, the security of critical infrastructures and the preservation of biodiversity, require a thorough understanding of the network structure connecting the units of the underlying complex systems [43–45]. One of the obstacles systematically encountered in the analysis and modeling of real-world networks is the simultaneous presence of structures at multiple interacting scales. Estab-

lishing a consistent representation of a graph at multiple scales is a long-standing problem whose solution would enable considerable progress in the description, modeling, and control of real-world complex systems. In the language of statistical physics, achieving a proper multiscale description of a network requires the introduction of a renormalization scheme whereby a network can be coarse-grained iteratively by partitioning nodes into 'block-nodes' either horizontally, i.e. at homogeneous levels of the hierarchy, or across hierarchical levels, thus allowing block-nodes to contain possibly very different numbers of nodes. The traditional block-renormalization scheme (whereby equally sized blocks of neighbouring nodes in a regular lattice are replaced by identical block-nodes, leading to a reduced lattice with the same geometry) is feasible for geometrically embedded networks where the coordinates of nodes naturally induce a definition of block-nodes of equal size. However, it becomes ill-defined in arbitrary graphs where node coordinates are not necessarily defined, and particularly problematic in real-world networks with broad degree distribution (which makes the neighbourhoods of nodes very heterogeneous in size and not good candidates as block-nodes) and small-world property (which limits the iterability of coarse-grainings based on shortest paths). Several renormalization schemes for complex networks have been proposed to deal with this inherent complications [28, 34, 38, 42, 46–49]. For instance, in analogy with fractal analysis, a box-covering technique defining block-nodes as certain sets of neighbouring nodes has been defined [34, 46, 50, 51]. Alternative coarse-graining schemes have been proposed based on the identification of communities [28] or motifs [52]. Another notable approach is the geometrical embedding of networks in a hidden euclidean [47] or hyperbolic [21, 38, 39] metric space, followed by the coarse-graining of nearby nodes. Hyperbolically embedded graphs have many realistic properties, including scale-free degree distributions and large clustering, that are preserved upon geometric renormalization [38]. Other methods are based on the preservation of certain spectral properties of the original network via the identification of (approximate) equivalence classes of structurally similar nodes [48, 49]. A notable recent contribution is a diffusion-based coarse-graining scheme

that detects spatio-temporal scales in heterogeneous networks via the Laplacian operator for graphs [42].

The above approaches have not yet focused on the problem of defining *the most general graph model* that remains consistent across different coarse-grainings of the network, i.e. that keeps describing the same system coherently (possibly with renormalized parameters) at all scales without requiring any further assumption e.g. regarding the embedding of nodes with geometric coordinates. Indeed, the available approaches require the existence of specific topological properties (e.g., community structure [28, 52], hyperbolicity [21, 38, 39], scale-freeness [34, 46, 51], approximate structural equivalence [48, 49], non-trivial Laplacian susceptibility [42]) and are often irreducible to the ordinary renormalization scheme defined for lattices or (random) Euclidean graphs, which on the other hand are a clear example of scale-invariant network models. Crucially, the requirement that the renormalization can act flexibly across hierarchical levels in a multiscale fashion is not explicitly enforced in any of the available methods.

Here we propose a general renormalization scheme based on a random network model that remains invariant across all scales, for *any desired (horizontal or vertical) partition of nodes into block-nodes*. In a certain 'quenched' setting, the model can guide the renormalization of generic graphs, from regular lattices to realistic complex networks with node attributes and (optionally, but not necessarily) dyadic properties such as distances and/or community structure. In a different 'annealed' setting, it can generate realistic scale-free networks spontaneously, without fine-tuning and without geometry, only as a result of the requirement of scale-invariance imposed both on graph probability and on parameters distribution, leading to one-sided α-stable random variables.

The rest of the chapter is organized as follows. In Sec. 2.2 we introduce the graph renormalization framework, identify the resulting scale-invariant network model, discuss several theoretical properties of the resulting networks, and highlight the differences with respect to other existing models. In Sec. 2.3 we consider the quenched setting, where the

model parameters are considered fixed and identifiable with empirical features, and show an application leading to a remarkably consistent model of the International Trade Network across arbitrary geographical partitions. In Sec. 2.4 we consider the annealed setting, where the model parameters are themselves regarded as random variables subject to a scale-invariance requirement, and show how this leads spontaneously to a model of scale-free networks with interesting realistic features, including finite local clustering even in absence of any notion of metric distance. In Sec. 2.5 we propose some concluding remarks. Finally, Sec. 2.6 includes technical details complementary to our analysis that have been set apart from the main text to enhance overall readability.

2.2 Graph renormalization and scale-invariant network model

Consider a binary undirected graph with N_0 'fundamental' nodes (labeled as $i_0 = 1, N_0$) and its $N_0 \times N_0$ adjacency matrix $\mathbf{A}^{(0)}$ with entries $a^{(0)}_{i_0,j_0} = 1$ if the nodes i_0 and j_0 are connected, and $a^{(0)}_{i_0,j_0} = 0$ otherwise. We do not allow for multiple edges but we do allow for self-loops, i.e., each diagonal entry can take values $a^{(0)}_{i_0,i_0} = 0, 1$.

Now, let us introduce the scheme that will allow defining the network at a coarse-grained level. Consider a non-overlapping partition $\mathbf{\Omega}_0$ of the nodes of G_{N_0} that maps the original N_0 nodes to a set of N_1 blocks, with $N_1 \leq N_0$. As will be discussed later, $\mathbf{\Omega}_0$ must be *non-overlapping* and *surjective* (i.e. each node at level 0 is assigned to *at most one* block and each block originates from *at least* one node at level 0) but is *arbitrary* on other respects. Then all the nodes within the same block are merged to form a block-node (labeled as $i_1 = 1, \ldots, N_1$), and each pair of block-nodes (or nodes at level 1) are connected if *at least one link* is present between the nodes across the two blocks, as illustrated in Fig. 1.

Therefore, the coarse-grained graph is described by the $N_1 \times N_1$ adjacency matrix $\mathbf{A}^{(1)}$ with entries $a^{(1)}_{i_1,j_1} = 1 - \prod_{i_0 \in i_1} \prod_{j_0 \in j_1} (1 - a^{(0)}_{i_0,j_0})$, where $i_0 \in i_1$ denotes that the chosen partition $\mathbf{\Omega}_0$ maps the original node i_0 onto the block-node i_1, i.e. $i_1 = \mathbf{\Omega}_0(i_0)$. Note that we do not require

$i_1 \neq j_1$, as we keep allowing for self-loops as we coarse-grain (a self-loop at a block-node represents the existence of at least one link or self-loop in the subgraph connecting the original 'internal' nodes).
In the following, we will denote the graph G_{N_ℓ} and its nodes at level ℓ as *ℓ-graph* and *ℓ-nodes*, respectively, for any considered level of aggregation ℓ.

Iterating the coarse-graining ℓ times produces a hierarchy of 'blocks of blocks', with the partition $\mathbf{\Omega}_\ell$ leading to a $(\ell+1)$-graph with $N_{\ell+1}$ $(\ell+1)$-nodes and adjacency matrix $\mathbf{A}^{(\ell+1)}$ with entries

$$a^{(\ell+1)}_{i_{\ell+1},j_{\ell+1}} = 1 - \prod_{i_\ell \in i_{\ell+1}} \prod_{j_\ell \in j_{\ell+1}} \left(1 - a^{(\ell)}_{i_\ell,j_\ell}\right) \tag{2.1}$$

where i_ℓ and j_ℓ are ℓ-nodes, while $i_{\ell+1} = \mathbf{\Omega}_\ell(i_\ell)$ and $j_{\ell+1} = \mathbf{\Omega}_\ell(j_\ell)$ are $(\ell+1)$-nodes.

The hierarchy $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$ of desired partitions can be uniquely parametrized in terms of a dendrogram, akin to the toy representation in Fig. 2.

Our first objective is the identification of a random graph model that can be renormalized under *any* partition obtained from $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$ via either a 'horizontal' (left panel of Fig. 2) or a 'multi-scale' (right panel of Fig. 2) cut of the dendrogram. Note that, since any 'multi-scale' coarse-graining is ultimately another partition of the same 0-nodes, we can equivalently produce it 'horizontally' as well, but on a certain modified hierarchy $\{\mathbf{\Omega}'_\ell\}_{\ell\geq 0}$ obtained from $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$. Therefore, requiring that the model is scale-invariant for any specified hierarchy of partitions automatically allows for multi-scale coarse-grainings as well. To enforce this requirement, we fix some $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$ and regard the initial 0-graph $\mathbf{A}^{(0)}$ not as deterministic, but as generated by a random process with some probability $P_0\big(\mathbf{A}^{(0)};\mathbf{\Theta}_0\big)$ normalized so that $\sum_{\mathbf{A}^{(0)}\in\mathcal{G}_{N_0}} P_0\big(\mathbf{A}^{(0)};\mathbf{\Theta}_0\big) = 1$, where $\mathbf{\Theta}_0$ denotes all parameters of the model (including N_0) and $\mathcal{G}_N$ denotes the set of all binary undirected graphs with N nodes. A given partition $\mathbf{\Omega}_0$ will in general map multiple 0-graphs $\{\mathbf{A}^{(0)}\}$ onto the same

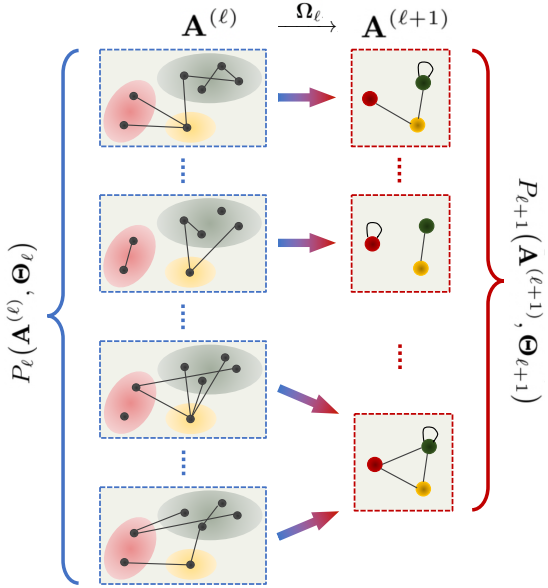

Figure 1: Schematic example of the graph coarse-graining and induced ensembles. Nodes of an ℓ-graph $\mathbf{A}^{(\ell)}$ (left) are grouped together, via a given partition $\mathbf{\Omega}_\ell$, to form the block-nodes of the coarse-grained $(\ell+1)$-graph $\mathbf{A}^{(\ell+1)}$ (right). Note that, in general, block-nodes can contain different numbers of nodes. A link between two block-nodes (or a self-loop at a single block-node) is drawn whenever a link is present between any pair of constituent nodes. Different realizations of the ℓ-graph are mapped onto realizations of the $(\ell+1)$-graph via $\mathbf{\Omega}_\ell$. Multiple realizations of the ℓ-graph may end up in the same realization of the $(\ell+1)$-graph. The scale-invariant requirement is obtained by viewing the realizations of the ℓ-graph as the outcome of a random graph generating process with probability $P_\ell\big(\mathbf{A}^{(\ell)};\mathbf{\Theta}_\ell\big)$, where $\mathbf{\Theta}_\ell$ is a set of parameters, and imposing that the induced probability $P_{\ell+1}\big(\mathbf{A}^{(\ell+1)};\mathbf{\Theta}_{\ell+1}\big)$ at the next level has the same functional form as $P_\ell\big(\mathbf{A}^{(\ell)};\mathbf{\Theta}_\ell\big)$, with renormalized parameters $\mathbf{\Theta}_{\ell+1}$.

coarse-grained 1-graph $\mathbf{A}^{(1)}$, and the notation $\{\mathbf{A}^{(0)}\} \xrightarrow{\mathbf{\Omega}_0} \mathbf{A}^{(1)}$ will denote such surjective mapping. Therefore $P_0\big(\mathbf{A}^{(0)};\mathbf{\Theta}_0\big)$ will induce a random process at the next level (see Fig. 1), generating each possible 1-graph $\mathbf{A}^{(1)}$ with probability $\sum_{\{\mathbf{A}^{(0)}\}\xrightarrow{\mathbf{\Omega}_0}\mathbf{A}^{(1)}} P_0\big(\mathbf{A}^{(0)};\mathbf{\Theta}_0\big)$, where the sum

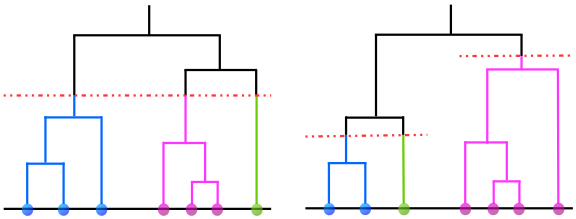

Figure 2: Horizontal vs multiscale renormalization. Left: the desired hierarchy of coarse-grainings can be represented as a dendrogram where the 0-nodes are the bottom 'leaves' and the ℓ-nodes are the 'branches' cut out by a horizontal line placed at a suitable height. Right: if the dendrogram is cut at different heights, one obtains a multiscale renormalization scheme with block-nodes defined across multiple hierarchical levels. This is ultimately another partition of the 0-nodes and is therefore readily implemented in our approach, which is designed to work for *any* partition.

runs over all 0-graphs that are projected onto $\mathbf{A}^{(1)}$ by $\mathbf{\Omega}_0$. Iterating ℓ times, we induce a process generating the ℓ-graph $\mathbf{A}^{(\ell)}$ with probability $\sum_{\{\mathbf{A}^{(0)}\}\xrightarrow{\mathbf{\Omega}_{\ell-1}\cdots\mathbf{\Omega}_0}\mathbf{A}^{(\ell)}} P_0\big(\mathbf{A}^{(0)};\mathbf{\Theta}_0\big)$, where $\mathbf{\Omega}_{\ell-1}\cdots\mathbf{\Omega}_0$ is the composition of the ℓ partitions $\{\mathbf{\Omega}_m\}_{m=0}^{\ell-1}$, which is ultimately a partition of the 0-nodes.

We now enforce a scale-invariant random graph model that, for any level ℓ, can generate the possible ℓ-graphs in two equivalent ways: either *hierarchically*, i.e. by first generating the 0-graphs with probability $P_0\big(\mathbf{A}^{(0)};\mathbf{\Theta}_0\big)$ and then coarse-graining them ℓ times via the partitions $\{\mathbf{\Omega}_k\}_{k=0}^{\ell-1}$, or *directly*, i.e. with a certain probability $P_\ell\big(\mathbf{A}^{(\ell)};\mathbf{\Theta}_\ell\big)$ that depends on ℓ only through a set $\mathbf{\Theta}_\ell$ of renormalized parameters that can be obtained, in their turn, hierarchically from $\mathbf{\Theta}_0$ (by means of renormalization rules still to be defined) or directly at level ℓ. This scale-invariance requirement means that, apart from the different dimensionality of their domains, $P_0(\cdot,\cdot)$ and $P_\ell(\cdot,\cdot)$ should have the same functional form, which

we denote by $P(\cdot,\cdot)$, and be such that, for any pair ℓ, m (with $\ell > m$),

$$P\big(\mathbf{A}^{(\ell)};\boldsymbol{\Theta}_\ell\big) = \sum_{\{\mathbf{A}^{(m)}\}\xrightarrow{\boldsymbol{\Omega}_{\ell-1}\cdots\boldsymbol{\Omega}_m}\mathbf{A}^{(\ell)}} P\big(\mathbf{A}^{(m)};\boldsymbol{\Theta}_m\big) \tag{2.2}$$

where the renormalized parameters $\boldsymbol{\Theta}_\ell$ are obtained only from $\boldsymbol{\Theta}_m$, given $\boldsymbol{\Omega}_{\ell-1}\cdots\boldsymbol{\Omega}_m$.

Let us look for the general solution in the case of random graphs with independent links. In this case, $P\big(\mathbf{A}^{(\ell)};\boldsymbol{\Theta}_\ell\big)$ factorizes as

$$\prod_{i_\ell=1}^{N_\ell}\prod_{j_\ell=1}^{i_\ell}\big[p_{i_\ell,j_\ell}\big(\boldsymbol{\Theta}_\ell\big)\big]^{a^{(\ell)}_{i_\ell,j_\ell}}\big[1-p_{i_\ell,j_\ell}\big(\boldsymbol{\Theta}_\ell\big)\big]^{1-a^{(\ell)}_{i_\ell,j_\ell}}, \tag{2.3}$$

where $p_{i_\ell,j_\ell}\big(\boldsymbol{\Theta}_\ell\big)$ is the probability that two ℓ-nodes i_ℓ and j_ℓ are linked. In this case it is natural to require that $\boldsymbol{\Theta}_\ell$ contains (besides N_ℓ) an overall constant δ_ℓ (which will set the global link density), a set of additive node-specific 'fitness' parameters $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ (which will distribute the total number of links heterogeneously among nodes), and an (optional) set of dyadic (pair-specific) parameters $\{d_{i_\ell,j_\ell}\}_{i_\ell,j_\ell=1}^{N_\ell}$ (which include, when $i_\ell = j_\ell$, the 'self-interaction' of a node with itself). We can therefore use the notation $p_{i_\ell,j_\ell}\big(\boldsymbol{\Theta}_\ell\big) = p_{i_\ell,j_\ell}(\delta_\ell)$ where we keep only δ_ℓ explicit in the argument of p_{i_ℓ,j_ℓ}, because the dependence on the other variables x_{i_ℓ}, x_{j_ℓ}, d_{i_ℓ,j_ℓ} is already implicitly denoted by the subscripts i_ℓ, j_ℓ (indeed, p_{i_ℓ,j_ℓ} depends on i_ℓ and j_ℓ only through x_{i_ℓ}, x_{j_ℓ}, d_{i_ℓ,j_ℓ}).

As we prove in Sec. 2.6.1, if the graph probability factorizes as in Eq. (2.3) and the fitness x is assumed to be additive upon coarse-graining of nodes, then there is a unique solution to Eq. (2.2), given by the connection probability

$$p_{i_\ell,j_\ell}(\delta) = \begin{cases} 1-e^{-\delta\, x_{i_\ell}\, x_{j_\ell}\, f(d_{i_\ell,j_\ell})} & \text{if} \quad i_\ell \neq j_\ell \\ 1-e^{-\frac{\delta}{2}\, x_{i_\ell}^2\, f(d_{i_\ell,i_\ell})} & \text{if} \quad i_\ell = j_\ell \end{cases} \tag{2.4}$$

where $\delta > 0$, $x_{i_\ell} \geq 0$ for all i_ℓ, f is an arbitrary positive function and the

following renormalization rules apply:

$$\delta_{\ell+1} \equiv \delta_\ell \equiv \delta, \tag{2.5}$$

$$x_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} x_{i_\ell}, \tag{2.6}$$

$$f\big(d_{i_{\ell+1},j_{\ell+1}}\big) \equiv \frac{\sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in j_{\ell+1}} x_{i_\ell} x_{j_\ell} f\big(d_{i_\ell,j_\ell}\big)}{\sum_{i_\ell \in i_{\ell+1}} x_{i_\ell} \sum_{j_\ell \in j_{\ell+1}} x_{j_\ell}} \tag{2.7}$$

which means δ is scale-invariant, x is node-additive and f renormalizes as a specific fitness-weighted average.
If the fitness is assumed to have a different renormalization rule (e.g. multiplicative rather than additive), then a corresponding modified solution is obtained (e.g. with x replaced by $\log x$). So, up to a redefinition of the fitness, the solution above is general.
Note that Eq. (2.7) applies also to the 'diagonal' terms with $i_{\ell+1} = j_{\ell+1}$, in which case it represents the renormalized self-interaction of node $i_{\ell+1}$ with itself, determining the probability of the presence of the corresponding self-loop.

Equations (2.4)-(2.7) are our key result. One of their consequences is that, while the dependence of the connection probability $p_{i_\ell,j_\ell}(\delta)$ on the dyadic factor d_{i_ℓ,j_ℓ} can be switched off entirely without destroying the scale-invariant properties of the model (for example by taking f to be a constant function, whereby Eq. (2.7) is automatically fulfilled), the dependence on the node-specific factors $x_{i_\ell}\ x_{j_\ell}$ cannot be switched off unless the model is made deterministic by formally requiring that $f(d_{i_\ell,j_\ell})$ takes only the two values $f = 0$ (implying $p_{i_\ell,j_\ell}(\delta) = 0$) or $f = +\infty$ (implying $p_{i_\ell,j_\ell}(\delta) = 1$). We consider examples of both situations below. For now, we would like to remark that *the dependence on dyadic factors (including geometric distances) is optional, while that on node-specific factors is necessary.* This is a general result following only from the enforcement of scale-invariance. More specific results are discussed below.

2.2.1 Scale-invariance of graph probability and partition function

Equations (2.4)-(2.7) have been derived using the scale-invariant requirement imposed in Eq. (2.2), under the assumption of edge independence stated in Eq. (2.3). Indeed, inserting Eq. (2.4) into Eq. (2.3) yields the explicit expression for the (scale-invariant) graph probability:

$$
\begin{aligned}
P\big(\mathbf{A}^{(\ell)};\delta\big) &= \prod_{i_\ell=1}^{N_\ell}\prod_{j_\ell=1}^{i_\ell}\big[1-p_{i_\ell,j_\ell}(\delta)\big]\left[\frac{p_{i_\ell,j_\ell}(\delta)}{1-p_{i_\ell,j_\ell}(\delta)}\right]^{a^{(\ell)}_{i_\ell,j_\ell}} \qquad (2.8)\\
&= \prod_{i_\ell=1}^{N_\ell}\frac{\big[e^{\frac{\delta}{2}x_{i_\ell}^2 f(d_{i_\ell,i_\ell})}-1\big]^{a^{(\ell)}_{i_\ell,i_\ell}}}{e^{\frac{\delta}{2}x_{i_\ell}^2 f(d_{i_\ell,i_\ell})}}\prod_{j_\ell=1}^{i_\ell-1}\frac{\big[e^{\delta x_{i_\ell}x_{j_\ell} f(d_{i_\ell,j_\ell})}-1\big]^{a^{(\ell)}_{i_\ell,j_\ell}}}{e^{\delta x_{i_\ell}x_{j_\ell} f(d_{i_\ell,j_\ell})}}\\
&= \frac{\prod_{i_\ell=1}^{N_\ell}\big[e^{\frac{\delta}{2}x_{i_\ell}^2 f(d_{i_\ell,i_\ell})}-1\big]^{a^{(\ell)}_{i_\ell,i_\ell}}\prod_{j_\ell=1}^{i_\ell-1}\big[e^{\delta x_{i_\ell}x_{j_\ell} f(d_{i_\ell,j_\ell})}-1\big]^{a^{(\ell)}_{i_\ell,j_\ell}}}{Q^{-1}(\delta)}
\end{aligned}
$$

where we have introduced the quantity

$$
\begin{aligned}
Q(\delta) &\equiv \prod_{i_\ell=1}^{N_\ell} e^{-\frac{\delta}{2}x_{i_\ell}^2 f(d_{i_\ell,i_\ell})}\prod_{j_\ell=1}^{i_\ell-1} e^{-\delta x_{i_\ell}x_{j_\ell} f(d_{i_\ell,j_\ell})}\\
&= \prod_{i_\ell=1}^{N_\ell}\prod_{j_\ell=1}^{N_\ell} e^{-\frac{\delta}{2}x_{i_\ell}x_{j_\ell} f(d_{i_\ell,j_\ell})}\\
&= e^{-\frac{\delta}{2}\sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1}^{N_\ell} x_{i_\ell}x_{j_\ell} f(d_{i_\ell,j_\ell})}\\
&= e^{-\frac{\delta}{2}x_{i_\infty}^2 f(d_{i_\infty,i_\infty})}\\
&= 1-p_{i_\infty,i_\infty}(\delta), \qquad (2.9)
\end{aligned}
$$

with

$$
\begin{aligned}
x_{i_\infty} &\equiv \sum_{i_\ell=1}^{N_\ell} x_{i_\ell}, \qquad (2.10)\\
f(d_{i_\infty,i_\infty}) &\equiv \sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1}^{N_\ell}\frac{x_{i_\ell}x_{j_\ell} f\big(d_{i_\ell,j_\ell}\big)}{x_{i_\infty}^2} \qquad (2.11)
\end{aligned}
$$

representing the total fitness of all nodes and the fitness-weighted average of f over all pairs of nodes, respectively. Our notation above is meant to suggest that x_{i_∞} and $f(d_{i_\infty,i_\infty})$ can be interpreted as the fitness and self-interaction of a *supernode* i_∞ representing the only coarse-grained node remaining after applying an infinite sequence of partitions, or equivalently after applying the trivial partition Ω_∞ that places all nodes in the same supernode i_∞ (such that $N_\infty = 1$). Indeed, when applied to such supernode, Eqs. (2.6) and (2.7) produce exactly the values x_{i_∞} and $f(d_{i_\infty,i_\infty})$ defined in Eqs. (2.10) and (2.11), respectively. These quantities are clearly scale-invariant, in that they can be calculated from the values taken by x and $f(d)$ at any hierarchical level ℓ. Therefore $Q(\delta)$ is a constant term that depends neither on the realized ℓ-graph $\mathbf{A}^{(\ell)}$ or, owing to Eq. (2.7), on the hierarchical level ℓ being considered. Consequently, as desired, $P\big(\mathbf{A}^{(\ell)};\delta\big)$ depends on ℓ only through the parameters $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ and $\{d_{i_\ell,j_\ell}\}_{i_\ell,j_\ell=1}^{N_\ell}$, which renormalize as stated in Eqs. (2.6) and (2.7).
Note that since $p_{i_\infty,i_\infty}(\delta)$ in Eq. (2.9) represents the probability of a self-loop at the supernode i_∞, i.e. the probability of having at least one link in the graph at any hierarchical level, then $Q(\delta) = 1 - p_{i_\infty,i_\infty}(\delta)$ formally represents the probability of having zero links in the network.

We can recast the above result in a way that has an explicit connection with the real-space renormalization framework in statistical physics [31, 32]. To do so, we rewrite the graph probability in Eq. (2.8) in terms of an effective Hamiltonian $\mathscr{H}_{\text{eff}}^{(\ell)}$ and a partition function $\mathscr{Z}^{(\ell)}$:

$$P(\mathbf{A}^{(\ell)};\delta) = \frac{e^{-\mathscr{H}_{\text{eff}}^{(\ell)}(\mathbf{A}^{(\ell)},\delta)}}{\mathscr{Z}^{(\ell)}(N_\ell,\delta)} \tag{2.12}$$

where we have defined:

$$\begin{aligned}
\mathscr{H}_{\text{eff}}^{(\ell)}(\mathbf{A}^{(\ell)},\delta) &= -\sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1}^{i_\ell} a_{i_\ell,j_\ell}\log\left[\frac{p_{i_\ell,j_\ell}(\delta)}{1-p_{i_\ell,j_\ell}(\delta)}\right] \qquad (2.13)\\
&= -\sum_{i_\ell=1}^{N_\ell}\Big[a_{i_\ell,i_\ell}\log\left(e^{\frac{\delta}{2}x_{i_\ell}x_{i_\ell}f(d_{i_\ell,i_\ell})}-1\right)\\
&\quad + \sum_{j_\ell=1}^{i_\ell-1} a_{i_\ell,j_\ell}\log\left(e^{\delta x_{i_\ell}x_{j_\ell}f(d_{i_\ell,j_\ell})}-1\right)\Big]
\end{aligned}$$

and

$$\mathcal{Z}^{(\ell)}(N_\ell, \delta) \equiv \sum_{\mathbf{A}^{(\ell)}} e^{-\mathcal{H}_{\text{eff}}^{(\ell)}(\mathbf{A}^{(\ell)}, \delta)}. \tag{2.14}$$

Note that, while Eq. (2.12) is formally identical to the expression for graph probabilities in the Exponential Random Graphs approach [22–24, 53, 54], in our case it does not exhibit a *sufficient statistic*, i.e. the Hamiltonian in Eq. (2.13) cannot be written as a simpler function of graph properties and, to be evaluated, requires the knowledge of the entire adjacency matrix of the graph. A straightforward calculation yields

$$\begin{aligned}
\mathcal{Z}^{(\ell)}(N_\ell, \delta) &= \sum_{\mathbf{A}^{(\ell)}} \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} \left[\frac{p_{i_\ell,j_\ell}(\delta)}{1-p_{i_\ell,j_\ell}(\delta)}\right]^{a_{i_\ell,j_\ell}^{(\ell)}} \\
&= \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} \sum_{a_{i_\ell,j_\ell}=0}^{1} \left[\frac{p_{i_\ell,j_\ell}(\delta)}{1-p_{i_\ell,j_\ell}(\delta)}\right]^{a_{i_\ell,j_\ell}^{(\ell)}} \\
&= \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} \frac{1}{1-p_{i_\ell,j_\ell}(\delta)} \\
&= \frac{1}{1-p_{i_\infty,i_\infty}(\delta)} \\
&= Q^{-1}(\delta).
\end{aligned} \tag{2.15}$$

As noticed above, Q only depends on δ, which is invariant under renormalization. We can therefore drop the superscript and denote the partition function as $\mathcal{Z}(\delta)$.

Indeed, since the effective Hamiltonian $\mathcal{H}_{\text{eff}}^{(\ell)}$ has the same form given in Eq. (2.13), independently of the resolution level ℓ, recalculating $\mathcal{Z}^{(m)}(N_m, \delta)$ from Eq. (2.14) for any other coarse-graining level $m \neq \ell$ and number N_m of nodes would return exactly the same value $Q^{-1}(\delta)$:

$$\mathcal{Z}^{(m)}(\delta) = \mathcal{Z}^{(\ell)}(\delta) \equiv \mathcal{Z}^{(\ell)}(\delta) = Q(\delta), \quad \forall\, \ell, m. \tag{2.16}$$

This means that, akin to Kadanoff's construction [31], *the partition function is invariant along the coarse-graining flow*. This property follows crucially from the functional form of the connection probability $p_{i_\ell,j_\ell}(\delta)$ given in Eq. (2.4): any other functional form, including those considered in

ERGs [22–24, 53, 54], would in general not lead to an invariant partition function. On the other hand, precisely because of this scale-invariance, in our model here the effective Hamiltonian for a realization (say, $\mathbf{A}^{(m)}$) of the graph at a coarse-grained level m can be evaluated exactly without knowing the microscopic details of any finer-grained version $\mathbf{A}^{(\ell)}$ (with $\ell < m$) of the same realized graph $\mathbf{A}^{(m)}$. This is not possible in ERGs, and suggests that the topology of $\mathbf{A}^{(m)}$ represents in some sense a sort of sufficient statistic for the model since the probability of $\mathbf{A}^{(m)}$ can be estimated without explicitly summing over the compatible topologies of any finer version of the same network, i.e. over the realizations $\mathbf{A}^{(\ell)}$ such that $\mathbf{A}^{(\ell)} \xrightarrow{\Omega_{m-1}...\Omega_\ell} \mathbf{A}^{(m)}$ for some sequence $\Omega_{m-1}...\Omega_\ell$ of partitions.

Node-specific fitness The connection probability p_{i_ℓ,j_ℓ} increases as x_{i_ℓ} and/or x_{j_ℓ} increase. Therefore, as in the Fitness Model (FM) [20] and in the inhomogeneous random graph model (IRGM) [55], x_{i_ℓ} can be viewed as a hidden variable or 'fitness' that characterizes the intrinsic tendency of the ℓ-node i_ℓ to form connections. Here, the fitness is defined across multiple hierarchical levels and renormalizability ensures that it is also an additive quantity summing up to the value in Eq. (2.6) when ℓ-nodes are merged onto an $(\ell+1)$-node. This ensures that the total fitness x_{i_∞} defined in Eq. (2.10) is preserved by the renormalization. For instance, if one starts with $x_{i_0} = 1$ for all i_0, then x_{i_ℓ} will simply count how many 0-nodes are found within the ℓ-node i_ℓ, and $x_{i_\infty} = N_0$. More interesting outcomes are obtained by using heterogeneous distributions of the fitness, as we illustrate later.

We will consider both the 'quenched' case where the fitness is fixed and possibly identified with some empirical quantity (thereby allowing for the renormalization of real-world networks irrespective of their scale-free behaviour), and then an opposite 'annealed' scenario that spontaneously leads to scale-invariant *and* scale-free networks with a density-dependent cut-off (thereby providing a generic mechanism for the emergence of scale-free networks from scale-invariance, *without geometry*).

Dyadic properties The quantity d_{i_ℓ,j_ℓ} is a dyadic factor (such as distance, similarity, co-membership in the same community, etc.) associated with the node pair (i_ℓ, j_ℓ). Although we are free to do otherwise, we may regard d_{i_ℓ,j_ℓ} as a distance, in which case it may make sense to assume that f is a decreasing function, ensuring that more distant nodes are less likely to be connected.

It is easy to realize that, if d_{i_0,j_0} is an *ultrametric distance*, i.e., such that the 'stronger' triangle inequality $d_{i_0,j_0} \leq \max\{d_{i_0,k_0}, d_{j_0,k_0}\}$ holds for every triple i_0, j_0, k_0 of 0-nodes [56] and is consistent with the hierarchy of coarse-grainings (i.e., such that all distances can be represented as the heights of the branching points of the dendrogram shown in Fig. 2), then $d_{i_\ell,j_\ell} = d_{i_0,j_0}$ and hence $f(d_{i_\ell,j_\ell}) = f(d_{i_0,j_0})$ whenever the 0-nodes i_0 and i_0 map onto the ℓ-nodes i_ℓ and j_ℓ respectively, that is, whenever $i_\ell = \mathbf{\Omega}_{\ell-1} \cdots \mathbf{\Omega}_0(i_0)$ and $j_\ell = \mathbf{\Omega}_{\ell-1} \cdots \mathbf{\Omega}_0(j_0)$. In such a case, Eq. (2.7) reduces to $f\big(d_{i_{\ell+1},j_{\ell+1}}\big) = f\big(d_{i_\ell,j_\ell}\big)$ with $i_{\ell+1} = \mathbf{\Omega}_\ell(i_\ell)$ and $j_{\ell+1} = \mathbf{\Omega}_\ell(j_\ell)$, showing that *if the distances among the* 0*-nodes are ultrametric on the dendrogram induced by the hierarchy of partitions, they decouple from the hidden variables and remain invariant* across the entire coarse-graining process, just like the global parameter δ. Reversing the point of view, we may equivalently say that, *given an ultrametric distance among the* 0*-nodes, any hierarchy of partitions induced by the associated dendrogram keeps the distances scale-invariant*. In weaker form, this also means that one may use d_{i_0,j_0} to specify the dendrogram parametrizing the desired hierarchy of partitions that will keep the distances scale-invariant. The hierarchy may coincide with e.g. a nested community structure that one may want to impose. In any case, we stress that, although ultrametricity is an attractive property (especially in the annealed scenario that we introduce later or in the one where directed graphs are approached - as detailed in Chap. 3), we do not require it as a necessary condition in general.

2.2.2 Recovering the lattice case

We may now discuss a simple but important extreme case, where the graph is constructed only as a function of distance, and our approach

reduces to the traditional scheme for renormalizing regular lattices.
For instance, assume that the 0-nodes have coordinates at the sites of a 2-dimensional grid with lattice spacing τ_0 and that d_{i_0,j_0} is the Euclidean distance between these coordinates. If we set $f \equiv +\infty$ if $d_{i_\ell,j_\ell} \leq 2^\ell \tau_0$ and $f \equiv 0$ otherwise, then the 0-graph will be deterministically the grid itself and the ℓ-graph will be the usual renormalized lattice with spacing $\tau_\ell = 2^\ell \tau_0$ obtained through an appropriate partition $\mathbf{\Omega}_{\ell-1}$ that maps each square block of 4 nearest $(\ell-1)$-nodes onto a single ℓ-node sitting at the center of the square. In this case, each vertical line of the dendrogram of hierarchical partitions branches regularly into 4 'daughter' lines and $\tau_\ell = 2^\ell \tau_0$ is the height of the branching points splitting $(\ell+1)$-nodes into ℓ-nodes. The renormalized distances d_{i_ℓ,j_ℓ} can be mapped exactly to this dendrogram, thereby retrieving the standard lattice coarse-graining scheme as a special case of our approach.

2.2.3 Relation to other network models

In the opposite extreme, the dependence on the dyadic factors can be switched off. In particular, if we set $f \equiv 1$, Eq. (2.4) reduces to

$$p_{i_\ell,j_\ell}(\delta) = \begin{cases} 1 - e^{-\delta\, x_{i_\ell}\, x_{j_\ell}} & \text{if} \quad i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2}\, x_{i_\ell}^2} & \text{if} \quad i_\ell = j_\ell \end{cases} \tag{2.17}$$

Depending on whether the fitness is considered to be quenched or annealed (a distinction that we will study in detail below), this model can also be viewed as a unique specification of the FM [20] or of the IRGM [55, 57], respectively. In particular, the specific form of connection probability in Eq. (2.17) has been studied in previous works [58, 59]. However, both our quenched (deterministic fitness) and annealed (random fitness) approaches yield quite different scenarios with respect to those previous studies, where the considered fitness (under the different names of 'capacity' [58] or 'sociability' [59]) are assumed to be random variables drawn from distributions with finite mean, whereas our fitness is either deterministic (and taken to be some fixed value measured from real data) or random but with infinite mean (the infinite-mean case being irreducible to the finite-mean one).

In the 'sparse' and 'bounded' case, i.e. for $\delta \ll x_{\max}^{-2}$ and $x_{\max} < +\infty$ where $x_{\max}$ is the maximum realized (in the quenched case) or expected (in the annealed case) value of the fitness, Eq. (2.17) reduces to $p_{i_\ell,j_\ell}(\delta) \approx \delta x_{i_\ell} x_{j_\ell}$, which includes the Chung-Lu [60] or 'sparse' Configuration Model (CM) ($p_{i,j} \approx \delta x_i x_j$ with $x_i = k_i$ and $\delta = (2L)^{-1}$, where k_i is the degree of node i and L is the total number of links). Indeed, it is possible to prove the asymptotic equivalence (or a weaker form of asymptotic contiguity) of these models under certain assumptions on the expected network sparsity and on the moments of the distribution of the hidden variables [57]. Similarly, in the same limit Eq. (2.4) reduces to $p_{i_\ell,j_\ell}(\delta) \approx \delta x_{i_\ell} x_{j_\ell} f(d_{i_\ell,j_\ell})$, which includes the sparse degree-corrected Stochastic Block-Model (dcSBM) [61] ($p_{i,j} \approx \delta x_i x_j B_{i,j}$ where $\mathbf{B}$ is a block matrix) and the Hyperbolic Model (HM) [21, 39] (where x_i is a 'hidden degree' related to the radial coordinate of node i and $d_{i,j}$ to the angular separation between nodes i and j). The CM, dcSBM, and HM are among the most popular network models and find diverse applications including community detection [62], pattern recognition [54] and network reconstruction [63]. They are instances of maximum-entropy random graph ensembles [54], which are obtained by maximizing the entropy under constraints on certain expected structural properties [23, 24, 53, 64]. To generate scale-free networks with power-law degree distribution, the CM and the dcSBM are usually constructed by drawing the fitness from a power-law distribution with *the same exponent* [65] of the target degree distribution (equivalently, in the HM the fitness distribution is realized via suitably sprinkling points in hyperbolic space). In the sparse regime, the fitness distribution and the degree distribution are therefore identical. In the dense regime, the degree distribution has still the same power-law regime as the fitness distribution, but it additionally features a size-dependent upper cut-off, corresponding to the largest degrees approaching their maximum value [65].

However, Eq. (2.17) is in general *not* equivalent to the aforementioned models. Firstly, in the quenched case, even if we start from a sufficiently sparse 0-graph for which these models are consistent with Eq. (2.4), successive coarse-grainings will unavoidably increase $x_{\max}$ and bring the

network to the dense regime where the CM, dcSBM and HM are described by their 'full' probability $p_{i,j} = \delta x_i x_j B_{i,j}/(1+\delta x_i x_j B_{i,j})$ [21, 53, 65]. Since the difference between the values of $p_{i,j}$ in Eq. (2.4) and the corresponding ones in the dcSBM or HM, and similarly between those in Eq. (2.17) and the corresponding ones in the CM, are now of finite order, these models are no longer equivalent [57] in the dense regime. Secondly, in the annealed case, we will find that all moments of the distribution of the hidden variables in our approach necessarily diverge. Remarkably, this property breaks the equivalence of the different models even in the sparse case, as the conditions on the moments of the fitness distribution required for equivalence and contiguity [57] no longer hold. As we show later, notable and useful consequences of this inequivalence are a non-linear dependence of the degree on the fitness (hence *different* exponents of the fitness and degree distributions) and a non-vanishing local clustering coefficient even in the sparse regime. The above considerations indicate that the multiscale model is in general *not* equivalent to the CM and the dcSBM, which are not renormalizable. Similar considerations apply to the traditional (non-degree-corrected) SBM [66] (for which $p_{i,j} = B_{i,j}$), to the Erdős-Rényi (ER) model [7] (for which $p_{i,j} = p$ for all i, j) and to growing network models based on preferential attachment (PA) [13] (where nodes enter sequentially into the network and the time at which a node enters determines its expected topological properties). There is no straightforward way to coarse-grain these models by defining block-nodes (possibly across different hierarchical levels) that respect the different expected properties of the nodes they contain.

These considerations indicate that scale-invariant networks are consistent with a unique specification of the FM, possibly enhanced by dyadic factors, while they are incompatible with the CM, (dc)SBM, ER, and PA models. As for the HM, while the renormalization scheme proposed in [38] does address the consistency of the graph probability across scales, the connection probability remains congruent with the hidden metric space (i.e. retains the same functional form across coarse-grainings) only if the density of links is kept sufficiently low, such that the contribution of multi-edges terms can be neglected. To maintain this condition enforced

across multiple agglomeration levels, the HM requires a progressive pruning of links, making the scheme different from the one considered here.

2.2.4 Scale-free versus scale-invariant networks

The above discussion inspires a few considerations around the distinction between *scale-free* networks (*i.e.* graphs with power-law tails in the degree distribution, as usually appearing in the CM, dcSBM, and PA models) and *scale-invariant* networks (i.e. graphs designed to be renormalizable as defined here).

The early renormalization approaches reminiscent of fractal analysis [34, 46, 51] relied on the idea that real-world networks can be interpreted as scale-invariant, precisely because of their scale-free property. However, the degrees, even when power-law distributed, cannot be renormalized exactly because they are neither preserved nor additively transformed upon renormalization. The non-renormalizability of the CM, (dc)SBM, ER and PA models originates precisely from the fact that their defining quantities are the node degrees. *Unlike fractals, the self-similarity of scale-free networks applies to a topological property (the degree), not to a metric one.* The absence of an embedding metric space, which would provide an 'ambient' dimensionality to harbour fractality in the first place, is also the reason why arbitrary networks cannot be easily renormalized using metric coordinates.

In general, scale-invariance as intended here is not due to the scale-free property, but to the compatibility with Eq. (2.4). As mentioned above, in the quenched case, and only if δ is small enough and the fitness is not too broadly distributed (so that $x_{\max} < +\infty$), there may be a sparse regime where Eq. (2.17) reduces to $p_{i_\ell,j_\ell} \approx \delta x_{i_\ell}, x_{j_\ell}$ with $k_{i_\ell} = x_{i_\ell}$, so that degrees are approximately additive. However, even in this case, degrees are rigorously additive only if each $(\ell+1)$-node is obtained as a set of ℓ-nodes that have no mutual connection among themselves. This prescription is opposite to the more natural scheme of merging nodes that are tightly connected, e.g. because they are in the same community [28] or motif [52]. If mutually connected nodes are mapped onto the same

block-node, the degree of the latter is strictly smaller than the sum of the degrees of the original nodes. We may say that the coarse-graining of a network is often designed in such a way that the additivity of degrees is maximally violated. Along the coarse-graining flow, the sparse regime progressively vanishes into the dense one, eventually breaking the approximate additivity of degrees and producing an unavoidable upper cut-off in the degree distribution. We will show that this aspect mirrors a suggestive outcome of the annealed case, where the proportionality between fitness and degree does not hold, even in the sparse regime. In that case, scale-invariance naturally leads to intrinsically non-additive degrees throughout the entire spectrum of network density.

2.3 Quenched fitness

In the quenched case, the fitness of each 0-node i_0 is assigned a fixed value x_{i_0} and the only randomness resides in the construction of the random graph ensemble, given the fitness values. For instance, when modeling real-world networks, the observed nodes can be identified with the 0-nodes and x_{i_0} can be taken to be the value of some measurable additive empirical quantity attached to the 0-node i_0. Then, after choosing a hierarchy of partitions and consistently with Eq. (2.6), the fitness $x_{i_{\ell+1}}$ of each $(\ell+1)$-node $i_{\ell+1}$ (with $\ell > 0$) is calculated iteratively by summing the fitness of all the ℓ-nodes mapped onto $i_{\ell+1}$. For each pair (i_0, j_0) of 0-nodes, a distance d_{i_0,j_0} may also be specified (and possibly measured from empirical data as well) and used to determine $f(d_{i_0,j_0})$. Consistently, the quantity $f(d_{i_{\ell+1},j_{\ell+1}})$ between each pair $(i_{\ell+1}, j_{\ell+1})$ of $(\ell+1)$-nodes is calculated via Eq. (2.7). Together, fitness and distance determine the probability (2.4) of connection between nodes at all scales. Once f is specified, the only free parameter is δ, controlling the overall density of the random network. When considering real-world networks for which fitness and distance can be measured from empirical data separately from the network structure, we may use the quenched model in order to check whether Eq. (2.4) reproduces the observed topological properties of the

0-graph itself and, if this is the case, to provide a testable multi-scale model of the renormalized network at any higher level of aggregation.

2.3.1 The International Trade Network

To illustrate the procedure, we consider the empirical International Trade Network (ITN), using the BACI-Comtrade dataset [67] which reports the international trade flows (imports and exports) between all pairs of world countries. We show the results for the year 2011, although we have obtained similar results for the other years available in the database.
We select this particular network because previous research has clarified that the topology of the ITN is strongly dictated by the GDP of countries [68–71]. Moreover, the economics literature has extensively shown that both GDP and geographical distance are key determinants of international trade, leading to the so-called 'Gravity Model' of trade [72, 73]. The additivity of the GDP (i.e. the aggregate GDP of two countries is the sum of their GDPs) makes the ITN a perfect candidate for our analysis and allows us to introduce a novel renormalization scheme for this important economic network across arbitrary levels of geographical aggregation. In particular, our aim is twofold. On the one hand, we want to introduce a multiscale model of the ITN derived from first principles, i.e. using the unique combination of GDP and geographical distances dictated by Eq. (2.4), rather than arbitrary or data-driven combinations. On the other hand, we want to check whether the empirical topology of the ITN is consistent with the multiscale model not only at the country level at which it is usually studied (here, the 0-graph) but also across different hierarchical levels using the renormalization rules in Eqs. (2.6) and (2.7). Further information on the dataset and on how we built on it can be found in the complementary Sec. 2.6.2, while here we detail the implementation of our method.

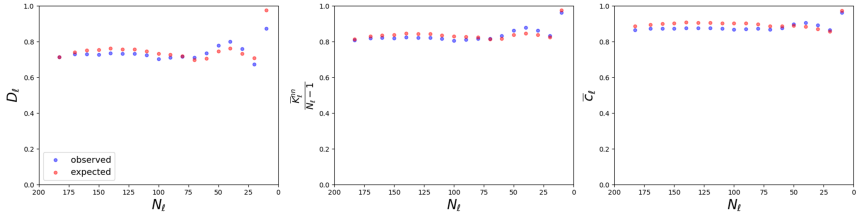

Figure 3: Prediction of global topological properties of the renormalized ITN across the full spectrum of geographical aggregation using the multiscale model. The panels show the agreement between the empirical and expected values of the link density D_ℓ including possible self-loops (left), node-averaged rescaled average nearest neighbour degree $\bar{k}_\ell^{nn}/(N_\ell - 1)$ (middle) and node-averaged local clustering coefficient $\bar{c}_\ell$ (right) as functions of the number N_ℓ of ℓ-countries, for all the 18 hierarchical levels considered ($\ell = 0, 17$).

Definition of the multiscale model

Let us identify each 0-node i_0 with a specific country for which there are GDP data available from the World Bank [74] in the considered year. This results in $N_0 = 183$ 0-nodes. Then, we set the fitness x_{i_0} of each 0-node equal to the empirical value of the GDP: $x_{i_0} = \mathrm{GDP}_{i_0}$, $i_0 = 1, N_0$.
For each pair (i_0, j_0) of countries, we set the distance d_{i_0,j_0} equal to the empirical geographical distance between the corresponding countries, using the BACI-CEPII GeoDist [75, 76] database that reports population-averaged inter-country distances.

Next, we use these distances to induce a hierarchy of partitions $\{\mathbf{\Omega}_\ell\}_{\ell \geq 0}$ that define the possible coarse-grainings of the ITN. Technically, this is done by merging geographically close countries into 'block-countries' following a single-linkage hierarchical clustering algorithm based on the GeoDist distances $\{d_{i_0,j_0}\}_{i_0,j_0=1}^{N_0}$. The output of this algorithm is a dendrogram (shown in Sec. 2.6.2) like the one in Fig. 2, where the leaves are the original countries (0-nodes), the branching points are the block-countries, and the height of each branching point represents the ultrametric geographical distance between pairs of countries across the corresponding two branches. Note that the ultrametric distances

$\{d^{<}_{i_0,j_0}\}^{N_0}_{i_0,j_0=1}$ obtained via the single-linkage clustering are known as *subdominant ultrametric* distances and ensure the smallest possible distortion among all possible ultrametric distances approximating the original metric distances 'from below' [56].

Cutting the dendrogram at a fixed height h_ℓ defines the hierarchical level ℓ and identifies a unique partition $\mathbf{\Omega}_\ell$ of countries into a certain number N_ℓ of 'ℓ-countries'. This partition can be regarded as a multiscale aggregation of countries into groups of varying size, following from actual geographical closeness rather than pre-imposed regional or political criteria. Cutting the dendogram at multiple heights $\{h_\ell\}_{\ell\geq 0}$ (with $h_0 = 0$) identifies a set $\{\ell\}$ of hierarchical levels, a geography-induced hierarchy $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$ of partitions, and a corresponding sequence $\{N_\ell\}_{\ell\geq 0}$ of numbers of block-countries. We considered 18 hierarchical levels (from $\ell = 0$ to $\ell = 17$), such that the number of block-countries is $N_\ell = 183$ for $\ell = 0$ and $N_\ell = 180 - 10\ell$ for $\ell = 1, 17$. For each of these levels, the additivity of GDP ensures that Eq. (2.6) holds as a definition for the empirical aggregate GDP of block-countries:

$$\mathrm{GDP}_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} \mathrm{GDP}_{i_\ell}. \tag{2.18}$$

We then fix the function f in Eq. (2.4) as $f(d) = d^{-1}$, so that the renormalized geographical distances equal

$$d^{-1}_{i_{\ell+1},j_{\ell+1}} \equiv \frac{\sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in j_{\ell+1}} \mathrm{GDP}_{i_\ell}\, \mathrm{GDP}_{j_\ell}\, d^{-1}_{i_\ell,j_\ell}}{\sum_{i_\ell \in i_{\ell+1}} \mathrm{GDP}_{i_\ell} \sum_{j_\ell \in j_{\ell+1}} \mathrm{GDP}_{j_\ell}}, \tag{2.19}$$

which is the GDP-averaged equivalent of the population-averaged distances commonly used in geography and in the GeoDist database itself [75] (as detailed in Sec. 2.6.2). In this way, $d_{i_{\ell+1},j_{\ell+1}}$ represents a sort of distance between the 'barycenters' of the block-countries $i_{\ell+1}$ and $j_{\ell+1}$, where the barycenter of each $(\ell+1)$-country is defined via the internal GDP distribution across the constituent ℓ-countries.

Putting all the above ingredients together, we arrive at the follow-

ing multi-scale model for the ITN:

$$p_{i_\ell,j_\ell}(\delta) = \begin{cases} 1 - e^{-\delta\,\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}/d_{i_\ell,j_\ell}} & \text{if} \quad i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2}\,\mathrm{GDP}^2_{i_\ell}/d_{i_\ell,i_\ell}} & \text{if} \quad i_\ell = j_\ell \end{cases} \tag{2.20}$$

where δ is the only free parameter and where the renormalization rules are given by Eqs. (2.18) and (2.19).

Once our multi-scale model of the ITN is defined, we build the corresponding instances of the real network at the chosen 18 levels of aggregation. To this end, we construct the empirical 0-graph $\tilde{\mathbf{A}}^{(0)}$ by drawing an undirected link between each pair of countries that have a positive trade relationship in either direction in the BACI-Comtrade dataset (this is detailed as well in Sec. 2.6.2). Then, we use the distance-induced partitions $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$ defined above in order to construct the ℓ-graph according to Eq. (2.1) for each level ℓ. This procedure creates a sequence $\{\tilde{\mathbf{A}}^{(\ell)}\}_{\ell\geq 0}$ of empirical coarse-grained versions of the ITN, each one representing the existence of trade among ℓ-countries.

To test the multiscale model defined by Eq. (2.20) against the real data $\{\tilde{\mathbf{A}}^{(\ell)}\}_{\ell\geq 0}$, we first calibrate it by setting δ to the unique value $\tilde{\delta}$ that produces the same link density D_0 as the real ITN, i.e. such that the expected number of links in the 0-graph (that is a monotonically increasing function of δ) equals the empirical value observed in $\tilde{\mathbf{A}}^{(0)}$. Details on the procedure are provided in Sec. 2.6.3, together with the definition of all the topological properties taken into consideration.
After this single parameter choice, all the probabilities in Eq. (2.20) are uniquely determined at all hierarchical levels and we can test the model by comparing the empirical and expected value of various topological properties of the ITN at different coarse-grainings. In particular, for each level ℓ we consider the *link density* D_ℓ (including possible self-loops) and, for each ℓ-node i_ℓ, the *degree* k_{i_ℓ}, the *average nearest neighbour degree* $k^{nn}_{i_\ell}$ [65] and the *local clustering coefficient* c_{i_ℓ} [10] (see Sec. 2.6.3 for all definitions).

Capturing structural properties at multiple aggregation levels

As a first *global* test of the model, in Fig. 3 we plot, for each hierarchical level ($\ell = 0, 17$), the link density D_ℓ, the normalized overall average nearest neighbour degree $\bar{k}_\ell^{nn}/(N_\ell - 1)$ and the overall local clustering coefficient $\bar{c}_\ell$ as a function of the number N_ℓ of ℓ-nodes (the bar over a quantity denoting an average over all ℓ-nodes). Note that all these global quantities are normalized on the same interval $[0, 1]$, irrespective of ℓ. We find captivating the accordance between the model's predictions and the empirical counterparts, for such wide range of hierarchical levels, given that the model has only one free parameter (δ), which was calibrated uniquely to match the density D_0 of the 0-graph, while the agreement holds for the other quantities as well and across multiple levels. Interestingly, all the rescaled quantities remain roughly constant as the level increases (i.e. as N_ℓ decreases). In line with our previous discussion about the non-equivalence between Eq. (2.4) and the CM and dcSBM, the large values of density confirm that our model is necessarily different from the model that would be obtained by inserting the GDP into the equations for the CM or dcSBM.

As a more stringent test of the model, in Fig. 4 we confirm the prediction that the *local* topological properties of the individual (block-)countries, and in particular k_{i_ℓ}, $k_{i_\ell}^{nn}$ and c_{i_ℓ}, should depend strongly on the empirical value of GDP_{i_ℓ}, in a way that is governed by Eq. (2.20) at all levels. From the moment that, as already noted, the single parameter δ was used to match only the density of the 0-graph, which is a global property defined at a single hierarchical level, the agreement between observations and model expectations at the level of individual nodes and across all hierarchical levels is particularly interesting.
Note that results similar to those shown in Figs. 3 and 4 are retrieved if δ is initially fixed in order to match the empirical density of $\tilde{\mathbf{A}}^{(\ell)}$ for levels $\ell > 0$.

The above results suggest that there is a profound difference between scale-invariant and scale-free networks: the ITN is indeed not a scale-

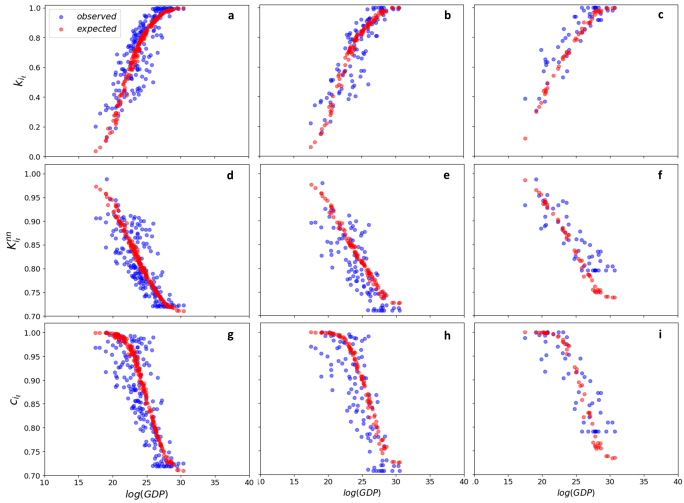

Figure 4: Prediction of local topological properties of the renormalized ITN across the full spectrum of geographical aggregation using the multi-scale model. Top panels (a,b,c): empirical (blue) and expected (red) degree k_{i_ℓ} vs $\ln(\mathrm{GDP}_{i_\ell})$ for all N_ℓ nodes, for three representative hierarchical levels ($\ell_1 = 0$, $\ell_2 = 8$, $\ell_3 = 13$) such that $N_{\ell_1} = 183$ (left), $N_{\ell_2} = 100$ (center) and $N_{\ell_3} = 50$ (right). Middle panels (d,e,f): empirical (blue) and expected (red) average nearest-neighbour degree $k^{nn}_{i_\ell}$ vs $\ln(\mathrm{GDP}_{i_\ell})$ for all N_ℓ nodes, for the same three hierarchical levels. Bottom panels (g,h,i): empirical (blue) and expected (red) local clustering coefficient c_{i_ℓ} vs $\ln(\mathrm{GDP}_{i_\ell})$ for all N_ℓ nodes, for the same three hierarchical levels.

free network (its degree distribution is not power-law [68–71], and in any case could be turned into virtually any distribution via an *ad hoc* coarse-graining), yet its structure turns out to be scale-invariant.

2.4 Annealed fitness

In the annealed case we regard not only the graph structure but also the fitness as a random variable. At the 0-th level, this means that, for all $i_0 = 1, \ldots, N_0$, the value x_{i_0} is drawn from a certain probability density function (PDF) $\rho_{i_0}(x; \mathbf{\Gamma}_{i_0})$ with positive support, where $\mathbf{\Gamma}_{i_0}$ denotes all

parameters of the PDF.
As for the randomness in the topology, we impose that the randomness in the fitness, induced from $\{x_{i_0}\}_{i_0=1}^{N_0}$ to $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ at all higher levels $\ell > 0$ by the additivity property in Eq. (2.6), should be scale-invariant. This means that we should be able to produce the possible values of x_{i_ℓ}, with exactly the same probability, by proceeding along two equivalent ways: hierarchically, by sampling each value x_{i_0} from its PDF $\rho_{i_0}(x;\mathbf{\Gamma}_{i_0})$ and summing up these values for all the 0-nodes that are mapped onto i_ℓ through the partition $\mathbf{\Omega}_{\ell-1}\cdots\mathbf{\Omega}_0$, or directly, by drawing x_{i_ℓ} from a certain PDF $\rho_{i_\ell}(x;\mathbf{\Gamma}_{i_\ell})$ that should have the same functional form of $\rho_{i_0}(x;\mathbf{\Gamma}_{i_0})$ and a set of *renormalized* parameters $\mathbf{\Gamma}_{i_\ell}$ obtainable from $\{\mathbf{\Gamma}_{i_0}\}_{i_0=1}^{N_0}$ only through the knowledge of $\mathbf{\Omega}_{\ell-1}\cdots\mathbf{\Omega}_0$. In other words, the fitness values can be resampled at each scale ℓ from a universal distribution with scale-invariant functional form and possibly scale-dependent parameters. This requirement is equivalent to imposing that $\rho_{i_\ell}(x;\mathbf{\Gamma}_{i_\ell})$ belongs to the family of α-stable distributions [77–79]. A comprehensive introduction of the subject can be found in [79], here we briefly recall the notions that are essential to our purposes. This broad family of distributions emerged in probability theory in the framework of the study of stability properties of sums of independent random variables. After Bernoulli's and De Moivre's embrionic formulations of the Law of Large Numbers and of Central Limit Theorem [78, 80, 81], respectively, the following problem was there: how wide is the class of distribution functions that can play the role of the limiting law? The major contribution in this respect came with Lévy, who in [77, 82] progressively derived the most general functional equations whose solutions define now the family of stable laws.
All the stable laws, except the degenerate ones, are absolutely continuous, which implies that the corresponding distribution functions $F(x)$ possess the densities $\rho(x) = F'(x)$ but, other than a few exceptions, neither the distribution functions nor the densities can be explicitly expressed in terms of elementary functions. Instead, the stable laws can be described in terms of the corresponding characteristic functions:

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x) = \int_{-\infty}^{+\infty} e^{itx}\rho(x)dx \tag{2.21}$$

Indeed the functional equations that define the stable laws become particularly simple in terms of their characteristic functions, and lead to a description of this family of distributions conveyed by four parameters: the *characteristic* $\alpha \in (0, 2]$, the *skew* parameter $\beta \in [-1, 1]$, the *shift* parameter $\mu \in [0, \infty)$ and the *scale* parameter $\gamma \in (-\infty, \infty)$. The only stable distributions that can be written explicitly are the the Gaussian ($\alpha = 2$), the Cauchy ($\alpha = 1, \beta = 0$), and the Lévy ($\alpha = 1/2, \beta = 1$). All the others are expressed through the corresponding characteristic functions, that take the form:

$$\varphi(t; \alpha, \beta, \gamma, \mu) = \begin{cases} e^{it\mu - |\gamma t|^{\alpha}\left[1 - i\beta \mathrm{sign(t)} \tan\frac{\pi\alpha}{2}\right]} & \text{if } \alpha \neq 1, \\ e^{it\mu - |\gamma t|\left[1 + i\beta\frac{2}{\pi}\mathrm{sign(t)} \ln|\mathrm{t}|\right]} & \text{if } \alpha = 1. \end{cases}$$

Stable laws are often convenient in modelling data with long tail behaviour. Indeed it is well known that whenever $\alpha \neq 2$, the asymptotic tail of a stable distribution is given exactly by a power-law: $\rho_{i_\ell}(x; \mathbf{\Gamma}_{i_\ell}) \propto |x|^{-1-\alpha}$ for x large. When $\alpha = 2$, $\rho_{i_\ell}(x; \mathbf{\Gamma}_{i_\ell})$ is instead Gaussian.
The essential property of stable distributions is that sums of stable random variables still follow a stable law with parameters that can be obtained from the old ones by proper rescaling (except for the stability parameter α, which is invariant). This statement can be reframed in our context by going back to the renormalization rule in Eq. (2.6) established for the fitness. Indeed, if the set $\{x_{i_\ell}\}_1^{N_{i_\ell}}$ is composed by random variables with stable distribution $\rho_{i_\ell}(x_{i_\ell}; \alpha, \beta_{i_\ell}, \gamma_{i_\ell}, \mu_{i_\ell})$, then the coarse-grained fitness $\{x_{i_{\ell+1}}\}_1^{N_{i_{\ell+1}}}$ are random variables with stable distribution $\rho_{i_{\ell+1}}(x_{i_\ell+1}; \alpha, \beta_{i_{\ell+1}}, \gamma_{i_{\ell+1}}, \mu_{i_{\ell+1}})$, where:

$$\alpha_{i_{\ell+1}} \equiv \alpha, \tag{2.22}$$

$$\beta_{i_{\ell+1}} \equiv \frac{\sum_{i_\ell \in i_{\ell+1}} \beta_{i_\ell} \gamma_{i_\ell}^{\alpha}}{\sum_{i_\ell \in i_{\ell+1}} \gamma_{i_\ell}^{\alpha}}, \tag{2.23}$$

$$\gamma_{i_{\ell+1}}^{\alpha} \equiv \sum_{i_\ell \in i_{\ell+1}} \gamma_{i_\ell}^{\alpha}, \tag{2.24}$$

$$\mu_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} \mu_{i_\ell}. \tag{2.25}$$

When $0 < \alpha < 1$ and $\beta_{i_\ell} = 1$, the support of α-stable distributions is $[\mu_{i_\ell}, +\infty)$. In order to ensure non-negative fitness values at all scales $\ell \geq 0$ (as required in the connection probability p_{i_ℓ,j_ℓ}), we therefore start from $\ell = 0$ and set $0 < \alpha < 1$, $\beta_{i_0} = 1$ and $\mu_{i_0} = 0$ for all $i_0 = 1, \ldots, N_0$. Note that we might set $\mu_{i_0} > 0$ as well, but in that case Eq. (2.25) would imply an increase of μ_{i_ℓ} with ℓ, while we do not want to progressively restrict the possible values of the fitness as ℓ increases; in other words, we want to keep the support of the fitness distribution scale-invariant. With this choice, Eqs. (2.22)-(2.25) imply that, at all higher levels,

$$\begin{aligned} \alpha_{i_{\ell+1}} &\equiv \alpha \in (0,1), \\ \beta_{i_{\ell+1}} &\equiv 1, \\ \gamma^\alpha_{i_{\ell+1}} &\equiv \sum_{i_\ell \in i_{\ell+1}} \gamma^\alpha_{i_\ell}, \\ \mu_{i_{\ell+1}} &\equiv 0, \end{aligned}$$

showing that α, β and μ are scale-invariant, while γ^α is node-additive.
It is important to note here that the requirement $\alpha \in (0,1)$ implies that *all moments of the fitness distribution diverge* (including the mean). The above rules, combined with the form of $\varphi_{i_\ell}(t; \mathbf{\Gamma}_{i_\ell})$ given above, finally lead to the scale-invariant CF of the fitness, for all $\alpha \in (0,1)$ and for all $\gamma_{i_\ell} > 0$:

$$\varphi_{i_\ell}(t; \alpha, \gamma_{i_\ell}) = e^{-|\gamma_{i_\ell} t|^\alpha \left[1 - i\,\mathrm{sign(t)} \tan \frac{\pi\alpha}{2}\right]}. \tag{2.26}$$

This choice corresponds to the so-called class of *one-sided stable distributions* [83]. For this particular class it is also known that, up to a scale transformation reabsorbed in the value of $\gamma_\alpha \equiv [\cos(\alpha\pi/2)]^{1/\alpha}$, the Laplace transform (LT) $\lambda_{i_\ell}(t; \alpha, \gamma_\alpha)$ of the PDF $\rho_{i_\ell}(x; \alpha, \gamma_\alpha)$ equals

$$\lambda_{i_\ell}(t; \alpha, \gamma_\alpha) \equiv \int_0^{+\infty} e^{-tx} \rho_{i_\ell}(x; \alpha, \gamma_\alpha) \mathrm{d}x = e^{-t^\alpha} \tag{2.27}$$

which is a stretched exponential [83–86].
As a first attempt to characterize the random graphs generated by these stable random fitness, we will restrict to the case in which the fitness variables are drawn according to the only stable distribution known in

closed form within the above constraints, *i.e.* the Lévy distribution, for which $\alpha = 1/2$:

$$\rho_{i_\ell}(x; 1/2, \gamma_{i_\ell}) = \sqrt{\frac{\gamma_{i_\ell}}{2\pi}} \frac{e^{-\gamma_{i_\ell}/(2x)}}{x^{3/2}} \quad (x > 0), \tag{2.28}$$

where we have restored the arbitrary parameter $\gamma_{i_\ell} > 0$, which is the only remaining free parameter and is subject to renormalization rule given by Eq. (2.24).

In summary, in the annealed scenario, at any hierarchical level ℓ the fitness of each ℓ-node is a random variable described by the CF $\varphi_{i_\ell}(t; \alpha, \gamma_{i_\ell})$ in Eq. (2.26), or, equivalently, by the LT $\lambda_{i_\ell}(t; \alpha, \gamma_\alpha)$ in Eq. (2.27). If $\alpha = 1/2$, the PDF is known explicitly from Eq. (2.28) and such that $\rho_{i_\ell}(x; 1/2, \gamma_{i_\ell}) \propto x^{-3/2}$ for x large, while for general $\alpha \in (0, 1)$ we know that $\rho_{i_\ell}(x; \alpha, \gamma_{i_\ell}) \propto x^{-1-\alpha}$ for x large, even if the explicit form is not known.
Then, given a realization of fitness values, the network is generated with probability $P\big(\mathbf{A}^{(\ell)}; \delta\big)$ given by Eq. (2.8), i.e. by connecting pairs of ℓ-nodes with connection probability $p_{i_\ell, j_\ell}(\delta)$ given by Eq. (2.4).
This construction is entirely self-consistent across all hierarchical levels: the ℓ-graph can either be built bottom-up, starting from level 0 and coarse-graining the 0-graph up to level ℓ, or directly at the ℓ-th level, by sampling the fitness at that level and generating the resulting ℓ-graph immediately. Note that, up to this point, the connection probability p_{i_ℓ, j_ℓ} can still depend on the distances d_{i_ℓ, j_ℓ} as long as the latter are ultrametric on the histogram of desired coarse grainings (and therefore decoupled from the fitness). If the distances between 0-nodes are not ultrametric, Eq. (2.7) would make the renormalized distances fitness-dependent and hence random, which is something that can be subtle to handle. Therefore for now we refrain to include such a possibility.

2.4.1 From semi-group to group

A notable property of the annealed case is that the renormalization procedure defines not only a *semi-group* proceeding bottom-up from the

0-graph to higher levels as in usual schemes, but also a *group*: it can proceed top-down as well, by resolving the 0-graph into a graph with any number of nodes larger than N_0, indefinitely and in a scale-invariant manner. This possibility is ensured by the fact that stable distributions are *infinitely divisible*, i.e. they can be expressed as the probability distribution of the sum of an arbitrary number of i.i.d. random variables from the same family. This property implies that we can disaggregate each ℓ-node (including $\ell = 0$) with fitness x_{i_ℓ} into any desired number of $(\ell-1)$-nodes, each with its own fitness.
This possibility allows us to perform the *upscaling* of the network, in a way that is conceptually similar to, but physically different from, the approach in Ref. [87] (which assumes a geometric embedding of nodes). We can therefore attach no particular meaning to the level $\ell = 0$ and consider any 'negative' level $m < 0$ (stretching all the way down to $m = -\infty$) as well, provided that the (ultrametric) distances between all pairs of m-nodes are given and consistent with the higher-level ones, i.e. such that $f(d_{i_\ell,j_\ell}) = f(d_{i_m,j_m})$ whenever $i_\ell = \mathbf{\Omega}_{\ell-1} \cdots \mathbf{\Omega}_m(i_m)$ and $j_\ell = \mathbf{\Omega}_{\ell-1} \cdots \mathbf{\Omega}_m(j_m)$ for all $\ell > m$. This requirement is always ensured in two cases: *i)* if distances are ultrametric and the associated dendrogram is used to define which m-nodes branch into which $(m-1)$-nodes as we go deeper in the hierarchy of partitions; *ii)* in the distance-free case $f \equiv 1$. We consider the latter an instructive example and discuss it in the rest of the section.

In general, we may start from $\ell = 0$ and assign to each 0-node i_0 a different value of γ_{i_0}, then specify a hierarchy of coarse-grainings (and even fine-grainings) and calculate the corresponding values of γ_{i_ℓ} for all ℓ-nodes and the resulting properties of the network, for all $\ell \neq 0$. This leaves a lot of flexibility, in principle allowing us to tailor the resulting properties of the network to any degree of heterogeneity. However, to avoid making ad hoc assumptions, we put ourselves in the simplest situation, where the following premises are settled:

- distances are switched off (*i.e.* $f \equiv 1$), so that the model is governed by Eq. (2.17) and is entirely *non-'geometric'*;

- all 0-*nodes are statistically equivalent* (*i.e.* $\gamma_{i_0} \equiv \gamma_0$ for all $i_0 = 1, \ldots, N_0$);

- The dendrogram of coarse-grainings is m_ℓ-*regular*: at each level ℓ, the N_ℓ ℓ-nodes are merged into a number $N_{\ell+1}$ of $(\ell+1)$-nodes (each formed by exactly m_ℓ ℓ-nodes), given by

$$N_{\ell+1} = \frac{N_\ell}{m_\ell} = \cdots = \frac{N_0}{\prod_{q=0}^{\ell} m_q} \tag{2.29}$$

This is the most homogeneous choice, as it preserves the statistical equivalence of all the N_ℓ ℓ-nodes at every hierarchical level, implying $\gamma_{i_\ell} \equiv \gamma_\ell$ for all $i_\ell = 1, \ldots, N_\ell$ where

$$\gamma_{\ell+1} = {b_\ell}^{1/\alpha}\gamma_\ell = \cdots = \prod_{m=0}^{\ell} b_m^{1/\alpha}\gamma_0 = \left(\frac{N_0}{N_{\ell+1}}\right)^{1/\alpha}\gamma_0 \tag{2.30}$$

(with $\alpha = 1/2$ here), as ensured by Eq. (2.24). This means that, for any ℓ, the fitness values $\{x_{i_\ell}\}_{\ell=1}^{N_\ell}$ are i.i.d. with common distribution

$$\rho_\ell(x; 1/2, \gamma_\ell) = \sqrt{\frac{\gamma_\ell}{2\pi}}\frac{e^{-\gamma_\ell/(2x)}}{x^{3/2}} \quad (x > 0), \tag{2.31}$$

effectively reducing a multivariate problem to a univariate one.

Note that the resulting probability of generating a graph at a given hierarchical level ℓ does not depend on the labeling of nodes, i.e., it is unchanged upon permutations of the nodes' labels. This property is known as *exchangeability* [88, 89] and can be a desirable property of random graph models [90, 91].

The above prescriptions make the model similar to an annealed version of the FM [20] or equivalently to the class of (rank-1) IRGM [55], with two special requirements: *i)* here the fitness is defined at all hierarchical levels simultaneously and *ii)* the connection probability can only take the scale-invariant form given by Eq. (2.17).

Note that the fitness distribution depends on the hierarchical level ℓ

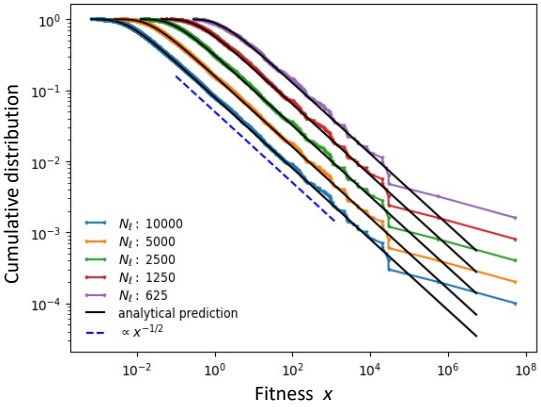

Figure 5: Fitness distribution in the annealed scale-invariant model. The points show the cumulative distribution of the node fitness x across five different hierarchical levels ($\ell = 0, 1, 2, 3, 4$), for a single realization from an α-stable distribution with parameter choice $\alpha = 1/2$, $N_0 = 10^4$, $b = 2$. The solid lines are the corresponding analytical α-stable cumulative distributions obtained integrating Eq. (2.31) with $\gamma_{\ell+1} = b^{1/\alpha}\gamma_\ell$, $\alpha = 1/2$ and $b = 2$. The dashed line is a power-law with exponent $-1/2$, confirming that the non-cumulative fitness distribution has power-law tails with exponent $-3/2$. Note that there is no upper cut-off to this tail, despite the increasing network density for higher hierarchical levels, because the fitness of a node has no bounds.

through the parameter γ_ℓ, which, as clear from Eq. (2.30), cannot decrease from the moment N_ℓ cannot increase. This implies an overall shift towards larger fitness values as nodes are coarse-grained. For instance, if we take $b_\ell = b$ (i.e., the branching ratio is level-independent), then Eq. (2.30) implies $\gamma_\ell = b^{\ell/\alpha}\gamma_0 = b^{2\ell}\gamma_0$ and the corresponding behaviour of the fitness distribution is illustrated in Fig. 5. Irrespective of the rightward shift, the tail of the fitness distribution is always a pure power-law proportional to $x^{-1-\alpha}$, independently of ℓ. We will adopt the choice $b_\ell = b$ throughout the rest of the chapter, although all the results that we obtain for a fixed hierarchical level ℓ hold irrespective of this choice and are therefore general. Indeed the choice of a level-independent b only affects how the calculated quantities change across hierarchical levels.

In what follows, we focus on the analysis of certain topological properties of the network realizations drawn from the SIM ensemble, as above described.
Since the fitness is annealed, the expected local topological properties involving each node i_ℓ, when averaged over the randomness of the fitness, will be identical. However, what interests us is deriving typical structural patterns relating, irrespective of the particular realization of the fitness (and hence surving after averaging over such realizations), the correlation between different local properties of nodes. For instance, we are interested in the expected degree $k_\ell(x)$ of an ℓ-node whose realized fitness is x at level ℓ, since all ℓ-nodes with the same value x of the realized fitness are statistically equivalent in the random realizations of the network, so the expected degree only depends on x. In this way, we necessarily lose the identity of the node (since each ℓ-node i_ℓ is assigned different values of the fitness x_{i_ℓ} in different realizations) but we keep the structural relationship between degree and fitness. We can therefore drop the subscript i_ℓ accompanying any local topological property (such as k_{i_ℓ}, $k^{nn}_{i_\ell}$, c_{i_ℓ}) and replace it with the dependence of the expected value of that property on the realized fitness x.

In the next section, we will use Eq. (2.31) to provide an analytical characterization of the annealed model when $\alpha = 1/2$, although we will also retrieve similar results for all $\alpha \in (0, 1)$ either analytically or through numerical sampling of the fitness. Further progress is possible by replacing the $\alpha - stable$ PDF with a pure Pareto one with the same tail exponent $-1-\alpha$, as pursued in Chap. 4.

2.4.2 Scale-free networks from scale-invariance without geometry

In Sec. 2.2.4 some reflections on why and in which sense scale-free and scale-invariant networks should be regarded as distinct concepts are expounded. Here we consider a special case of our annealed scale-invariant model that spontaneously leads to scale-free networks,

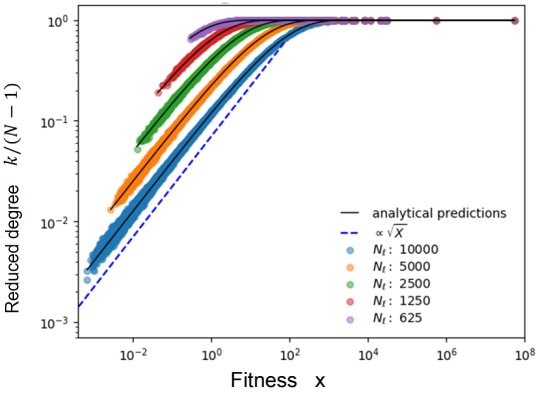

Figure 6: Reduced degree as a function of fitness in the annealed scale-invariant model. The circles represent the reduced degree κ_ℓ of each ℓ-node as a function of the corresponding fitness x in numerical simulations of the model across five different hierarchical levels ($\ell = 0, 1, 2, 3, 4$), for the parameter choice $\alpha = 1/2$, $N_0 = 10^4$, $b = 2$. The solid lines are the expected theoretical relationship $\kappa_\ell(x)$ obtained via Eq. (2.38) for the same parameter values. The dashed line is proportional to the square root of the fitness, emphasizing the behaviour $\kappa_\ell(x) \approx \sqrt{2\delta\gamma_\ell x}$ of the (reduced) degree of nodes with small fitness. For generic $\alpha \in (0, 1)$, the (reduced) degree of nodes with small fitness is proportional to x^α.

thus connecting the two concepts and providing a nontrivial recipe for generating scale-freeness purely from scale invariance.

Let us start by deriving the functional form of the expected degree distribution. To this end, for any fixed hierarchical level ℓ we adapt the procedure outlined in Ref. [20] to compute, for a typical realization of the fitness values, the distribution $P_\ell(k)$ of expected (over the realizations of the network) degrees from the PDF of the fitness $\rho_\ell(x)$ and the connection probability p_{i_ℓ,j_ℓ}, written as a function $p_{i_\ell,j_\ell} = f(x_{i_\ell}, x_{j_\ell})$ of the fitness of the nodes involved, where in our case

$$f(x, y) = 1 - e^{-\delta\, x\, y}. \tag{2.32}$$

We first notice that since $f(x, y)$ is an increasing function of both its argu-

ments, the expected degree $\langle k_{i_\ell} \rangle$

$$\langle k_{i_\ell} \rangle = \sum_{j_\ell \neq i_\ell} p_{i_\ell, j_\ell} = \sum_{j_\ell \neq i_\ell} f(x_{i_\ell}, x_{j_\ell}) \tag{2.33}$$

is an increasing function of the fitness x_{i_ℓ}.
For a large number N_ℓ of ℓ-nodes, the above discrete sum can be approximated by an integral over the number $(N_\ell - 1)\rho_\ell(y; \alpha, \gamma_\ell)$ of ℓ-nodes (except i_ℓ itself) with fitness in a neighbourhood of y: if $k_\ell(x)$ denotes the expected degree of a node with fitness x at level ℓ, we have

$$\begin{aligned} k_\ell(x) &= (N_\ell - 1) \int_0^\infty f(x, y) \rho_\ell(y; \alpha, \gamma_\ell) \mathrm{d}y \\ &= (N_\ell - 1) \left(1 - \int_0^\infty e^{-\delta x y} \rho_\ell(y, \alpha, \gamma_\ell) \mathrm{d}y \right) \\ &= (N_\ell - 1) \left(1 - \lambda_\ell(\delta x; \alpha, \gamma_\ell) \right) \end{aligned} \tag{2.34}$$

where $\lambda_\ell(t; \alpha, \gamma_\ell)$ denotes the LT of $\rho_\ell(x, \alpha, \gamma_\ell)$ as in Eq. (2.27).
When $\alpha = 1/2$, the LT can be calculated explicitly as

$$\begin{aligned} \lambda_\ell(\delta x; 1/2, \gamma_\ell) &= \int_0^\infty e^{-\delta x y} \sqrt{\frac{\gamma_\ell}{2\pi}} \frac{e^{-\gamma_\ell/(2y)}}{y^{3/2}} \mathrm{d}y \\ &= e^{-\sqrt{2\delta\gamma_\ell x}}, \end{aligned} \tag{2.35}$$

so that

$$k_\ell(x) = (N_\ell - 1) \left(1 - e^{-\sqrt{2\delta\gamma_\ell x}} \right). \tag{2.36}$$

It is convenient to rescale the degree k_ℓ by $N_\ell - 1$, thereby defining the *reduced degree*

$$\kappa_\ell \equiv \frac{k_\ell}{N_\ell - 1} \in [0, 1], \tag{2.37}$$

whose range is independent of ℓ and whose node-averaged value $\bar{\kappa}_\ell$ coincides with the network density *excluding self-loops*.
Clearly, Eq. (2.36) is equivalent to

$$\kappa_\ell(x) = 1 - e^{-\sqrt{2\delta\gamma_\ell x}}, \tag{2.38}$$

an exact calculation that is confirmed by numerical simulations, as shown in Fig. 6.

To aid the comparison with other network models, it is worth noting that, for x sufficiently small, Eq. (2.36) is approximated by $k_\ell(x) \approx (N_\ell - 1)\sqrt{2\delta\gamma_\ell x}$, i.e., the expected degree of nodes with small fitness is proportional to the *square root* of the fitness, not the fitness itself (as also illustrated in Fig. 6). The latter remark can be extended to the general case $\alpha \in (0, 1)$, although approximately, through the last equality in Eq. (2.34): for x sufficiently small, we have $\kappa_\ell(x) \propto x^\alpha$. This result reflects the aforementioned key difference between the annealed scale-invariant model and the CM: even for very small values of the fitness, Eq. (2.17) does *not* reduce to $p_{i_\ell,j_\ell}(\delta) \approx \delta x_{i_\ell} x_{j_\ell}$ and the expected degree is *not* proportional to the fitness. This is due to the divergence of all moments of the fitness in the annealed case (due to the restriction $0 < \alpha < 1$), which implies $\max_{i_\ell}\{x_{i_\ell}\} = +\infty$ and makes the regime $\delta \ll (\max_{i_\ell}\{x_{i_\ell}\})^{-2}$ (usually assumed in the sparse CM) impossible, irrespective of the hierarchical level ℓ.

By inverting Eq. (2.36), we find that the fitness x_ℓ of an ℓ-node with expected degree k at level ℓ is

$$x_\ell(k) = \frac{1}{2\delta\gamma_\ell} \ln^2\left(\frac{N_\ell - 1}{N_\ell - 1 - k}\right), \tag{2.39}$$

which implies

$$\frac{\mathrm{d}}{\mathrm{d}k} x_\ell(k) = \frac{\ln\left(\frac{N_\ell - 1}{N_\ell - 1 - k}\right)}{\delta\gamma_\ell(N_\ell - 1 - k)}. \tag{2.40}$$

We can use the above expressions to obtain the distribution $P_\ell(k)$ of the expected degrees from the distribution $\rho_\ell(x; \alpha, \gamma_\ell)$ of the corresponding fitness. Indeed, starting from the fundamental equation

$$P_\ell(k)\mathrm{d}k = \rho_\ell\big(x_\ell(k); \alpha, \gamma_\ell\big)\mathrm{d}x_\ell(k) \tag{2.41}$$

relating the probability distributions of the two random variables k and x, and using Eqs. (2.31), (2.39) and (2.40), we arrive at the explicit form of

the distribution of expected degrees:

$$P_\ell(k) = \frac{2\gamma_\ell \sqrt{\frac{\delta}{\pi}} \exp\left[\frac{-\delta\gamma_\ell^2}{\ln^2\left(1-\frac{k}{N_\ell-1}\right)}\right]}{(N_\ell - 1 - k)\ln^2\left(1-\frac{k}{N_\ell-1}\right)} \tag{2.42}$$

for $k \geq 0$, and $P_\ell(k) = 0$ otherwise.

The above degree distribution shows a twofold dependence on the hierarchical level ℓ, as there are two contrasting tendencies as ℓ increases. On the one hand, the number of nodes N_ℓ decreases, hence the possible range of values $[1, N_\ell - 1]$ for the degree k shrinks: this implies a tendency for the degree to decrease. On the other hand, the ongoing coarse-graining is such that, on average, ℓ-nodes acquire more and more links as ℓ increases: this implies a tendency for the degree to increase. We can remove the effect of the first tendency by considering the probability distribution $Q_\ell(\kappa)$ for the reduced degree κ_ℓ, which is easily calculated from $P_\ell(k)$ as

$$Q_\ell(\kappa) = \frac{P_\ell[(N_\ell - 1)\kappa]}{1/(N_\ell - 1)} = \frac{2\gamma_\ell \sqrt{\frac{\delta}{\pi}} \exp\left[\frac{-\delta\gamma_\ell^2}{\ln^2(1-\kappa)}\right]}{(1-\kappa)\ln^2(1-\kappa)}. \tag{2.43}$$

We see that the distribution has a residual dependence on the level ℓ through the parameter γ_ℓ. As a consequence, the reduced degree distributions obtained for different hierarchical levels do not collapse upon each other, as confirmed in Fig. 7 using the same parameter choice as above. This is purely the effect of the second tendency. Indeed we see that, as ℓ increases, there is a more and more pronounced accumulation of values of the reduced degree κ_ℓ close to the maximum value 1. This is a saturation effect cutting off the tail of the degree distribution.

Importantly, for values of the degree that are sufficiently lower than the upper cut-off, the distribution has a universal power-law trend proportional to κ^{-2}, for all values of $\alpha \in (0, 1)$ (hence without requiring a fine-tuning of α to a specific value in that interval). Indeed, one can show analytically (see Sec. 2.6.4) that the right tail of the reduced degree

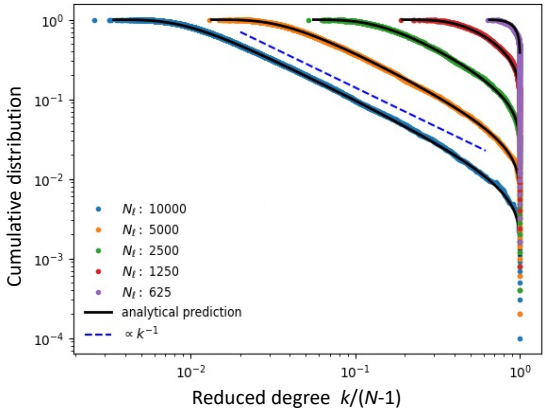

Figure 7: Degree distribution in the annealed scale-invariant model. Cumulative degree distribution (fraction of nodes with reduced degree $\geq \kappa$) across five different hierarchical levels ($\ell = 0, 1, 2, 3, 4$), for the parameter choice $\alpha = 1/2$, $N_0 = 10^4$, $b = 2$. The circles represent a single realization of the network, while the solid lines correspond to the theoretical prediction given by Eq. (2.43). The dashed line is a power-law with exponent -1, corresponding to a power-law $Q_\ell(\kappa) \propto \kappa^{-2}$ for the non-cumulative distribution. This exponent is universal for all $\alpha \in (0, 1)$ and different from the exponent $-1-\alpha$ of the corresponding non-cumulative fitness distribution. Another difference is the presence of an upper cut-off $C_\ell(\kappa)$ (due to the fact that κ cannot exceed 1) becoming stronger as the hierarchical level increases.

distribution behaves as

$$Q_\ell(\kappa) \approx \kappa^{-2} C_\ell(\kappa), \tag{2.44}$$

where $C_\ell(\kappa)$ is a cut-off function with a peak at values of κ that increase towards 1 as ℓ increases. The cut-off function captures stronger and stronger finite-size effects as the network size shrinks under the effect of coarse-graining. In the opposite direction (decreasing ℓ), one can always reach the sparse regime through fine-graining, i.e. by subdividing each ℓ-node into multiple $(\ell-1)$-nodes and so on. In such a regime, the effect of the cut-off function practically disappears and the network is essentially scale-free with universal degree exponent -2. This conjecture will be made more rigorous in chapter 4, where the hierarchical level at which this becomes

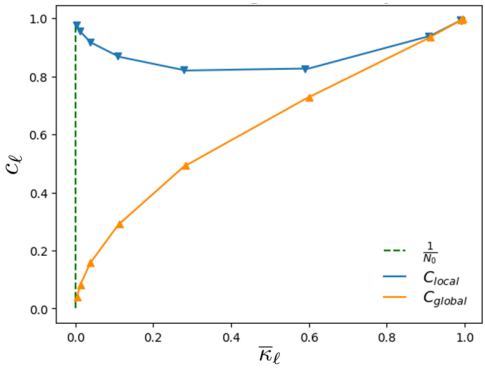

Figure 8: Local and global clustering coefficient as a function of density along the coarse-graining flow. The average local (c_ℓ^{local}) and global (c_ℓ^{global}) clustering coefficients are shown as a function of the network density (excluding self-loops) $\bar{\kappa}_\ell$ for different coarse-grainings of the scale-invariant model with $\alpha = 1/2$, $N_0 = 10^4$, $b = 2$. Triangles refer to a single realization of the (coarse-grained) network, while the solid lines show the expected values. The dashed line is a reference corresponding to a density $1/N_0 = 10^{-4}$.

true is made explicit. As anticipated, this exponent is different from the tail exponent $-1-\alpha \in (-2,-1)$ of the corresponding fitness distribution $\rho_\ell(x)$, as a consequence of the divergence of all moments of $\rho_\ell(x)$ and the related non-linear dependence between degree and fitness, even for small fitness values.
A mechanism producing the universal exponent -2 has been advocated previously ([21, 92]) and, more generally, the latter falls within the empirical range of exponents observed for the vast majority of networks, which are found in the interval $(-3,-2]$ [24].

2.4.3 Assortativity and clustering

Here we focus on the assortativity and clustering patterns, including a nonvanishing average local clustering coefficient, which we observe in our ensemble.

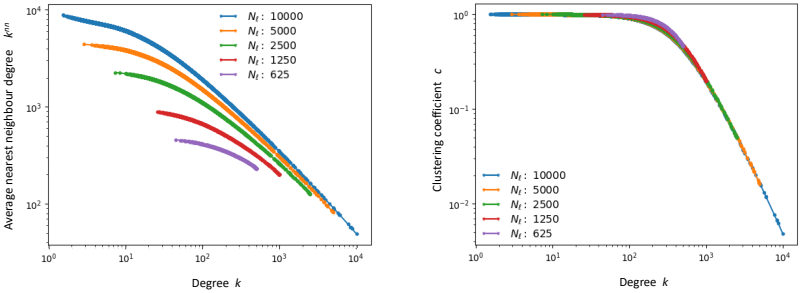

Figure 9: Local assortativity and clustering properties. Average nearest neighbour degree k_i^{nn} (left) and local clustering coefficient c_i (right) versus degree k_i in the annealed scale-invariant model across different hierarchical levels, for the parameter choice $\alpha = 1/2$, $N_0 = 10^4$, $b = 2$.

In Fig. 8 we show, as a function of the link density (or equivalently the average reduced degree $\bar{\kappa}_\ell$), the node-averaged local clustering coefficient $c_\ell^{local} \equiv \bar{c}_\ell = \sum_{i_\ell=1}^{N_\ell} c_{i_\ell}/N_\ell$ and the global clustering coefficient $c_\ell^{\text{global}} \equiv \Delta_\ell/\Lambda_\ell$ obtained for different hierarchical levels. The latter is defined as the ratio of the overall number Δ_ℓ of realized triangles (each counted three times) to the number Λ_ℓ of *wedges*, i.e. *potential* triangles [93–96] (the exact formula can be found in Sec. 2.6.3). Various studies have shown that, apart from cases where the network is sufficiently homogeneous [93], the average local and global clustering coefficients can be quite different [94–97]. In particular, an empirically widespread property of real networks is their "large" overall local clustering, defined as a non-vanishing (strictly positive) node-averaged local clustering coefficient even in the sparse regime where the network density $\bar{k}_\ell$(without self-loops), and possibly the global clustering coefficient, vanishes as the number of nodes increases.

In our model, we can numerically assess the behaviour of both clustering coefficients as functions of the link density $\bar{\kappa}_\ell$, whose expected value can easily be written in a concise way for $\alpha = 1/2$, as we show in the following lines. The expected link density $\bar{\kappa}_\ell$ (excluding self-loops) is

given by

$$\begin{aligned} \langle \bar{\kappa}_\ell \rangle &= \frac{1}{N_\ell - 1} \int_0^{N_\ell - 1} P_\ell(k) k \, \mathrm{d}k \\ &= \left(1 - \frac{2}{\pi} \int_0^\infty e^{-t_\ell^2 - \gamma_\ell \sqrt{\delta}/t_\ell} \mathrm{d}t_\ell \right), \end{aligned} \tag{2.45}$$

where we have changed variables by introducing

$$t_\ell = \frac{\gamma_\ell \sqrt{\delta}}{\ln \frac{N_\ell - 1}{N_\ell - 1 - k}}.$$

The integral in Eq. (2.45) corresponds to one of the Meijer-G functions $G^{\,m,n}_{p,\,q}\left({a_1,\ldots,a_p \atop b_1,\ldots,b_q} \middle| z \right)$, thus yielding:

$$\langle \bar{\kappa}_\ell \rangle = 1 - \frac{\gamma_\ell \sqrt{\delta}}{2\pi} G^{\,3,0}_{0,3}\left({-1/2, \dot{0}, 0 \atop } \middle| \delta \gamma_\ell^2/4 \right). \tag{2.46}$$

As a side remark, note that $G^{\,3,0}_{0,3}\left({-1/2, \dot{0}, 0 \atop } \middle| \delta \gamma_\ell^2/4 \right)$ is a decreasing function of the combination $\delta \gamma_\ell^2$, therefore the network density increases as the level ℓ increases.

Now, in the annealed scenario considered here, we are simultaneously generating graphs at all scales $\ell = -\infty, \ldots, +\infty$, ranging from the fully connected regime ($\langle \bar{\kappa}_{+\infty} \rangle = 1$) to the fully disconnected one ($\langle \bar{\kappa}_{-\infty} \rangle = 0$). We can therefore inspect the expected average local clustering coefficient c_ℓ^{local} and expected global clustering coefficient c_ℓ^{global} as a function of the network density $\bar{\kappa}_\ell$. The outcome is shown in Fig. 8, which illustrates how c_ℓ^{global} decreases as the density decreases (that is, as the level ℓ decreases), while c_ℓ^{local} retains finite values. In particular we see that, even for the particular hierarchical level(s) ℓ^* corresponding to the sparse regime $\langle \bar{\kappa}_{\ell^*} \rangle \propto 1/N_{\ell^*}$, $c_{\ell^*}^{\text{local}}$ remains finite despite the asymptotic vanishing of the network density. In random graph models, a non-vanishing local clustering coefficient and a vanishing global clustering coefficient in the sparse regime have been found also in the CM [98] (in the limit where the tail exponent of the degree distribution approaches the value -2 found here), in a class of 'windmill' graphs [95] and in the hyperbolic geometric

model [96, 97]. In real-world networks, a typical tendency of the global clustering coefficient to be significantly smaller than the average local clustering coefficient (and even vanishing) has also been documented [95].

We finally consider the average nearest neighbour degree $k_{i_\ell}^{nn}$ [65] and local clustering coefficient c_{i_ℓ} [10] as a function of the degree k_{i_ℓ} of each ℓ-node. These quantities are plotted in Fig. 9 for $\alpha = 1/2$. The plots show decreasing trends for both $k_{i_\ell}^{nn}$ and c_{i_ℓ} as k_{i_ℓ} increases. Together with the presence of a power-law degree distribution with a cut-off, these properties are widespread in real-world networks [10, 24, 65].

2.5 Discussion

We proposed a renormalization scheme based on the identification of a scale-invariant random graph model. The functional form of the probability for two nodes to be connected is independent on the hierarchical level being considered. At each level, the model can generate any network in two possible ways, with exactly the same probability: either hierarchically, by generating the finest-grained network and then coarse-graining it via progressive non-overlapping (but otherwise arbitrary) partitions, or directly, using appropriately renormalized parameters. These parameters include a global scale-invariant density parameter, a necessary set of hidden 'fitness' variables attached to each (block-)node, and, if useful, a set of dyadic factors representing distances or communities. It turns out that the model possesses scale-invariance without postulating the existence of node coordinates in an underlying metric space.
If the fitness values are treated as quenched, the model can guide the renormalization of real-world graphs. In this case, the parameters of the model can be identified with empirical quantities attached to nodes and dyads. In our application to the ITN, we found that a one-parameter fit of the model to the observed network density is enough to accurately replicate many local topological properties of individual nodes, even across several hierarchical levels of resolution. This result exemplifies the deep *a priori* conceptual distinction between scale-free networks (in the sense of

power-law degree distributions, which are absent in the ITN) and scale-invariant networks (in the sense of the network formation mechanisms being consistent across scales, as found in the ITN) highlighted by the model.

If the fitness values are annealed, the model naturally leads to one-sided Lévy-stable fitness distributions, which are characterized by a tail exponent $-1-\alpha \in (-2,-1)$. The properties of stability and infinite divisibility of these distributions allow for the definition of a renormalization scheme in both forward (coarse-graining) and backward (fine-graining) directions. At the same time, the divergence of all moments of these distributions implies that the multiscale model is not asymptotically equivalent to the CM and dcSBM, showing that those models are not renormalizable. The annealed version of the model has also the property of exchangeability, which means that graph probabilities are unchanged upon relabelling of nodes. It turns out that the requirement of scale invariance spontaneaously leads to scale-free networks with degree distribution featuring a universal power-law decay $P(k_\ell) \propto k^{-2}$ (which does not require a fine-tuning of α) followed by a density-dependent cut-off and with realistic assortativity and clustering properties, without postulating mechanisms such as growth, preferential attachment or hyperbolic embedding. In particular, in the sparse regime the model appears simultaneously scale-free and locally clustered, with no need for metric distances producing clustering as a result of triangular inequalities as postulated in the hyperbolic model [21]. Importantly, the desirable topological properties generated by our annealed model differ from those of random graphs that, while being defined through the same connection probability as in Eq. (2.17), are characterized by fitness variables with finite mean [58, 59]. Indeed, those studies did not consider a scale-invariant random graph setting, and consequently they did not demand that in the annealed setting the fitness variables are α-stable random variables, hence with $0 < \alpha < 1$ because of the positivity of the fitness.

2.6 Supporting material

2.6.1 Determining the scale-invariant connection probability

Here we show how the scale-invariance requirement stated in Eq. (2.2), for any model with independent links as formulated in Eq. (2.3), leads to the unique form of the connection probability given by Eq. (2.4).

Let us consider a partition $\mathbf{\Omega}_\ell$ that maps an ℓ-graph with N_ℓ ℓ-nodes and adjacency matrix $\mathbf{A}^{(\ell)}$ to an $(\ell+1)$-graph with $N_{\ell+1}$ $(\ell+1)$-nodes and adjacency matrix $\mathbf{A}^{(\ell+1)}$. The relation between the entries of the matrices $\mathbf{A}^{(\ell)}$ and $\mathbf{A}^{(\ell+1)}$ is given by Eq. (2.1). Now, for any random graph model with independent links as stated in Eq. (2.3), $a^{(\ell)}_{i_\ell,j_\ell}$ is a Bernoulli random variable equal to 1 with probability $p^{(\ell)}_{i_\ell,j_\ell}$ and equal to 0 with probability $1-p^{(\ell)}_{i_\ell,j_\ell}$. Similarly, $a^{(\ell+1)}_{i_{\ell+1},j_{\ell+1}}$ is a Bernoulli random variable equal to 1 with probability $p^{(\ell+1)}_{i_{\ell+1},j_{\ell+1}}$ and equal to 0 with probability $1-p^{(\ell+1)}_{i_{\ell+1},j_{\ell+1}}$. Now, the scale-invariance requirement in Eq. (2.2) demands that we should create, *with equal probability*, any of the possible realizations of the adjacency matrix $\mathbf{A}^{(\ell+1)}$ either by: *i)* generating the possible realizations of the matrix $\mathbf{A}^{(\ell)}$ (using the associated probabilities $\{p^{(\ell)}_{i_\ell,j_\ell}\}$) and then aggregating the corresponding ℓ-graphs into $(\ell+1)$-graphs, or *ii)* directly generating all the possible realizations of the matrix $\mathbf{A}^{(\ell+1)}$ (using the associated probabilities $\{p^{(\ell+1)}_{i_{\ell+1},j_{\ell+1}}\}$). Scale-invariance also demands that $p^{(\ell)}_{i_\ell,j_\ell}$ depends on ℓ only through its parameters. Assuming that these parameters are a combination of global (δ_ℓ), node-specific (x_{i_ℓ}, x_{j_ℓ}) and dyadic (d_{i_ℓ,j_ℓ}) factors, we can write $p^{(\ell)}_{i_\ell,j_\ell}(\delta_\ell) = p_{i_\ell,j_\ell}(\delta_\ell)$. Enforcing scale-invariance means finding not only the functional form of p_{i_ℓ,j_ℓ}, but also the renormalization rules connecting $\delta_\ell, x_{i_\ell}, x_{j_\ell}, d_{i_\ell,j_\ell}$ to their next-level counterparts $\delta_{\ell+1}, x_{i_{\ell+1}}, x_{j_{\ell+1}}, d_{i_{\ell+1},j_{\ell+1}}$.

To enforce the scale-invariance requirement, we first consider the case when the connection at the coarse-grained level $\ell+1$ involves two distinct blocks $i_{\ell+1} \neq j_{\ell+1}$. In this case, since a link between the pair $(i_{\ell+1}, j_{\ell+1})$ of $(\ell+1)$-nodes is present if and only if there is at least one link present

between any pair (i_ℓ, j_ℓ) of ℓ-nodes such that $i_\ell \in i_{\ell+1}$ and $j_\ell \in j_{\ell+1}$, the probability that $i_{\ell+1}$ and $j_{\ell+1}$ are *not* connected is equal, according to the procedure *ii)* described above, to the probability that none of the pairs of underlying ℓ-nodes is connected. Since links are independent, this probability equals $\prod_{i_\ell \in i_{\ell+1}} \prod_{j_\ell \in j_{\ell+1}} [1 - p_{i_\ell, j_\ell}(\delta)]$. On the other hand, according to the procedure *i)*, the same event occurs with probability $1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)$. Enforcing the equality between the two probabilities leads to the condition

$$1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta) = \prod_{i_\ell \in i_{\ell+1}} \prod_{j_\ell \in j_{\ell+1}} [1 - p_{i_\ell, j_\ell}(\delta)] \,. \tag{2.47}$$

Taking the logarithm of both sides of Eq. (2.47), we obtain

$$\ln \left[1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)\right] = \sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in j_{\ell+1}} \ln \left[1 - p_{i_\ell, j_\ell}(\delta)\right] , \tag{2.48}$$

from which we can now derive the scale-invariant form of the connection probability.

Note that Eq. (2.47) is consistent with taking the expected values of both sides of Eq. (2.1). However, it cannot be derived directly in that way, because the two expected values are taken with respect to different probability distributions having different support, i.e. $P\big(\mathbf{A}^{(\ell+1)}; \mathbf{\Theta}_{\ell+1}\big)$ and $P\big(\mathbf{A}^{(\ell)}; \mathbf{\Theta}_{\ell}\big)$ respectively. Let us first consider the case where the connection probability p_{i_ℓ, j_ℓ} does not depend on any dyadic factor d_{i_ℓ, j_ℓ}. In this case, the only functional form of $p_{i_{\ell+1}, j_{\ell+1}}$ compatible with Eq. (2.48) for every pair of $(\ell+1)$-nodes is such that

$$\ln \left[1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)\right] = -\delta \, g(x_{i_{\ell+1}}) \, g(x_{j_{\ell+1}}) \tag{2.49}$$

where $g(x)$ is a positive function such that

$$g(x_{i_{\ell+1}}) = \sum_{i_\ell \in i_{\ell+1}} g(x_{i_\ell}) \tag{2.50}$$

and δ is positive and ℓ-independent. The positivity of δ and $g(x)$ follows from the fact that, since $0 \leq p_{i_{\ell+1}, j_{\ell+1}}(\delta) \leq 1$, $\ln \left[1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)\right]$ has to be non-positive. On the other hand, $g(x)$ has to have the same sign for

all nodes, otherwise for some pair of nodes the product $g(x_{i_{\ell+1}})\, g(x_{j_{\ell+1}})$ will be negative. Interpreting $g(x)$ as the impact of the fitness x on the connection probability, it makes sense to choose the positive sign for $g(x)$ (and, incidentally, to assume that $g(x)$ is monotonically increasing with x). For similar reasons, δ has to be positive as well. Now, if the quantity x is node-additive (e.g. because it is identified with some empirical additive quantity, like the GDP in our model of the ITN), then the fitness of each $(\ell+1)$-node $x_{i_{\ell+1}}$ should be consistently obtained as a sum $\sum_{i_\ell \in i_{\ell+1}} x_{i_\ell}$ over the underlying ℓ-nodes. This implies that, after reabsorbing any (positive) proportionality factor into δ, the only possible choice for $g(x)$ in the additive case is $g(x) = x$. By contrast, if we do not require x being node-additive, we can always invoke the desired monotonicity of $g(x)$ and redefine $x \leftarrow g(x)$ (indeed, there is no *a priori* reason why x_{i_ℓ}, rather than $g(x_{i_\ell})$, should be regarded as the 'natural' node-specific factor affecting the connection probabilities involving i_ℓ). This makes the redefined fitness x additive by construction.

In summary, by redefining the node-specific factor x in a way that makes it node-additive, and reabsorbing any global constant into δ, the only possible functional form for p_{i_ℓ, j_ℓ} under the requirement of scale-invariance (and in the absence of dyadic factors) is such that

$$\ln\left[1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)\right] = -\delta\, x_{i_{\ell+1}} x_{j_{\ell+1}}, \tag{2.51}$$

for $i_{\ell+1} \neq j_{\ell+1}$, or equivalently

$$p_{i_\ell, j_\ell}(\delta) = 1 - e^{-\delta x_{i_\ell} x_{j_\ell}}, \quad \delta, x_{i_\ell}, x_{j_\ell} > 0, \quad i_\ell \neq j_\ell, \tag{2.52}$$

where δ is scale-invariant and $x_{i_{\ell+1}} = \sum_{i_\ell \in i_{\ell+1}} x_{i_\ell}$.

Now we consider the connection probability between a block $i_{\ell+1}$ and itself, i.e. the self-loop at the coarse-grained level. In this case, to avoid double counting the internal pairs of nodes, Eq. (2.48) should be replaced by the expression

$$\ln\left[1 - p_{i_{\ell+1}, i_{\ell+1}}(\delta)\right] = \sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in i_{\ell+1}, j_\ell \leq i_\ell} \ln\left[1 - p_{i_\ell, j_\ell}(\delta)\right]. \tag{2.53}$$

Now, by isolating the terms corresponding to self-loops in the quantity on the right hand side, we can rewrite the remaing terms as in Eq. (2.51)

and obtain:

$$\sum_{i_\ell \in i_{\ell+1}} \left[\sum_{j_\ell \in i_{\ell+1}, j_\ell < i_\ell} \ln\left[1 - p_{i_\ell, j_\ell}(\delta)\right] + \ln\left[1 - p_{i_\ell, i_\ell}(\delta)\right] \right]$$
$$= \sum_{i_\ell \in i_{\ell+1}} \left[\frac{1}{2} \sum_{j_\ell \in i_{\ell+1}, j_\ell \neq i_\ell} \ln\left[1 - p_{i_\ell, j_\ell}(\delta)\right] + \ln\left[1 - p_{i_\ell, i_\ell}(\delta)\right] \right] \tag{2.54}$$
$$= \sum_{i_\ell \in i_{\ell+1}} \left[-\frac{\delta}{2} \sum_{j_\ell \in i_{\ell+1}, j_\ell \neq i_\ell} x_{i_\ell} x_{j_\ell} + \ln\left[1 - p_{i_\ell, i_\ell}(\delta)\right] \right].$$

As argued above, the only solution for p_{i_ℓ, i_ℓ} compatible with the requirement $x_{i_{\ell+1}} = \sum_{i_\ell \in i_{\ell+1}} x_{i_\ell}$ involves a function $\tilde{g}(x_{i_\ell})$ such that $\tilde{g}(x_{i_{\ell+1}}) = \sum_{i_\ell \in i_{\ell+1}} \tilde{g}(x_{i_\ell})$. Take $\tilde{g}(x_{i_\ell}) = \sqrt{\eta}\, x_{i_\ell}$ for some $\eta > 0$. Then the requirement in Eq. (2.53) finally takes the form

$$\eta \sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in i_{\ell+1}} x_{i_\ell} x_{j_\ell} = \sum_{i_\ell \in i_{\ell+1}} \left[\frac{\delta}{2} \sum_{j_\ell \in i_{\ell+1}, j_\ell \neq i_\ell} x_{i_\ell} x_{j_\ell} + \eta x_{i_\ell}^2 \right] \tag{2.55}$$

where we have used

$$x_{i_{\ell+1}}^2 = \left(\sum_{i_\ell \in i_{\ell+1}} x_{i_\ell} \right)^2 = \sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in i_{\ell+1}} x_{i_\ell} x_{j_\ell}.$$

Clearly, the only possible solution for Eq. (2.55) is given by $\eta = \frac{\delta}{2}$, yielding:

$$p_{i_\ell, i_\ell} = 1 - e^{-\frac{\delta}{2} x_{i_\ell}^2}, \quad \delta, x_{i_\ell}, x_{j_\ell} > 0. \tag{2.56}$$

Taken together, Eq. (2.52) and (2.56) coincide with what stated in Eq. (2.4) when $f \equiv 1$, i.e. with Eq. (2.17).

If we add dyadic factors, i.e. if we allow p_{i_ℓ, j_ℓ} to additionally depend on some positive function $f(d)$ of the dyadic quantity d, while at the same time preserving the bilinear dependence of $\ln\left[1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)\right]$ on x_{i_ℓ} and x_{j_ℓ} (i.e. preserving the additivity of the fitness), then Eq. (2.51) has to be generalized to

$$\ln\left[1 - p_{i_{\ell+1}, j_{\ell+1}}(\delta)\right] = -\delta\, x_{i_{\ell+1}}\, x_{j_{\ell+1}}\, f(d_{i_{\ell+1}, j_{\ell+1}}) \tag{2.57}$$

for $i_{\ell+1} \neq j_{\ell+1}$ and

$$\ln\left[1 - p_{i_{\ell+1},i_{\ell+1}}(\delta)\right] = -\frac{\delta}{2}\, x_{i_{\ell+1}}\, x_{i_{\ell+1}}\, f(d_{i_{\ell+1},i_{\ell+1}}) \tag{2.58}$$

otherwise, where $f(d_{i_\ell,j_\ell})$ renormalizes as

$$x_{i_{\ell+1}}\, x_{j_{\ell+1}}\, f(d_{i_{\ell+1},j_{\ell+1}}) = \sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in j_{\ell+1}} x_{i_\ell} x_{j_\ell} f\big(d_{i_\ell,j_\ell}\big). \tag{2.59}$$

Equations (2.57), (2.58) and (2.59) coincide with Eqs. (2.4) and (2.7), thus completing our proof.

As a final remark, note that in principle the constant δ may be entirely reabsorbed into the fitness x (as mentioned above) or even into the function $f(d)$, however it is useful to keep it separate as a single parameter controlling the overall density of the graph. Also note that if the dyadic factor d is interpreted as a feature enhancing the connection probability (e.g. because it represents similarity, correlation, co-affiliation, etc.), then $f(d)$ has to be an increasing function. By contrast, if d suppresses the connection probability (e.g. because it represents distance or dissimilarity), then $f(d)$ has to be a decreasing function, as in our model of the ITN.

2.6.2 GDP, distance and Trade data

In our analysis of the ITN, the fundamental hierarchical level $\ell = 0$ is the one where each 0-node i_0 corresponds to a country in the world and the fitness x_{i_0} corresponds to the GDP of that country. Similarly, the distance d_{i_0,j_0} corresponds to the geographic distance between the two countries i_0 and j_0 and a realized link ($a_{i_0,j_0} = 1$) corresponds to the existence of a trade relation (in either direction) between i_0 and j_0.

GDP data are taken from the World Bank dataset [74] and are expressed in US Dollars. The results reported in Sec. 2.3 use data for year 2011. The number of countries for which GDP data are available in that year is 183. Note that, unlike the international trade data (see below), the World Bank GDP dataset covers a slightly smaller number of countries as it does not include very small ones (typically islands).

Geographic distance data are taken from the BACI-CEPII GeoDist database [75]. It reports bilateral inter-country distances measured as

population-based averages among the most populated pairs of cities across each pair of countries. The database uses the general formula

$$d_{i_0,j_0} = \left(\frac{\sum_{k \in i_0} \sum_{l \in j_0} \mathrm{POP}_k \mathrm{POP}_l \, d_{k.l}^{\theta}}{\sum_{k \in i_0} \sum_{l \in j_0} \mathrm{POP}_k \mathrm{POP}_l} \right)^{1/\theta} \tag{2.60}$$

developed by Head and Mayer [76] for calculating the distance d_{i_0,j_0} between country i_0 and country j_0 as a population-based average of the distances $d_{k,l}$ between pairs of internal agglomerations (cities, towns and places) across i_0 and j_0. The symbol $k \in i_0$ denotes that k runs over the agglomerations inside country i_0, and POP_k denotes the demographic population of agglomeration k. In the GeoDist database, population data were taken from the World Gazetteer (`https://www.world-gazetteer.com`) website. Note that $d_{i_0,i_0} > 0$, i.e. the 'distance' of a country to itself is non-zero (therefore it is not a proper metric distance). This is consistent with the fact that, at higher hierarchical levels, the distance between a block-node to itself is necessarily positive as a result of the renormalization rule. The exponent θ measures the sensitivity of trade flows to bilateral distance. As noted in the BACI-CEPII GeoDist documentation, selecting $\theta = -1$ corresponds to the usual coefficient estimated from gravity models of bilateral trade flows. Such a choice results in the calculation of d_{i_0,j_0} as a population-based average analogous to the GDP-based average used later in our own renormalization procedure when coarse-graining the network. The agreement between our model and the ITN data actually suggests that, for the study of international trade, a better definition of inter-country distances could presumably be obtained by replacing POP with GDP in the above formula, to make inter-country distances fully consistent with our GDP-averaged renormalized distances at higher levels. Unfortunately, GDP data at the agglomeration level are much more difficult to obtain than the corresponding population data. For this reason, we used population-averaged distances in our analysis at level $\ell = 0$, and their GDP-averaged renormalized values at higher levels $\ell > 0$. Given the pairwise geographical distances $\{d_{i_0,j_0}\}_{i_0,j_0=1}^{N_0}$ at level $\ell = 0$, we constructed the dendrogram of nested partitions $\{\mathbf{\Omega}_\ell\}_{\ell \geq 0}$ of world countries (shown in Fig. 10) using single-linkage hierarchical

clustering, which produces *subdominant ultrametric distances* $\{d^{<}_{i_0,j_0}\}_{i_0,j_0=1}^{N_0}$ as explained in Sec. 2.3. A straight cut in the dendrogram induces a hierarchical level ℓ and a corresponding partition of countries into ℓ-countries. The renormalized GDPs and distances are then calculated using Eqs. (2.18) and (2.19) (using the original distances).

For the construction of the International Trade Network, we used the BACI-Comtrade dataset [67]. The dataset reports the international trade flows between 207 countries for the years 2008 to 2011. From the full set of countries, we selected the 183 countries for which we could find matching GDP data in the World Bank database (as explained above). In the BACI-Comtrade dataset, trade is disaggregated into 96 commodity classes labeled at a 2-digit resolution level and is expressed in thousands of dollars. The database is the result of an adjustment procedure [67] which reconciles unbalanced trade values as reported by importers and exporters. For the purpose of this study, we first merged the disaggregated data into a unique aggregate undirected network, where the monetary flows between countries is the total trade (both import and export) in all the 96 commodities, and then considered its binary (i.e. unweighted) projection. Therefore, a binary link in the 0-graph of the ITN is present if the two countries at its endpoints have a positive trade (either import or export) in any commodity, consistently with similar analyses of the topology of the ITN constructed from different datasets [68–71]. This procedure defines the empirical adjacency matrix $\tilde{\mathbf{A}}^{(0)}$ of the 0-graph of the ITN. The empirical matrices $\tilde{\mathbf{A}}^{(\ell)}$ for $\ell > 0$ are obtained via coarse-graining the empirical 0-graph (following the general procedure illustrated in Fig. 1) using the nested partitions $\{\mathbf{\Omega}_\ell\}_{\ell \geq 0}$ induced by the dendrogram in Fig. 10.

2.6.3 Network properties: empirical and expected values

Here we define the key topological properties considered in our analysis and modeling of the ITN.

Each such property is a function $Y(\mathbf{A}^{(\ell)})$ of the $N_\ell \times N_\ell$ adjacency matrix $\mathbf{A}^{(\ell)}$ (with entries $a^{(\ell)}_{i_\ell,j_\ell} = 0, 1$) of the generic ℓ-graph. Note that this matrix is symmetric and can contain non-zero entries along the diagonal,

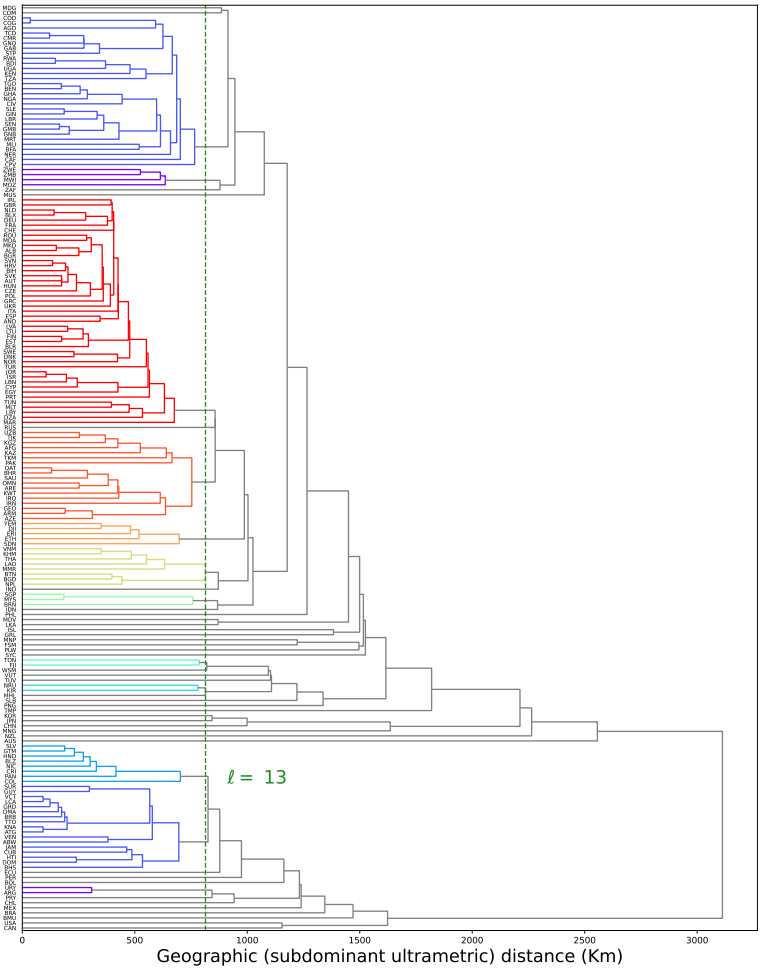

Figure 10: Dendrogram of world countries from their geographical distances using single-linkage hierarchical clustering. The dendrogram can be used to produce any desired sequence $\{\mathbf{\Omega}_\ell\}_{\ell\geq 0}$ of geographically nested partitions, via either single-scale (straight) or multi-scale (non-straight, but monophyletic) 'cuts' as illustrated in Fig. 2. In our analysis we considered 18 straight cuts at various ultrametric distances $\{h_\ell\}_{\ell=0}^{17}$ (with $h_0 = 0$) producing a hierarchy $\{\mathbf{\Omega}_\ell\}_{\ell=0}^{17}$ of 18 partitions and a corresponding sequence of block-countries with $N_0 = 183$ and $N_\ell = 180 - 10\ell$ for $\ell = 1, 17$. For instance, a cut at level $\ell = 13$ (dashed line) yields 50 block-countries that correspond to the 50 branches drawn in different colors.

representing self-loops. These self-loops may or may not be present in the 0-graph, but are in any case eventually generated by the coarse graining procedure if the nodes mapped onto the same block-node are connected among themselves.
When analysing the ITN, the relevant matrix $\mathbf{A}^{(\ell)}$ is the empirical matrix $\tilde{\mathbf{A}}^{(\ell)}$ obtained at the hierarchical level ℓ from the BACI-Comtrade data in year 2011 as described above. The corresponding empirical value of each topological property Y of interest will be denoted as $\tilde{Y} \equiv Y(\tilde{\mathbf{A}}^{(\ell)})$. When considering the multiscale model, $\mathbf{A}^{(\ell)}$ is instead a random (symmetric) matrix whose entries $\{a^{(\ell)}_{i_\ell,j_\ell}\}$ are Bernoulli random variables with expected value

$$\langle a^{(\ell)}_{i_\ell,j_\ell}\rangle \equiv p_{i_\ell j_\ell}(\delta) = \begin{cases} 1 - e^{-\delta \frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}}} & \text{if} \quad i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2}\frac{\mathrm{GDP}^2_{i_\ell}}{d_{i_\ell,i_\ell}}} & \text{if} \quad i_\ell = j_\ell \end{cases} \tag{2.61}$$

where, consistently with the possible presence of self-loops, we allow for $i_\ell = j_\ell$.
Equation (2.61) allows us to calculate the expected value of several topological properties. For instance, the total number of ℓ-links (including possible self-loops) at level ℓ is given by

$$L_\ell(\mathbf{A}^{(\ell)}) = \sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1}^{i_\ell} a^{(\ell)}_{i_\ell,j_\ell}. \tag{2.62}$$

Before considering other properties, we note that we fix the only free parameter δ to the unique value $\tilde{\delta}$ such that the expected number

$$\langle L_0\rangle = \sum_{i_0=1}^{N_0}\sum_{j_0=1}^{i_0} p_{i_0,j_0}(\delta) \tag{2.63}$$

of links of the 0-graph equals the empirical value

$$\tilde{L}_0 = L_0(\tilde{\mathbf{A}}^{(0)}) = \sum_{i_0=1}^{N_0}\sum_{j_0=1}^{i_0} \tilde{a}^{(0)}_{i_0,j_0} = 12018 \tag{2.64}$$

observed in the ITN in year 2011. This selects the value $\tilde{\delta} = 3.6 \cdot 10^{-17}(\mathrm{USD})^{-2}$, where USD stands for US dollars (the unit of measure used in GDP data). Having fixed $\tilde{\delta}$, we can generate unbiased realisations $\{\mathbf{A}^{(\ell)}\}$ of the ℓ-graphs from the multiscale model at any desired hierarchical level ℓ by sampling ℓ-links independently with probability $\tilde{p}_{i_\ell,j_\ell} \equiv p_{i_\ell,j_\ell}(\tilde{\delta})$. By averaging the value $Y(\mathbf{A}^{(\ell)})$ of any topological property of interest over such realizations, we can efficiently estimate the corresponding expected value

$$\langle Y \rangle \equiv \sum_{\mathbf{A}^{(\ell)} \in \mathcal{G}_{N_\ell}} P\big(\mathbf{A}^{(\ell)}; \tilde{\delta}\big) Y(\mathbf{A}^{(\ell)}), \tag{2.65}$$

where $P\big(\mathbf{A}^{(\ell)}; \delta\big)$ is given by Eq. (2.8), without actually calculating the above sum explicitly. If $Y(\mathbf{A}^{(\ell)})$ is linear in $\mathbf{A}^{(\ell)}$, we can even calculate $\langle Y \rangle$ exactly by directly replacing $a^{(\ell)}_{i_\ell,j_\ell}$ with $\tilde{p}_{i_\ell,j_\ell}$ in the definition of $Y(\mathbf{A}^{(\ell)})$, without sampling any graph at all. This is indeed the case for the number of links in Eq. (2.62).

Given any ℓ-graph $\mathbf{A}^{(\ell)}$ (be it the empirical ℓ-graph or a random realization from the model), the main topological properties of interest to us are: the *link density*

$$D_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{2L_\ell(\mathbf{A}^{(\ell)})}{N_\ell(N_\ell+1)} = \frac{2\sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1}^{i_\ell} a^{(\ell)}_{i_\ell,j_\ell}}{N_\ell(N_\ell+1)} \tag{2.66}$$

(representing the ratio of realized to maximum number of links, including possible self-loops), the *degree*

$$k_{i_\ell}(\mathbf{A}^{(\ell)}) \equiv \sum_{j_\ell \neq i_\ell} a^{(\ell)}_{i_\ell,j_\ell} \tag{2.67}$$

(counting the number of links of the ℓ-node i_ℓ, excluding self-loops), the *rescaled degree*

$$\kappa_{i_\ell}(\mathbf{A}^{(\ell)}) \equiv \frac{1}{N_\ell - 1} \sum_{j_\ell \neq i_\ell} a^{(\ell)}_{i_\ell,j_\ell} \tag{2.68}$$

(which ranges in $[0,1]$, irrespective of the vertex and hierarchical level considered), the *average nearest neighbour degree* [65]

$$k^{nn}_{i_\ell}(\mathbf{A}^{(\ell)}) \equiv \frac{\sum_{j_\ell \neq i_\ell}\sum_{k_\ell \neq j_\ell} a^{(\ell)}_{i_\ell,j_\ell} a^{(\ell)}_{j_\ell,k_\ell}}{\sum_{j_\ell \neq i_\ell} a^{(\ell)}_{i_\ell,j_\ell}} \tag{2.69}$$

(representing the average degree of the neighbours of i_ℓ), and finally the local clustering coefficient [10]

$$c_{i_\ell}(\mathbf{A}^{(\ell)}) \equiv \frac{\sum_{j_\ell \neq i_\ell} \sum_{k_\ell \neq i_\ell, j_\ell} a^{(\ell)}_{i_\ell, j_\ell} a^{(\ell)}_{j_\ell, k_\ell} a^{(\ell)}_{k_\ell, i_\ell}}{\sum_{j_\ell \neq i_\ell} \sum_{k_\ell \neq i_\ell, j_\ell} a^{(\ell)}_{i_\ell, j_\ell} a^{(\ell)}_{k_\ell, i_\ell}} \tag{2.70}$$

(representing the number of triangles into which i_ℓ partipates, divided by the maximum realizable number of triangles, given the value of k_{i_ℓ}). All the above quantities can be averaged over nodes to obtain the following overall properties:

$$\bar{k}_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{1}{N_\ell} \sum_{i_\ell=1}^{N_\ell} k_{i_\ell}(\mathbf{A}^{(\ell)}), \tag{2.71}$$

$$\bar{\kappa}_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{1}{N_\ell} \sum_{i_\ell=1}^{N_\ell} \kappa_{i_\ell}(\mathbf{A}^{(\ell)}), \tag{2.72}$$

$$\bar{k}^{nn}_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{1}{N_\ell} \sum_{i_\ell=1}^{N_\ell} k^{nn}_{i_\ell}(\mathbf{A}^{(\ell)}), \tag{2.73}$$

$$\bar{c}_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{1}{N_\ell} \sum_{i_\ell=1}^{N_\ell} c_{i_\ell}(\mathbf{A}^{(\ell)}). \tag{2.74}$$

Note that $\bar{\kappa}_\ell(\mathbf{A}^{(\ell)}) \in [0, 1]$ in Eq. (2.72) coincides with the link density *excluding self-loops*, representative an alternative to the definition of density in Eq. (2.66) (where self-loops are included). Besides the average local clustering coefficient $\bar{c}_\ell(\mathbf{A}^{(\ell)})$, it is possible to define the *global clustering coefficient* [94–96]

$$c^{\text{global}}_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{\Delta_\ell(\mathbf{A}^{(\ell)})}{\Lambda_\ell(\mathbf{A}^{(\ell)})} \tag{2.75}$$

where

$$\Delta_\ell(\mathbf{A}^{(\ell)}) \equiv \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} \sum_{k_\ell \neq i_\ell, j_\ell} a^{(\ell)}_{i_\ell, j_\ell} a^{(\ell)}_{j_\ell, k_\ell} a^{(\ell)}_{k_\ell, i_\ell} \tag{2.76}$$

is the overall number of *realized* ('closed') triangles and

$$\Lambda_\ell(\mathbf{A}^{(\ell)}) \equiv \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} \sum_{k_\ell \neq i_\ell, j_\ell} a^{(\ell)}_{i_\ell, j_\ell} a^{(\ell)}_{k_\ell, i_\ell} \tag{2.77}$$

is the number of (Λ-shaped) *wedges*, i.e. *potential* (both 'open' and 'closed') triangles (note that each realized triangle is counted three times by both Δ_ℓ and Λ_ℓ).

It is important to stress that, of all the quantities defined in Eqs. (2.66)-(2.74) for each ℓ-node ($i_\ell = 1, N_\ell$) and/or all levels ($\ell = 0, 17$), only the overall density D_0 of the 0-graph is replicated by construction via the parameter choice $\delta = \tilde{\delta}$: indeed, having enforced $\langle L_0 \rangle = \tilde{L}_0$ by equating Eqs. (2.63) and (2.64) coincides with having required $\langle D_0 \rangle = \tilde{D}_0$. For all the other properties, including D_ℓ for all $\ell > 0$, the agreement between the model and the empirical network is highly nontrivial and hence notable.

2.6.4 The scale-free range with universal inverse square exponent

We first consider the case $\alpha = 1/2$ and rewrite the distribution of the reduced degree κ shown in Eq. (2.43) as

$$Q_\ell(\kappa) = A_\ell(\kappa) B_\ell(\kappa) \tag{2.78}$$

where

$$A_\ell(\kappa) \equiv \exp\left[\frac{-\delta\gamma_\ell^2}{\ln^2(1-\kappa)}\right], \tag{2.79}$$

$$B_\ell(\kappa) \equiv \frac{2\sqrt{\delta\gamma_\ell^2/\pi}}{(1-\kappa)\ln^2(1-\kappa)}. \tag{2.80}$$

The term $A_\ell(\kappa)$ is a lower cut-off that rapidly saturates to 1 as κ increases (see Fig. 11). On the other hand, $B_\ell(\kappa)$ has an intermediate power-law regime (for values of κ not too close to 1) and an upper cut-off (for κ closer to 1). This behaviour can be understood by using the expansion of $\ln(1-y) = -\sum_{n=1}^{\infty}(-y)^n/n = y + R(y)$ for $|y| < 1$, where $R(y) \equiv$

$-\sum_{n=2}^{\infty}(-y)^n/n$:

$$\begin{aligned} B_\ell(\kappa) &= \frac{2\gamma_\ell\sqrt{\delta/\pi}}{(1-\kappa)\ln^2(1-\kappa)} \\ &= \frac{2\gamma_\ell\sqrt{\delta/\pi}}{(1-\kappa)[\kappa+R(\kappa)]^2} \\ &\approx \begin{cases} \frac{2\gamma_\ell\sqrt{\delta/\pi}}{\kappa^2} & \kappa \ll 1 \\ +\infty & \kappa \to 1^- \end{cases} \\ &\approx \kappa^{-2} C_\ell(\kappa) \end{aligned} \tag{2.81}$$

where $C_\ell(\kappa)$ is a cut-off function being equal to $2\gamma_\ell\sqrt{\delta/\pi}$ for $\kappa \ll 1$ and diverging when $\kappa \to 1^-$. This is confirmed in Fig. 11. Putting the pieces together, the right tail of the reduced degree distribution behaves as

$$Q_\ell(\kappa) \approx \kappa^{-2} C_\ell(\kappa) \tag{2.82}$$

where $C_\ell(\kappa)$ is the cut-off function. This proves our statement in Eq. (2.44), and is confirmed by the numerical simulations in Fig. 7.

Now we can partly extend the above results to the general case $\alpha \in (0,1)$ using the following argument. We note from Eq. (2.34) that, for any $\alpha \in (0,1)$, the expected degree is uniquely determined by the LT of the fitness distribution. Even if the explicit form of $\rho_\ell(x;\alpha,\gamma_\ell)$ is not known for $\alpha \neq 1/2$ (apart from expressions involving integral representations [80, 83, 84]), the LT is known and given by Eq. (2.27). Using that formula, thereby selecting without loss of generality the value $\gamma_\alpha \equiv [\cos(\alpha\pi/2)]^{1/\alpha}$, we see that Eq. (2.34) can be rewritten as

$$\begin{aligned} k_\ell(x) &= (N_\ell - 1)\left(1 - \lambda_\ell(\delta x;\alpha,\gamma_\alpha)\right) \\ &= (N_\ell - 1)\left(1 - e^{-(\delta x)^\alpha}\right). \end{aligned} \tag{2.83}$$

Indeed, for $\alpha = 1/2$ and $\gamma_{1/2} = [\cos(\pi/4)]^2 = 1/2$, the above equation reduces exactly to Eq. (2.38). In complete analogy with the case $\alpha = 1/2$, Eq. (2.83) implies that, for small values of x, the expected degree behaves as

$$k_\ell(x) \propto x^\alpha \qquad (x \ll \delta^{-1}), \tag{2.84}$$

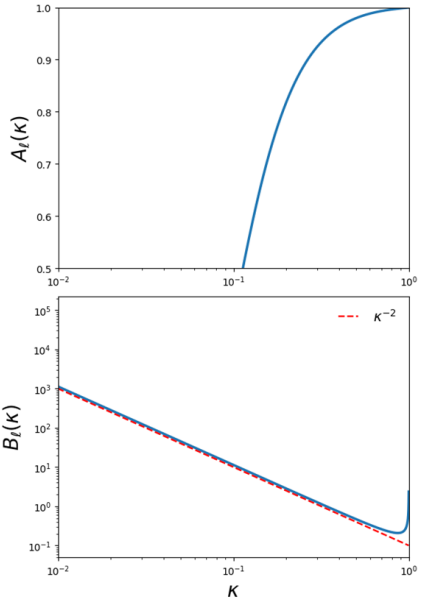

Figure 11: The two factors contributing to the cumulative distribution of the rescaled degree. Top: lower cut-off function $A_\ell(\kappa)$ defined in Eq. (2.79). The function rapidly saturates to $A_\ell(\kappa) \approx 1$ as the rescaled degree κ increases. Bottom: tail function $B_\ell(\kappa)$ defined in Eq. (2.80). The function behaves as a power law $B_\ell(\kappa) \approx \kappa^{-2}$ (red dashed line) for a wide range of κ and has an ℓ-dependent upper cut-off corresponding to nodes whose rescaled degree saturates to 1.

while for large values of x there is a saturation $k_\ell(x) \approx N_\ell - 1$ (as in Fig. 6) which produces the cut-off in the degree distribution $P_\ell(k)$. Therefore, in order to establish the behaviour of $P_\ell(k)$ before the cut-off appears (i.e. for $k \ll N_\ell - 1$), it is enough to invert Eq. (2.84) as $x_\ell(k) \propto k^{1/\alpha}$ and use it into Eq. (2.41) to obtain

$$\begin{aligned} P_\ell(k) &= \rho_\ell\big(x_\ell(k);\alpha,\gamma_\alpha\big)\frac{\mathrm{d}}{\mathrm{d}k}x_\ell(k) \\ &\propto \big(x_\ell(k)\big)^{-1-\alpha}k^{-1+1/\alpha} \\ &\propto k^{-1-1/\alpha}k^{-1+1/\alpha} \\ &\propto k^{-2} \qquad\qquad (k \ll N_\ell - 1) \end{aligned} \tag{2.85}$$

where we have used $\rho_\ell(x;\alpha,\gamma_\alpha) \propto x^{-1-\alpha}$ for large enough x. As clear from Eq. (2.84), the range of values of x for which both $x_\ell(k) \propto k^{1/\alpha}$ and $\rho_\ell(x;\alpha,\gamma_\alpha) \propto x^{-1-\alpha}$ are valid is larger when δ is smaller (correspondingly, the effect of the cut-off in the degree distribution is weaker). So for sparser networks the regime $P_\ell(k) \propto k^{-2}$ is valid for a larger fraction of the range of values of k. Correspondingly, the reduced degree distribution behaves as

$$Q_\ell(\kappa) \propto \kappa^{-2} \qquad (\kappa \ll 1) \tag{2.86}$$

and is followed by an upper cut-off for $\kappa \lesssim 1$. For sparser networks, Eq. (2.86) is valid for a larger range of values. The above results confirm Eq. (2.82), which was obtained for $\alpha = 1/2$, and extend it to the entire range $\alpha \in (0,1)$.

Chapter 3

Generalization of the SIM to directed networks: scale-invariance and reciprocity

In this chapter, based on the work [3] by M. Lalli and D. Garlaschelli, we focus on the possibility of enlarging our discussion to the family of directed graphs, i.e. networks with directed edges. In this case, a non-trivial reformulation of the SIM is needed to embed the additional information conveyed by asymmetric interactions, and the main objective of the following chapter is to provide such a framework.

3.1 Introduction

The original formulation of the SIM is restricted to undirected graphs, while several real-world networks, including economic [99–102], social [103, 104], biological [105] and material flow [106] networks are intrinsically directed. In general, moving from undirected to directed networks is nontrivial, because of the pervasive property of *reciprocity*, which refers to the non-random tendency of pairs of nodes to form mu-

tual connections [105, 107]. By varying the degree of reciprocity, one can range from undirected (perfectly reciprocal) through random (areciprocal) to totally asymmetric (perfectly antireciprocal) networks [105, 107]. Reciprocity has been found to classify directed real-world networks into consistent classes [105, 108], affect the abundances of directed triads [109–111], determine the spectral properties of adjacency matrices [112] and significantly impact processes taking place on networks [113–117].

In light of the aforementioned results, here we extend the SIM to the realm of directed networks with reciprocity, which include the undirected case as an extreme one.
It should be noted here that other approaches to network renormalization (discussed in the previous chapter of this thesis) do not easily lend themselves to directed networks. For instance, the geometric embedding method [38] is based on necessarily symmetric distances that, *per se*, cannot naturally explain the asymmetry of directed networks (indeed, directed geometric models require additional and more sophisticated ingredients [118, 119]). Similarly, the spectral method [42] requires symmetric adjacency matrices to define a Hermitian Laplacian operator. Importantly, we find that a naive reformulation of the original SIM, which merely makes the connection probabilities asymmetrical by using two features per node, fails in reproducing various topological and spectral properties of real-world networks, precisely because it does not capture the empirical patterns of reciprocity. On the contrary, a non-trivial extension can replicate reciprocity and the resulting properties.

The chapter is organized as follows. In Sec. 3.2 we introduce the Directed Scale-Invariant Model (DSIM). In Sec. 3.3, as a pedagogical benchmark, we present the simplest, completely homogeneous example of the model where all nodes are assumed to be statistically equivalent. Then, following the approach adopted in Chap. 2, we distinguish between a quenched scenario, where node features are treated as deterministic, and an annealed scenario, where they are interpreted as random variables. In the quenched scenario, the SIM can be used as a model of empirical networks where node features are observable and expected to determine

the topology. Then, unlike models that are not designed to remain consistent across different resolution levels, the SIM can make predictions about network properties observed at any scale. We illustrate this procedure in Sec. 3.4, where we provide a multiscale description of the international trade network that also captures its strongly reciprocal nature. In the annealed scenario, the model can be used to generate random networks built from the principle of scale invariance. We illustrate this approach in Sec. 3.5. Complementary details on data and calculations are provided in Sec. 3.7.

3.2 Construction of the Directed Scale-Invariant Model

In this section, we define the DSIM as a model of random directed graphs, with nontrivial reciprocity, characterized by the property that the connection probability between (blocks of) nodes has always the same functional form, irrespective of the coarse-graining (resolution level) adopted. As stressed in the previous chapter, this means that such probability depends on the chosen aggregation of nodes only through its parameters, which will obey appropriate renormalization rules. It is also worth stressing again that the coarse-graining can be arbitrarily heterogeneous and include *multi-scale* aggregations where certain blocks may be deliberately 'small' (and even coincide with the original microscopic nodes themselves) and other blocks may be very 'large' (i.e., containing several microscopic nodes). As it is clear, this graph model is a generalization of the undirected model introduced in Chap. 2 to which it will reduce in the extreme case of complete reciprocity, i.e. when all links are reciprocated (so that the network is effectively undirected).

Basic quantities and definitions. We consider a binary directed graph with N_0 'microscopic' nodes (labelled by the first N_0 positive integers $i_0 = 1, \dots, N_0$) and its $N_0 \times N_0$ (Boolean and, in general, asymmetric) adjacency matrix $\mathbf{A}^{(0)}$ with entries $a_{i_0 j_0} = 1$ if a directed link from node

i_0 to node j_0 exists, and $a_{i_0 j_0} = 0$ otherwise. Self-loops ($a_{i_\ell i_\ell} = 1$) are allowed. As is customary, we call this microscopic graph the 0-graph, and its N_0 nodes the 0-nodes. First, given an *arbitrary* surjective and non-overlapping partition Ω_0 of the original N_0 0-nodes onto $N_1 < N_0$ (coarser) nodes, labelled as $i_1 = 1, \ldots, N_1$, we define the coarse-grained directed graph (this graph will be called the 1-graph, and its N_1 nodes the 1-nodes) with $N_1 \times N_1$ adjacency matrix $\mathbf{A}^{(1)}$ as follows. A directed link from the 1-node i_1 to the 1-node j_1 is present if and only if, in the 0-graph, there is at least one directed link from any $i_0 \in i_1$ (i.e. from any of the 0-nodes 'inside' i_1) to any $j_0 \in j_1$, as illustrated in Fig. 12. This coarse-graining step can then be iterated to identify an arbitrary sequence of coarse-grained ℓ-graphs (each with a certain number N_ℓ of ℓ-nodes), by introducing any desired hierarchy $\{\Omega_\ell\}_{\ell \geq 0}$ of nested (surjective and non-overlapping) partitions. Each such ℓ-graph will be uniquely associated to an $N_\ell \times N_\ell$ Boolean (and in general still asymmetric) adjacency matrix $\mathbf{A}^{(\ell)}$, whose entries obey the following relationship:

$$a^{(\ell+1)}_{i_{\ell+1} j_{\ell+1}} = 1 - \prod_{i_\ell \in i_{\ell+1}} \prod_{j_\ell \in j_{\ell+1}} (1 - a^{(\ell)}_{i_\ell j_\ell}) \qquad \forall\, i_{\ell+1}, j_{\ell+1}. \tag{3.1}$$

Now, we consider a random graph model producing a specific realization $\mathbf{A}^{(\ell)}$ of the ℓ-graph with probability $P_\ell\big(\mathbf{A}^{(\ell)}; \mathbf{\Theta}_\ell\big)$, where $\mathbf{\Theta}_\ell$ is the set of all parameters of the model. We look for the functional form of this probability by imposing that, given a partition Ω_ℓ of the N_ℓ ℓ-nodes, the induced probability $P_{\ell+1}\big(\mathbf{A}^{(\ell+1)}; \mathbf{\Theta}_{\ell+1}\big)$ at the next level $\ell + 1$ has the same functional form as $P_\ell\big(\mathbf{A}^{(\ell)}; \mathbf{\Theta}_\ell\big)$, with appropriately renormalized parameters $\mathbf{\Theta}_{\ell+1}$ (as a result, we will drop the subscript ℓ from the P_ℓ that realizes the scale-invariance requirement). For simplicity, we restrict ourselves to models with *independent dyads* where the graph probability (for $\ell = 0$, and consequently for all $\ell > 0$) factorizes into dyadic connection probabilities over pairs of nodes, plus self-loop probabilities over single nodes. Adopting the formalism introduced in [107] and reprised in [54, 110, 111, 120, 121], if $\langle \cdot \rangle$ denotes an expected value with respect to the (yet to be determined) scale-invariant distribution $P\big(\mathbf{A}^{(\ell)}; \mathbf{\Theta}_\ell\big)$, then the probabilities of the four possible dyads between two nodes i_ℓ and j_ℓ are:

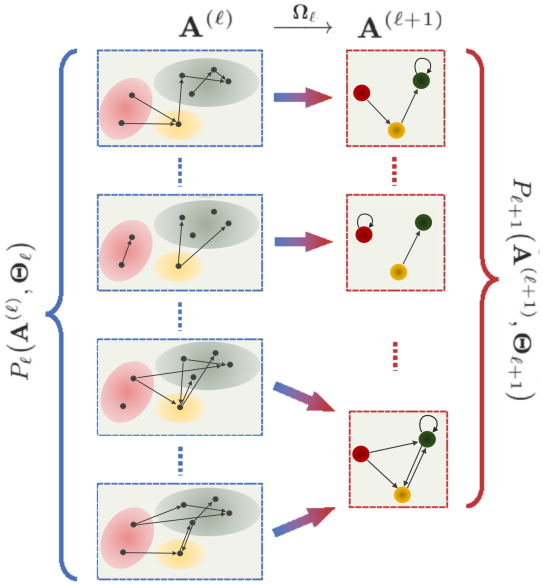

Figure 12: Schematic example of the graph coarse-graining and induced ensembles in the directed case. Given a probability distribution $P_\ell(\mathbf{A}^{(\ell)}; \mathbf{\Theta}_\ell)$ of graphs with adjacency matrix $\mathbf{A}^{(\ell)}$ (left), a given node partition $\mathbf{\Omega}_\ell$ is used to map sets of nodes onto 'block-nodes' of the resulting coarse-grained graphs with adjacency matrix $\mathbf{A}^{(\ell+1)}$ (right). A directed edge from $i_{\ell+1}$ to $j_{\ell+1}$ is drawn if an edge from i_ℓ to j_ℓ is present, for any $i_\ell \in i_{\ell+1}, j_\ell \in j_{\ell+1}$. This coarse-graining will induce a new probability distribution $P_{\ell+1}(\mathbf{A}^{(\ell+1)}; \mathbf{\Theta}_{\ell+1})$. Note that such a scheme is a straightforward generalisation of the one illustrated in Fig. 1 to the directed case, where the crucial difference resides in whether single or reciprocated links are created. As in the undirected case, the graph at level ℓ may end up in the same realization of the graph at level $\ell + 1$.

- $p^{\rightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) = \langle a^{\rightarrow}_{i_\ell j_\ell} \rangle$ with $a^{\rightarrow}_{i_\ell j_\ell} \equiv a^{(\ell)}_{i_\ell j_\ell}(1 - a^{(\ell)}_{j_\ell i_\ell})$ for the joint probability of a single directed link from i_ℓ to j_ℓ and no reciprocal link from j_ℓ to i_ℓ;

- $p^{\leftarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) = \langle a^{\leftarrow}_{i_\ell j_\ell} \rangle$ with $a^{\leftarrow}_{i_\ell j_\ell} \equiv a^{(\ell)}_{j_\ell i_\ell}(1 - a^{(\ell)}_{i_\ell j_\ell})$ for the joint probability of a single directed link from j_ℓ to i_ℓ and no reciprocal link from i_ℓ to j_ℓ;

- $p^{\leftrightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) = \langle a^{\leftrightarrow}_{i_\ell j_\ell} \rangle$ with $a^{\leftrightarrow}_{i_\ell j_\ell} \equiv a^{(\ell)}_{i_\ell j_\ell} a^{(\ell)}_{j_\ell i_\ell}$ for the joint probability of two reciprocal links between i_ℓ and j_ℓ;

- $p^{\nleftrightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) = \langle a^{\nleftrightarrow}_{i_\ell j_\ell} \rangle$ with $a^{\nleftrightarrow}_{i_\ell j_\ell} \equiv (1 - a^{(\ell)}_{i_\ell j_\ell})(1 - a^{(\ell)}_{j_\ell i_\ell})$ for the joint probability of no links between i_ℓ and j_ℓ.

Clearly, $p^{\rightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) + p^{\leftarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) + p^{\leftrightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) + p^{\nleftrightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) = 1$.
Note that, in order to generate reciprocity, i.e. the nontrivial occurrence of mutual links between the same two nodes, we do not assume that the joint probabilities further factorize over the marginal probabilities $p_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) \equiv \langle a^{(\ell)}_{i_\ell j_\ell} \rangle = p^{\rightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell) + p^{\leftrightarrow}_{i_\ell j_\ell}(\mathbf{\Theta}_\ell)$ for individual edges.
For self-loops, the only relevant probability is $p_{i_\ell i_\ell}(\mathbf{\Theta}_\ell) \equiv \langle a^{(\ell)}_{i_\ell i_\ell} \rangle = p^{\leftrightarrow}_{i_\ell i_\ell}(\mathbf{\Theta}_\ell) = 1 - p^{\nleftrightarrow}_{i_\ell i_\ell}(\mathbf{\Theta}_\ell)$ since $p^{\rightarrow}_{i_\ell i_\ell}(\mathbf{\Theta}_\ell) = p^{\leftarrow}_{i_\ell i_\ell}(\mathbf{\Theta}_\ell) = 0$ (self-loops necessarily reciprocate).

Before looking for the scale-invariant functional form of the above probabilities, we assume that the latter may depend on *global* (scalar) parameters η, δ, σ (determining, as we will see, the overall numbers of directed links, reciprocated links and self-loops respectively), *node-specific* (N_ℓ-dimensional) features $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{y_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{z_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{w_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ called *fitness* values (separately determining the intrinsic tendency of individual nodes of forming reciprocated pairs of links, unreciprocated out-going links, unreciprocated in-coming links and self-loops), and finally *dyadic* factors parametrized by an $N_\ell \times N_\ell$ matrix $\mathbf{D}^{(\ell)}$ (determining pairwise connection preferences not attributable to global or node-specific ones, e.g. similarities, distances, membership to communities, etc.). Having clarified our parametrization, and to ease the notation, from now on we omit the dependence on $\mathbf{\Theta}_\ell$ in all quantities.

3.2.1 Scale-invariant connection probabilities.

We can now look for the scale-invariant form of all probabilities. First, we demand that the undirected projection of our directed model reduces to the SIM considered in Chap. 2 for undirected graphs. Indeed, without this constraint, the directed model we are building here would violate scale-invariance when projected onto its undirected representation.

We can therefore constrain the functional form of the marginal probability $q_{i_\ell j_\ell}$ of nodes i_ℓ and j_ℓ being connected by at least one link, whatever the direction, to have the form already derived in Chap. 2. In formulae, this implies that the quantity defined by

$$q_{i_\ell j_\ell} \equiv p^{\rightarrow}_{i_\ell j_\ell} + p^{\leftarrow}_{i_\ell j_\ell} + p^{\leftrightarrow}_{i_\ell j_\ell} = 1 - p^{\nleftrightarrow}_{i_\ell j_\ell} \tag{3.2}$$

is given by

$$q_{i_\ell j_\ell} = \begin{cases} 1 - e^{-\delta x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})} & i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2} x_{i_\ell}^2 f(d_{i_\ell i_\ell}) - \sigma w_{i_\ell}} & i_\ell = j_\ell \end{cases} \tag{3.3}$$

where δ, $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{w_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{d_{i_\ell j_\ell}\}_{i_\ell,j_\ell=1}^{N_\ell}$ are all positive and f is a positive monotonic function. The global parameter δ tunes the overall density of *undirected* links, the node-specific fitness x_{i_ℓ} controls the tendency of node i_ℓ of forming undirected (i.e. of any directionality) links, and f incorporates the (optional) effects of the dyadic properties $\{d_{i_\ell j_\ell}\}_{i_\ell,j_\ell=1}^{N_\ell}$. Note that, as an addition to the undirected SIM, we have introduced an extra fitness w_{i_ℓ} (coupled to another global parameter σ) separately controlling for the tendency of i_ℓ of forming a self-loop more (if $\sigma > 0$) or less (if $\sigma < 0$) likely than otherwise determined by x_{i_ℓ} (note that self-loops are intrinsically reciprocated). To guarantee a positive $q_{i_\ell i_\ell}$, one must require

$$\sigma \geq -\min_{i_\ell} \left\{ \frac{\delta x_{i_\ell}^2 f(d_{i_\ell i_\ell})}{2 w_{i_\ell}} \right\}. \tag{3.4}$$

Also, to preserve the consistency between the directed network and its undirected projection, we need the matrix $\mathbf{D}^{(\ell)}$ to be symmetric ($d_{i_\ell j_\ell} = d_{j_\ell i_\ell}$) as in the undirected scenario.

Let us recall here that the key property of the probability $q_{i_\ell j_\ell}$ in Eq. (3.3) is that it preserves its functional form under arbitrary aggregations: if the coarse-grained network at the next level $\ell + 1$ is considered, then the probability $q_{i_{\ell+1} j_{\ell+1}}$ that the two $(\ell+1)$-nodes $i_{\ell+1}$ and $j_{\ell+1}$ are connected by a link (i.e. that any pair of their constituent ℓ-nodes are connected) is given by exactly the same expression, with parameters δ, σ remaining

unchanged and the other ones renormalizing as

$$x_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} x_{i_\ell}, \quad w_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} w_{i_\ell}, \tag{3.5}$$

$$f(d_{i_{\ell+1}j_{\ell+1}}) \equiv \frac{\sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in j_{\ell+1}} x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})}{\sum_{i_\ell \in i_{\ell+1}} x_{i_\ell} \sum_{j_\ell \in j_{\ell+1}} x_{j_\ell}}. \tag{3.6}$$

Note that (3.6) holds also for self-distances ($i_{\ell+1} = j_{\ell+1}$).

Next, we consider the other marginal, but 'directed', probability $p_{i_\ell j_\ell}$ introduced above, representing the probability of a directed edge being present from i_ℓ to j_ℓ, irrespective of the presence of the edge in the opposite direction. By demanding that $p_{i_\ell j_\ell}$ fulfils the same scale-invariance requirement that in the undirected case leads to the functional form (3.3), we arrive at the expression

$$p_{i_\ell j_\ell} = \begin{cases} 1 - e^{-\eta y_{i_\ell} z_{j_\ell} f(d_{i_\ell j_\ell})} & i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2} x_{i_\ell}^2 f(d_{i_\ell i_\ell}) - \sigma w_{i_\ell}} & i_\ell = j_\ell \end{cases} \tag{3.7}$$

where $\eta > 0$ tunes the overall density of directed links and, because of the possible asymmetry of $p_{i_\ell j_\ell}$ for $i_\ell \neq j_\ell$, we have introduced *two* sets of fitness variables $\{y_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ and $\{z_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ (representing the intrinsic tendency of each node of establishing out-going and in-coming links, respectively), while for $i_\ell = j_\ell$ we have necessarily enforced $p_{i_\ell i_\ell} \equiv q_{i_\ell i_\ell}$. Upon coarse-graining, η is unchanged while, in analogy with (3.5) and (3.6), the other parameters renormalize as

$$y_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} y_{i_\ell}, \quad z_{i_{\ell+1}} \equiv \sum_{i_\ell \in i_{\ell+1}} z_{i_\ell}, \tag{3.8}$$

$$f\big(d_{i_{\ell+1}j_{\ell+1}}\big) \equiv \frac{\sum_{i_\ell \in i_{\ell+1}} \sum_{j_\ell \in j_{\ell+1}} y_{i_\ell} z_{j_\ell} f\big(d_{i_\ell j_\ell}\big)}{\sum_{i_\ell \in i_{\ell+1}} y_{i_\ell} \sum_{j_\ell \in j_{\ell+1}} z_{j_\ell}}. \tag{3.9}$$

Note that Eq. (3.9) is not to be applied when $i_{\ell+1} = j_{\ell+1}$, in which case the requirement of scale-invariance bounces back to the undirected probabilities in Eq. (3.2) and thus to the renormalization rules in Eqs. (3.5) and (3.6). On the contrary, when $i_{\ell+1} \neq j_{\ell+1}$, Eqs (3.6)-(3.9) should be realized simultaneously. To ensure this, a convenient sufficient condition is that the dyadic factors $\{d_{i_\ell j_\ell}\}_{i_\ell, j_\ell=1}^{N_\ell}$ are *ultrametric* (or, more precisely, *metaultrametric*) distances compatible with the chosen hierarchy $\{\Omega_\ell\}_{\ell \geq 0}$

of nested partitions. *Metric* distances satisfy four axioms, for all i, j: positivity ($d_{ij} > 0$ if $i \neq j$), symmetry ($d_{ij} = d_{ji}$), triangular inequality ($d_{ij} \leq d_{ik} + d_{jk}\ \forall k$) and null self-distance ($d_{ij} = 0$ iff $i = j$). *Metametric* [122, 123] (also called *partial metric* [124] or *dislocated* [125]) distances relax the last axiom by requiring only $i = j$ if $d_{ij} = 0$, so they admit positive self-distance ($d_{ii} \geq 0$). *Ultrametric* distances [56] are metric distances that satisfy a stronger version of the triangular inequality, known as ultrametric inequality ($d_{ij} \leq \max_k\{d_{ik}, d_{jk}\}\ \forall k$). As a result, they can be arranged on a dendrogram whereby the distance between two 'leaves' equals the height of the closest branching point between them. In the rest of the chapter, we call *metaultrametric* distance a metametric that satisfies the ultrametric inequality. If d is metaultrametric over the dendrogram induced by the nested hierarchy $\{\Omega_\ell\}_{\ell\geq 0}$, then $f(d_{i_\ell j_\ell})$ comes out of the sums in Eqs. (3.6) and (3.9) and is therefore invariant, for any assignment of the fitness:

$$f(d_{i_{\ell+1}j_{\ell+1}}) = f(d_{i_\ell j_\ell}) \qquad \text{if} \quad i_\ell \in i_{\ell+1}, \quad j_\ell \in j_{\ell+1}, \tag{3.10}$$

with $i_{\ell+1} \neq j_{\ell+1}$. For $i_{\ell+1} = j_{\ell+1}$, $f(d_{i_{\ell+1}j_{\ell+1}})$ is still obtained through the rule (3.6). Allowing $d_{i_\ell i_\ell} > 0$ enables non-zero self-interactions ($p_{i_\ell i_\ell} > 0$) even when $\sigma = 0$. This is particularly relevant for coarse-grained configurations because any block of nodes connected by at least one link has a self-loop. From now on, we assume that the distances $\{\mathbf{D}^{(\ell)}\}_{\ell\geq 0}$ in our model are metaultrametric over $\{\Omega_\ell\}_{\ell\geq 0}$. Note that this requirement is unnecessary, although convenient, in the undirected case.

By rearranging the marginal probabilities $q_{i_\ell j_\ell}$ and $p_{i_\ell j_\ell}$ in Eqs. (3.3) and (3.7) respectively, we finally arrive at the fundamental expressions

for the joint probabilities:

$$p^{\rightarrow}_{i_\ell j_\ell} = \begin{cases} e^{-\eta z_{i_\ell} y_{j_\ell} f(d_{i_\ell j_\ell})} - e^{-\delta x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})} & i_\ell \neq j_\ell \\ 0 & i_\ell = j_\ell \end{cases}, \tag{3.11}$$

$$p^{\leftarrow}_{i_\ell j_\ell} = \begin{cases} e^{-\eta y_{i_\ell} z_{j_\ell} f(d_{i_\ell j_\ell})} - e^{-\delta x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})} & i_\ell \neq j_\ell \\ 0 & i_\ell = j_\ell \end{cases}, \tag{3.12}$$

$$p^{\leftrightarrow}_{i_\ell j_\ell} = \begin{cases} e^{-\delta x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})} & i_\ell \neq j_\ell \\ e^{-\frac{\delta}{2} x_{i_\ell}^2 f(d_{i_\ell i_\ell}) - \sigma w_{i_\ell}} & i_\ell = j_\ell \end{cases}, \tag{3.13}$$

$$p^{\leftrightarrow}_{i_\ell j_\ell} = \begin{cases} 1 - p^{\rightarrow}_{i_\ell j_\ell} - p^{\leftarrow}_{i_\ell j_\ell} - p^{\leftrightarrow}_{i_\ell j_\ell} & i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2} x_{i_\ell}^2 f(d_{i_\ell i_\ell}) - \sigma w_{i_\ell}} & i_\ell = j_\ell \end{cases}. \tag{3.14}$$

Note that the latter expression implies, whenever $i_\ell \neq j_\ell$, $p^{\leftrightarrow}_{i_\ell j_\ell} = 1 - e^{-\eta y_{i_\ell} z_{j_\ell} f(d_{i_\ell j_\ell})} - e^{-\eta z_{i_\ell} y_{j_\ell} f(d_{i_\ell j_\ell})} + e^{-\delta x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})}$. Also note that, in analogy with request (3.4), the following condition is needed to restrict all probabilities within the interval $[0, 1]$:

$$\delta^{(i_\ell, j_\ell)}_{\text{min}} \leq \delta \leq \delta^{(i_\ell, j_\ell)}_{\text{max}} \quad \forall i_\ell \neq j_\ell \tag{3.15}$$

where

$$\delta^{(i_\ell, j_\ell)}_{\text{max}} \equiv \begin{cases} -\frac{\ln\left(e^{-\eta y_{i_\ell} z_{j_\ell} f(d_{i_\ell j_\ell})} + e^{-\eta y_{j_\ell} z_{i_\ell} f(d_{i_\ell j_\ell})} - 1\right)}{x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})} & \text{if } 1 - p_{i_\ell j_\ell} - p_{j_\ell i_\ell} > 0 \\ +\infty & \text{if } 1 - p_{i_\ell j_\ell} - p_{j_\ell i_\ell} \leq 0 \end{cases}$$

while

$$\delta^{(i_\ell, j_\ell)}_{\text{min}} \equiv \frac{\eta \max\left(y_{i_\ell} z_{j_\ell}, y_{j_\ell} z_{i_\ell}\right)}{x_{i_\ell} x_{j_\ell}}.$$

Note that the above condition should hold for all pairs i_ℓ, j_ℓ simultaneously, which also means for all hierarchical levels $\ell \geq 0$. Whether it is possible to fulfill the condition, therefore, depends not only on the values of $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{y_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{z_{i_\ell}\}_{i_\ell=1}^{N_\ell}$, $\{w_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ of all nodes for a given hierarchical level ℓ, but also on the entire hierarchy $\{\Omega_\ell\}_{\ell \geq 0}$ of chosen partitions. Sufficient, although not necessary, conditions are the following:

$$\frac{\max(y_{i_{\ell+1}} z_{j_{\ell+1}}, y_{i_{\ell+1}} z_{j_{\ell+1}})}{x_{i_{\ell+1}} x_{j_{\ell+1}}} \leq \frac{\max(y_{i_\ell} z_{j_\ell}, y_{i_\ell} z_{j_\ell})}{x_{i_\ell} x_{j_\ell}}, \tag{3.16}$$

$$\frac{\left[e^{-\eta\, y_{i_{\ell+1}} z_{j_{\ell+1}} f(d_{i_{\ell+1} j_{\ell+1}})} + e^{-\eta\, y_{j_{\ell+1}} z_{i_{\ell+1}} f(d_{i_{\ell+1} j_{\ell+1}})} - 1\right]^{x_{i_\ell} x_{j_\ell} f(d_{i_\ell j_\ell})}}{\left[e^{-\eta\, y_{i_\ell} z_{j_\ell} f(d_{i_\ell j_\ell})} + e^{-\eta\, y_{j_\ell} z_{i_\ell} f(d_{i_\ell j_\ell})} - 1\right]^{x_{i_{\ell+1}} x_{j_{\ell+1}} f(d_{i_{\ell+1} j_{\ell+1}})}} \leq 1. \tag{3.17}$$

We highlight two important quantities characterizing the model. The first, local property is the *conditional* probability

$$r_{i_\ell j_\ell} \equiv \text{Prob}(i_\ell \to j_\ell | j_\ell \to i_\ell) = \frac{p^{\leftrightarrow}_{i_\ell j_\ell}}{p_{j_\ell i_\ell}} \tag{3.18}$$

that a link from i_ℓ to j_ℓ exists, given that the reciprocal link from j_ℓ to i_ℓ exists. Note that $r_{i_\ell i_\ell} = 1$, i.e. a self-loop is necessarily reciprocated (by itself).
The second, global property is the overall *reciprocity* $\langle r \rangle$, defined as the ratio of the expected value of the number $L^{\leftrightarrow} \equiv \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} a^{(\ell)}_{i_\ell j_\ell} a^{(\ell)}_{j_\ell i_\ell}$ of reciprocated links to the expected value of the number $L \equiv \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} a^{(\ell)}_{i_\ell j_\ell}$ of links in total, i.e.

$$\langle r \rangle \equiv \frac{\langle L^{\leftrightarrow} \rangle}{\langle L \rangle} = \frac{\sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} p^{\leftrightarrow}_{i_\ell j_\ell}}{\sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} p_{i_\ell j_\ell}}. \tag{3.19}$$

In the special case where all nodes are statistically equivalent (i.e., they are assigned the same fitness values), all probabilities do not depend on the specific pair of nodes and in particular $r_{i_\ell j_\ell} \equiv r = \langle r \rangle$ which equals the global reciprocity [105], motivating the choice of a common symbol for the two quantities.

Relevant cases. Before discussing more general settings, we briefly outline some relevant cases that represent useful benchmarks for later.

- *Maximal reciprocity* ($r_{i_\ell j_\ell} = 1$) is achieved when all the connections are bidirectional by construction and the model reduces to the undirected SIM, with the additional freedom of controlling self-loops separately from the other links. In particular, we get $q_{i_\ell j_\ell} = p_{i_\ell j_\ell} = p_{j_\ell i_\ell} = p^{\leftrightarrow}_{i_\ell j_\ell}$ and $p^{\rightarrow}_{i_\ell j_\ell} = p^{\leftarrow}_{j_\ell i_\ell} = 0 \; \forall \, i_\ell, j_\ell$. Clearly, in this case the conditions in Eq. (3.15) force the fitness variables to be related via the expression $y_{i_\ell} \equiv z_{i_\ell} \equiv x_{i_\ell} \sqrt{\delta/\eta} \; \forall \, i_\ell$.

- *Positive reciprocity* ($r_{i_\ell j_\ell} > p_{i_\ell j_\ell}$) is achieved when the network has a preference for the creation of reciprocal links, even when not maximally reciprocated. In particular, the conditional reciprocation

probability $r_{i_\ell j_\ell}$ is larger than the unconditional connection probability $p_{i_\ell j_\ell}$. Equivalently, the joint probability of a pair of links being reciprocated is larger than the probability of the same event occurring by chance if the two links were independent: $p^{\leftrightarrow}_{i_\ell,j_\ell} > p_{i_\ell j_\ell} p_{j_\ell i_\ell}$. Refs. [105, 107] report several real-world networks (in decreasing order of reciprocity, the World Trade Web, the World Wide Web, neural networks, email networks, word networks and metabolic networks) as belonging to this class of positive reciprocity.

- *Random reciprocity or areciprocity* ($r_{i_\ell j_\ell} = p_{i_\ell j_\ell}$) is achieved in the 'neutral' or trivial case when the two links between any two nodes are independent. Reciprocity is therefore the result of sheer chance, given the marginal connections probabilities: $p^{\leftrightarrow}_{i_\ell,j_\ell} = p_{i_\ell j_\ell} p_{j_\ell i_\ell}$. This case is achieved by fixing $\delta x_{i_\ell} x_{j_\ell} = \eta(y_{i_\ell} z_{j_\ell} + y_{j_\ell} z_{i_\ell}) \; \forall \, i_\ell, j_\ell$. This condition can be seen as an equivalent way to discriminate between a reciprocal ensemble where $\delta x_i x_j < \eta(y_i z_j + y_j z_i)$, and an antireciprocal ensemble (see below) where $\delta x_i x_j > \eta(y_i z_j + y_j z_i)$. The areciprocal case is useful to define the reference 'random' reciprocity

$$\langle r \rangle_{\text{rand}} = \frac{\sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} p_{i_\ell j_\ell} p_{j_\ell i_\ell}}{\sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell \neq i_\ell} p_{i_\ell j_\ell}} \tag{3.20}$$

  as the value achieved by Eq. (3.19) in a network where no dependency between mutual links is present.

- *Negative reciprocity or antireciprocity* ($r_{i_\ell j_\ell} < p_{i_\ell j_\ell}$) is achieved when the network tries to avoid the creation of reciprocal links. The conditional reciprocation probability $r_{i_\ell j_\ell}$ is in this case smaller than the unconditional connection probability $p_{i_\ell j_\ell}$. Equivalently, the joint probability of a pair of links being reciprocated is smaller than the probability of the same event occurring by chance if the two links were independent: $p^{\leftrightarrow}_{i_\ell,j_\ell} < p_{i_\ell j_\ell} p_{j_\ell i_\ell}$. Analogously to the case of positive reciprocity, Ref. [105] reports several real-world networks (e.g. corporate shareholding networks and food webs) as belonging to this class of negative reciprocity.

- *Minimal reciprocity (maximal antireciprocity)* corresponds, ideally, to a complete absence of mutual links, where all connections are either missing or unidirectional so to have $r_{i_\ell j_\ell} = p^{\leftrightarrow}_{i_\ell j_\ell} = 0$, $p^{\rightarrow}_{i_\ell j_\ell} = p_{i_\ell j_\ell}$, $p^{\leftarrow}_{j_\ell i_\ell} = p_{j_\ell i_\ell}$ $\forall\, i_\ell, j_\ell$. However, this extreme case cannot always be reached, as there might be no unique choice for δ achieving the fully antireciprocal limit $\delta = \delta^{\max}_{i_\ell j_\ell}$ for all node pairs simultaneously. This becomes increasingly difficult as the distribution of the fitness variables becomes broader (and, in general, as more *uneven* coarse-grainings are iterated): if $1 - p_{i_\ell j_\ell} - p_{j_\ell i_\ell} < 0$, then $p^{\leftrightarrow}_{i_\ell j_\ell} > 0$ irrespective of δ, since $p^{\leftrightarrow}_{i_\ell j_\ell} = p_{i_\ell j_\ell} + p_{j_\ell i_\ell} - q_{i_\ell j_\ell} > 1 - q_{i_\ell j_\ell} \geq 0$. This effect will be discussed more transparently in Sec. 3.3.

3.2.2 Graph probability and scale invariance.

Having derived all the dyadic connection probabilities in Eqs. (3.11) - (3.14), we can now use them to obtain the full probability $P(\mathbf{A}^{(\ell)})$ for the entire graph $\mathbf{A}^{(\ell)}$ at any hierarchical level ℓ. To do so, it is convenient to represent the graph through the mutually exclusive dyads $a^{\rightarrow}_{i_\ell j_\ell}$, $a^{\leftarrow}_{i_\ell j_\ell}$, $a^{\leftrightarrow}_{i_\ell j_\ell}$, $a^{\nleftrightarrow}_{i_\ell j_\ell}$ introduced above:

$$
\begin{aligned}
P(\mathbf{A}^{(\ell)}) &= \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} (p^{\rightarrow}_{i_\ell j_\ell})^{a^{\rightarrow}_{i_\ell j_\ell}} (p^{\leftarrow}_{i_\ell j_\ell})^{a^{\leftarrow}_{i_\ell j_\ell}} (p^{\leftrightarrow}_{i_\ell j_\ell})^{a^{\leftrightarrow}_{i_\ell j_\ell}} (p^{\nleftrightarrow}_{i_\ell j_\ell})^{a^{\nleftrightarrow}_{i_\ell j_\ell}} \\
&\equiv \frac{e^{-\mathcal{H}^{(\ell)}_{\text{eff}}(\mathbf{A}^{(\ell)})}}{\mathcal{Z}^{(\ell)}},
\end{aligned}
\tag{3.21}
$$

where, analogously to the description in Sec. 2.2.1, we have introduced the *effective Hamiltonian*

$$
\begin{aligned}
\mathcal{H}^{(\ell)}_{\text{eff}}(\mathbf{A}^{(\ell)}) = \; & - \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell=1}^{i_\ell} \Big[a^{\rightarrow}_{i_\ell j_\ell} \ln \frac{p^{\rightarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}} \\
& + \; a^{\leftarrow}_{i_\ell j_\ell} \ln \frac{p^{\leftarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}} \; + \; a^{\leftrightarrow}_{i_\ell j_\ell} \ln \frac{p^{\leftrightarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}} \Big]
\end{aligned}
\tag{3.22}
$$

and the *partition function*

$$\mathcal{Z}^{(\ell)} \equiv \sum_{\{\mathbf{A}^{(\ell)}\}} e^{-\mathcal{H}_{\text{eff}}^{(\ell)}\left(\mathbf{A}^{(\ell)}\right)} \tag{3.23}$$

$$\equiv \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} \sum_{\{(a^{\rightarrow}_{i_\ell j_\ell}, a^{\leftarrow}_{i_\ell j_\ell}, a^{\leftrightarrow}_{i_\ell j_\ell})\}} \left(\frac{p^{\rightarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}}\right)^{a^{\rightarrow}_{i_\ell j_\ell}} \left(\frac{p^{\leftarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}}\right)^{a^{\leftarrow}_{i_\ell j_\ell}} \left(\frac{p^{\leftrightarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}}\right)^{a^{\leftrightarrow}_{i_\ell j_\ell}}$$

$$\equiv \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} \left(\frac{p^{\rightarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}} + \frac{p^{\leftarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}} + \frac{p^{\leftrightarrow}_{i_\ell j_\ell}}{p^{\nleftrightarrow}_{i_\ell j_\ell}} + 1\right)$$

$$\equiv \prod_{i_\ell=1}^{N_\ell} \prod_{j_\ell=1}^{i_\ell} \frac{1}{p^{\nleftrightarrow}_{i_\ell j_\ell}}$$

$$\equiv e^{\frac{\delta}{2} x_{i_\infty}^2 f(d_{i_\infty, i_\infty}) + \sigma w_{i_\infty}}, \tag{3.24}$$

with $x_{i_\infty} \equiv \sum_{i_\ell=1}^{N_\ell} x_{i_\ell}$, $w_{i_\infty} \equiv \sum_{i_\ell=1}^{N_\ell} w_{i_\ell}$ and $f(d_{i_\infty,i_\infty}) \equiv x_{i_\infty}^{-2} \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell=1}^{N_\ell} x_{i_\ell} x_{j_\ell} f(d_{i_\ell,j_\ell})$ being scale-invariant quantities: $f(d_{i_\infty,i_\infty})$ is independent of the hierarchical level ℓ, while x_{i_∞} and w_{i_∞} are even independent of the hierarchy of partitions.

3.3 A simplified benchmark: the homogeneous case

As the simplest benchmark for the more complicated cases that we will consider later, we first illustrate the purely homogeneous case where the fitness variables are all equal and the dyadic factors are switched off ($f \equiv 1$), so that the only free parameters left are η and δ. Without loss of generality, we can then set $x_{i_0} \equiv y_{i_0} \equiv z_{i_0} \equiv w_{i_0} = 1 \; \forall \, i_0 = 1, N_0$ at level $\ell = 0$. Homogeneity will then be preserved at any subsequent level $\ell > 0$, provided that blocks of nodes are all equal in size for any ℓ. This model represents a scale-invariant reparametrization of the well-known *p1 model* by Holland and Leinhard [126], where the random graph is homogeneous but with nontrivial reciprocity. We do not consider self-loops at level $\ell = 0$ and we therefore set $\sigma = -\delta/2$, so that $p^{\leftrightarrow}_{i_0 i_0} = p^{\rightarrow}_{i_0 i_0} = p^{\leftarrow}_{i_0 i_0} = p_{i_0 i_0} =$

$q_{i_0 i_0} = 0$ for all $i_0 = 1, \ldots, N_0$.
In this scenario, all the connection probabilities are identical for different pairs of nodes (i_0, j_0) (with $i_0 \neq j_0$) and equal to

$$\begin{aligned} p &= 1 - e^{-\eta}, \\ q &= 1 - e^{-\delta}, \\ p^{\rightarrow} &= q - p = e^{-\eta} - e^{-\delta}, \\ p^{\leftrightarrow} &= 2p - q = 1 + e^{-\delta} - 2e^{-\eta}. \end{aligned} \tag{3.25}$$

Notably, the conditions in Eq. (3.15) become very transparent:

$$\eta \leq \delta \leq \delta_{\max}, \quad \delta_{\max} = \begin{cases} -\ln(2e^{-\eta} - 1), & \text{if } p < \frac{1}{2} \\ \infty, & \text{otherwise} \end{cases}$$

and are *global* (i.e. not pair-specific). This allows us to easily interpret the two extreme values of δ in terms of the resulting expected reciprocity, which in this case equals $\langle r \rangle = p^{\leftrightarrow}/p$. In particular, for $\delta = \delta_{\min} = \eta$ we recover maximal reciprocity ($\langle r \rangle = 1$), while for $\delta = \delta_{\max}$ we recover minimal reciprocity ($\langle r \rangle = \langle r \rangle_{\min}$). In the latter case, we have to distinguish between two possibilities: if $p > \frac{1}{2}$, then $\langle r \rangle_{\min}$ is necessarily larger than zero since $p^{\leftrightarrow}(\delta_{\max}) > 0$. This is correctly mirroring the fact that if the average link density (which in the homogeneous case equals p) is larger than $1/2$, then the minimum number $L^{\leftrightarrow}_{\min}$ of reciprocated links is given by twice the number of links exceeding the number of vertex pairs, i.e. $L^{\leftrightarrow}_{\min} = 2\,[L - N(N-1)/2]$. This implies that the realized value of r cannot be smaller than $L^{\leftrightarrow}_{\min}/L$ and, consequently, the minimum expected reciprocity takes the strictly positive value $\langle r \rangle_{\delta_{\max}} = 2 - \frac{1}{p}$. On the other hand, for $p \leq 1/2$ we can reach the complete absence of reciprocal links and get $\langle r \rangle_{\delta_{\max}} = 0$ for $\delta_{\max} = -\log(1 - 2p)$.
In what follows, we keep discussing the properties of the model at level $\ell = 0$. Clearly, the properties at the next levels $\ell > 0$ will depend on the details of the partitions chosen. Importantly, choosing block nodes of the same size at each level will keep the model homogeneous, with the same values of the global parameters δ, η and σ, but rescaled and increasingly larger values of the fitnesses (yet still equal for different blocks). In the

latter case, adjusting the results of this section to any desired hierarchical level is straightforward: coarser (finer) representations of the network effectively correspond to larger (smaller) values of the global parameters. Clearly, self-loops become more likely in coarser configurations. By contrast, partitions into heterogeneously sized blocks will effectively turn the model into a non-homogeneous one.

We are mainly interested in the amount of reciprocation generated by the model, compared with an areciprocal benchmark, or equivalently with the random reciprocity $\langle r \rangle_{\text{rand}} = p$ generated by the model itself in the particular case when $p^{\leftrightarrow} = p^2$, i.e. for $\delta_{\text{rand}} = 2\eta$ (this case simply reducing to a directed Erdős-Rényi model). Discounting for this 'baseline' reciprocity can be done transparently by adopting a modified metric of reciprocity [105] defined as the Pearson correlation coefficient ρ between the symmetric entries of the adjacency matrix, which for homogeneous random graphs takes the convenient form

$$\rho \equiv \frac{\sum_{i,j\neq i}(a_{ij} - \langle a_{ij} \rangle)(a_{ji} - \langle a_{ij} \rangle)}{\sum_{i,j\neq i}(a_{ij} - \langle a_{ij} \rangle)^2} = \frac{r-p}{1-p}.$$

Unlike r, ρ is an absolute quantity and takes values in the interval $[-1, 1]$. The value $\rho = 0$ points at the areciprocal case, where the couples a_{ij} and a_{ji} are indeed statistically independent. The sign of ρ immediately distinguishes then between reciprocal ($\rho > 0$) and antireciprocal ($\rho < 0$) networks. The explicit expressions for the expectations of r and ρ are

$$\begin{aligned} \langle r \rangle &= 2 - \frac{q}{p} = 2 - \frac{1-e^{-\delta}}{1-e^{-\eta}}, \\ \langle \rho \rangle &= \frac{(2-p)p - q}{(1-p)p} = \frac{e^{-(\delta-\eta)} - e^{-\eta}}{1-e^{-\eta}}. \end{aligned} \tag{3.26}$$

The above functions are illustrated in Fig. 13, where the dependence of $\langle r \rangle$ and $\langle \rho \rangle$ on the parameter δ is shown for four different values of η (i.e. of the expected link density p). It is interesting to note how the antireciprocal regime converges to the areciprocal one as $\eta \to 0$: in this limit, the graph gets so sparse that it becomes extremely unlikely for two nodes to establish mutual connections, and reciprocated links are effectively absent. This

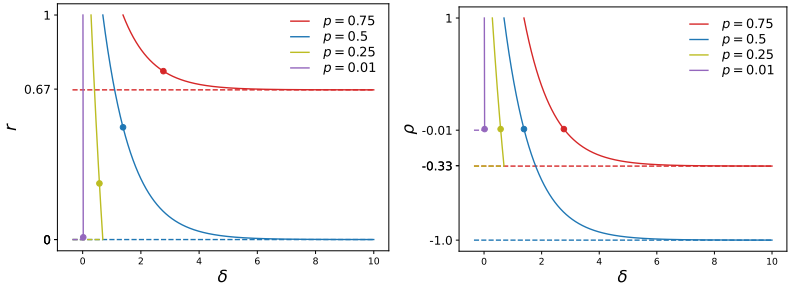

Figure 13: Reciprocity (r) and correlation coefficient (ρ) as functions of the density parameter (δ). The solid lines illustrate, for different values of the link density p, the behaviour of the functions $\langle r \rangle$) and $\langle \rho \rangle$ in equations (3.26), while the isolated points indicate the values achieved in the areciprocal case $\delta_{\text{rand}} = 2\eta$ (where $\langle r \rangle_{\text{rand}} = p$ and $\langle \rho \rangle_{\text{rand}} = 0$). The dashed lines indicate the minimum values of $\langle r \rangle$ and $\langle \rho \rangle$ attainable for the given values of p. The four values of p considered are: 1) $p = 0.75$, for which the maximum value $q_{\max}$ for q is 1, $\langle r \rangle_{\min} = \frac{2p-1}{1-p}$ and $\langle \rho \rangle_{\min} = \frac{2-p-1/p}{1-p}$; 2) $p = 0.5$, for which $q_{\max} = 1$, $\langle r \rangle_{\min} = 0$, and $\langle \rho \rangle_{\min} = -\frac{p}{1-p}$; 3) $p = 0.25$, for which $q_{\max} = 0.5$, $\langle r \rangle_{\min} = 0$, and $\langle \rho \rangle_{\min} = -\frac{p}{1-p}$; 4) $p = 0.01$, for which $q_{\max} \sim q_{\text{rand}}$, $\langle r \rangle_{\min} = 0$, and $\langle \rho \rangle_{\min} = -\frac{p}{1-p} \sim 0$.

can be confirmed by expanding the conditions (3.15) around $\eta = 0$: the value of δ for which $p^{\leftrightarrow}$ vanishes is $\delta_{\max} \overset{\eta \sim 0}{\sim} -\ln(1-2\eta) \sim 2\eta$, which coincides with the value of δ for which $p^{\leftrightarrow} = p^2$ (i.e. the areciprocal case). Note that, starting from such a sparse configuration at level $\ell = 0$ and progressively coarse-graining the network implies a departure from this scenario, as reciprocated links become more and more likely.

A remarkable effect of reciprocity is visible in the spectral density of the adjacency matrix of directed graphs, as pointed out in [112]. In our setting we can define the centred random matrix $\mathbf{C}$ with elements $C_{ij} = \frac{a_{ij} - p_{ij}}{\sqrt{N p_{ij}(1-p_{ij})}}$, which have zero mean and unit variance. Then, the covariance between its off-diagonal entries can be easily written in the suggestive form

$$\langle C_{ij} C_{ji} \rangle = \frac{p^{\leftrightarrow} - p^2}{Np(1-p)} = \rho \quad i \neq j.$$

According to [112], the spectral density of $\mathbf{C}$ is (asymptotically) uniform on the ellipse with axes $(1+\rho, 1-\rho)$. In terms of reciprocity, this statement has an intuitive meaning: the larger the value of ρ (i.e. the 'more symmetrical' the matrix $\mathbf{C}$), the more the spectral ellipse 'narrows' along the real axis, until it reaches the undirected (fully symmetric) limit, for which $\rho = 1$ and the Wigner semicircle law is recovered [127]. On the opposite, the smaller ρ, the more the ellipse concentrates on the imaginary axis. In the intermediate areciprocal case, we find a perfect circle ($1 + \rho = 1 - \rho = 1$) as expected for matrices with perfectly uncorrelated entries [128].
The above picture is illustrated in Fig. 14, where the outcome of numerical simulations is shown for a single realization of the homogeneous model and a sample of 100 realizations. Note that having centred the matrix $\mathbf{A}$ indeed removes the largest eigenvalue $\lambda_{\max}(\mathbf{A})$ (which is known to behave as $\sim Np$ for Erdős-Rényi random graphs, either directed o undirected) from the spectrum of $\mathbf{C}$, as illustrated in the right panels in Fig. 14. Moreover, rescaling by $\sqrt{Np_{ij}(1-p_{ij})}$ is convenient as N gets large.

The remarkable consistency between individual realizations and the theoretical elliptic distribution, already for moderate network sizes, justifies the use of the spectral density of the adjacency matrix of (necessarily finite) real-world networks as a powerful quantification of reciprocity, as we will see in the next section on a specific case study.

3.4 Quenched fitness

In this section, we move away from the homogeneous setting and address the general case where different nodes can have different fitness values, while still treating the latter as deterministic variables (which therefore represent some sort of 'intrinsic characteristics' of nodes). In particular, each node i_ℓ is provided with a fitness quadruplet $(x_{i_\ell}, y_{i_\ell}, z_{i_\ell}, w_{i_\ell})$ that will determine its tendency of establishing directed links with other nodes and with itself. We recall that when using the model to describe a real-world system, every element of the latter may be represented as a node at a certain level ℓ, while the fitness parameters may be identified

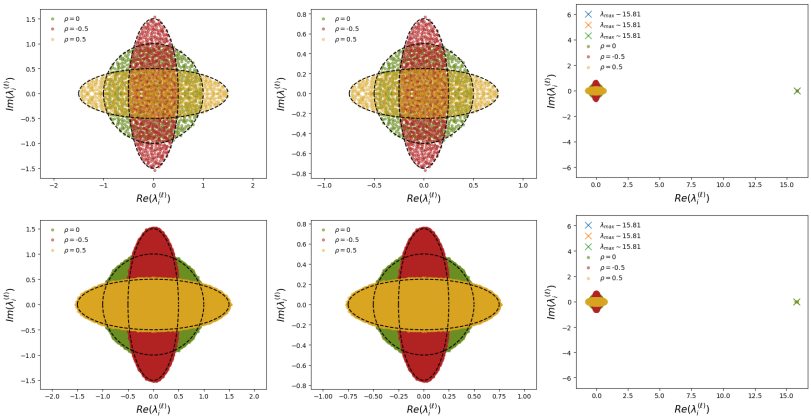

Figure 14: Spectral density for homogeneous directed networks. Complex spectra of the (centred and rescaled) adjacency matrix of the network for 1 (top panels) and 100 (bottom panels) realizations of the homogeneous DSIM with $N = 1000$ nodes. In all cases the parameter η is kept fixed to the value $\eta = 0.69315$, corresponding to the link density $d = 0.5$, which allows the parameter ρ to vary across its full range $[-1, 1]$ (see Fig. 13). The spectra are shown for the three values $\delta = 2.0794$ ($\rho = -0.5$, red), $\delta = 2\eta$ ($\rho = 0$, green), and $\delta = 0.9808$ ($\rho = 0.5$, yellow). Left: spectra of the matrix with entries $C_{ij} = \frac{a_{ij}-p}{\sqrt{Np(1-p)}}$, which are uniform over the Ellipse$[1+\rho, 1-\rho]$. centre: spectra of the matrix with entries $\frac{a_{ij}-p}{\sqrt{N}}$, which are uniform over the Ellipse$[(1+\rho)\sqrt{p(1-p)}, (1-\rho)\sqrt{p(1-p)}]$. Right: spectra of the matrix with entries a_{ij}, which consist of an elliptic 'bulk' and an isolated largest real eigenvalue $\lambda_{\max}$ (crosses). Note that $\lambda_{\max} \sim Np$ does not depend on δ and the resulting level of reciprocity.

with measurable node-specific quantities. Note that we may choose to tune only the total number of self-loops and not their individual node-specific probabilities, thereby relying solely on the parameter σ and fixing $w_{i_\ell} = 1 \ \forall i_\ell = 1, \dots N_\ell$. As already mentioned, we may also add a (symmetric) dyadic property d_{i_ℓ, j_ℓ} to each pair of nodes (i_ℓ, j_ℓ), to further encourage (or discourage) their connection. In this way, the only remaining free parameters are the global, scale-invariant ones (η, δ, σ). The latter can be adjusted by constraining the link density, the global reciprocity, and the overall number of self-loops, respectively. Once the

parameters are specified, the connection probabilities in Eqs. (3.11)-(3.14) are determined simultaneously at all scales, thereby providing a testable multi-scale model of the renormalized network at any higher level of aggregation. In the following subsection, we illustrate the procedure by considering the empirical International Trade Network (ITN).

3.4.1 The directed International Trade Network

In Chap. 2 we have shown how the undirected SI model can capture relevant features of the observed (symmetric) network of bilateral international trade, not only at the 'native' level of resolution (where nodes represent countries and links represent the existence of trade in any direction) but also consistently across different aggregation levels, e.g. between groups of countries, regions, etc.. Here we walk the same line while focusing on the finer topology induced by *directed* international trades, where a directed link represents export from one node to another. This is not a straightforward extension of the undirected case, precisely because of the strongly reciprocal nature of this particular directed network, which has been pointed out for example in [100, 105]. Capturing the reciprocity structure of the ITN is crucial, among other things, for correctly assessing the statistical significance of higher-order patterns such as triadic motifs [110]. We source the data from the expanded trade dataset developed by K. S. Gleditsch [99], which provides trade flow estimates among 185 countries for the years $1948 - 2000$. Here we show the results for the year 2000, but similar outcomes have been obtained for the other years as well (more details on the dataset are provided in Sec. 3.7.1).

We define our empirical network at the native level $\ell = 0$ by considering $N_0 = 185$ 0-nodes, each representing a country in the dataset [99]. The empirical adjacency matrix $\tilde{\mathbf{A}}^{(0)}$ (where the tilde indicates the specific empirical matrix in the ensemble of possible matrices that the model we are going to define can generate) is then built by setting the entry $\tilde{a}^{(0)}_{i_0 j_0} = 1$ if there is any reported (i.e. positive) export relationship from country i_0 to country j_0, and $\tilde{a}^{(0)}_{i_0 j_0} = 0$ otherwise. We set by convention all self-loops to zero: $\tilde{a}^{(0)}_{i_0 i_0} = 0, \ \forall\, i_0 = 1, \dots, N_0$.

Definition of the multiscale model

In what follows, we define our multiscale model for the directed ITN by using the probabilities introduced in Eqs (3.11)-(3.14) and specifying all the parameters therein.

We first discuss the choice of the fitness variables x_{i_0} for each country $i_0 = 1, N_0$. Building upon prior findings that highlight the significance of the Gross Domestic Product (GDP) in shaping the structure of the undirected ITN [68–71], we follow the approach taken in Chap. 2 and set $x_{i_0} = \text{GDP}_{i_0}$ for all $i_0 = 1, N_0$. As remarked in the undirected analysis, this choice is consistent by the additive nature of GDP upon aggregation (the total GDP of a set of countries is the sum of the individual GDPs), which compels with the renormalization rule for x in Eq. (3.5).

Next, we fix the dyadic parameters, for which a natural candidate is provided by the geographic distances between countries. Again, this builds upon established empirical knowledge of the role of distances in determining trade flows, as initially conceptualized by the traditional Gravity Model of trade [72] and more recently confirmed by various network analyses of the ITN [71, 129, 130]. On the tail of the analysis of the undirected ITN in Sec. 2.3, we specifically consider the population-averaged inter-country distances $\{d_{i_0 j_0}\}$ provided by the BACI-CEPII GeoDist database [75, 76] (see Sec. 2.6.2). However, as discussed in general in Sec. 3.2, we need our distances to be *metaultrametric*. For this reason, for all i_0 we set our diagonal metaultrametric distances $d^{\text{mu}}_{i_0 i_0}$ equal to the non-zero self-distances $d_{i_0 i_0} > 0$ as already reported in the GeoDist data (representing the population-averaged distance between agglomerations inside each country). Next, we construct the off-diagonal metaultrametric distances $d^{\text{mu}}_{i_0 j_0}$ (for $i_0 \neq j_0$) via a single-linkage hierarchical clustering algorithm, which produces the so-called *subdominant ultrametric* distances representing the closest ultrametric 'from below' to the original metric [56]. The subdominant ultrametric distances are obtained by using the original geographic distances $\{d_{i_0 j_0}\}$ as measure of dissimilarity. The output of this procedure is a dendrogram whose leaves correspond to the original countries (0-nodes) and the metaultrametric distance between each

pair is given by the height of the branching point for the corresponding branches. In particular, pairs of nodes that have the same ancestor have the same distance, so that the cardinality of different values in the matrix $\mathbf{D}^{\mathrm{mu}}$ is given by the number of branching points, that in any dendrogram amounts to at most $N-1$ (and is equal to $N-1$ if the elements are merged in pairs at each step of the algorithm). Then, we set the function f as $f(d^{\mathrm{mu}}) = 1/d^{\mathrm{mu}}$.
We then consider the fitness $\{w_{i_0}\}$. To replicate the absence of empirical self-loops at level 0 we set $w_{i_0} \equiv x_{i_0}^2 f(d^{\mathrm{mu}}_{i_0 i_0})$ for all $i_0 = 1, N_0$ and $\sigma \equiv -\delta/2$ (the latter is analogous to the choice made in Sec. 3.3). Note that, as we proceed to higher levels $\ell > 0$ through coarse-graining, self-loops will eventually emerge both in the empirical network and the model.
Finally, we consider the fitness values y_{i_0} and z_{i_0}. As already observed in [105] and confirmed for this specific dataset (see Fig. 15), each country in the ITN exhibits approximately equal in-degrees $k^{\mathrm{in}}_{i_0} = \sum_{j_0 \neq i_0} a_{j_0 i_0}$ and out-degrees $k^{\mathrm{out}}_{i_0} = \sum_{j_0 \neq i_0} a_{i_0 j_0}$. As a consequence, for this particular system, a natural choice is the symmetric one, where $p_{i_0 j_0} \equiv p_{j_0 i_0}$, i.e. $y_{i_0} \equiv z_{i_0}, \forall\, i_0, j_0$. Moreover, to keep the focus on GDP alone as the main node-specific driver of trade links, we make the convenient choice of taking the vectors $\vec{y}$ and $\vec{z}$ proportional to $\vec{x}$. Since the (positive) proportionality constant can be reabsorbed in the global parameter η, we can posit $\vec{x} \equiv \vec{y} \equiv \vec{z}$.

With the above specifications, the only free parameters left are η and δ, which we tune, respectively, so that both the expected number $\langle L_0^{\mathrm{nl}} \rangle$ of directed links (self-loops excluded) and the expected number $\langle L_0^{\leftrightarrow} \rangle$ of reciprocated links equal their measured counterparts $\tilde{L}_0^{\mathrm{nl}}$ and $\tilde{L}_0^{\leftrightarrow}$ in the empirical network, i.e. at level $\ell = 0$ (see Sec. 3.7.2).

We can now test the model against the real data at several levels of aggregation, through a sequence of iterative coarse-grainings induced by the dendrogram of metaultrametric distances discussed above. We follow the same procedure adopted in Sec. 2.3 and cut the dendrogram horizontally at 17 different levels, such that the number of block-countries, at any coarse-graining level ℓ, is given by $N_0 = 185$ for $\ell = 0$ and

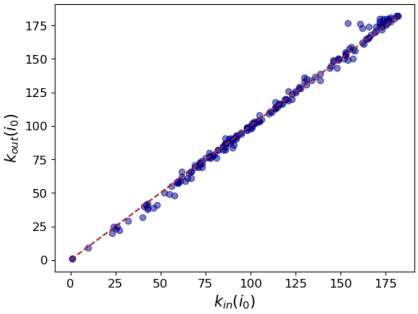

Figure 15: Symmetry in the WTW: the above scatter-plot illustrates the linear relation between in-degrees $\{k_{i_0}^{\text{in}}\}_1^{N_0}$ and out-degrees $\{k_{i_0}^{\text{out}}\}_1^{N_0}$ of $0-$nodes (i.e., countries) in the World Trade Web [99].

$N_\ell = 180 - 10\ell$, $\forall\ \ell = 1, \dots, 16$. This procedure generates the hierarchy of coarse-grained empirical networks $\{\tilde{\mathbf{A}}^{(\ell)}\}_{\ell=1,\dots,17}$ via the procedure illustrated in Fig. 12. In this way, each ℓ-node represents a geographic aggregate of nearby countries, and the additivity of GDPs ensures that the renormalized fitness always represents the empirical aggregate GDP of a block-country. Moreover, the transformation rules in (3.6) and (3.9) leave the off-diagonal metaultrametric distances $f(d^{\text{mu}}_{i_{\ell+1},j_{\ell+1}})$ invariant at any level ℓ, as specified in (3.10), while the diagonal terms $f(d^{\text{mu}}_{i_{\ell+1},i_{\ell+1}})$ renormalize as in (3.6). The values of the model parameters, together with their transformation rules across hierarchical levels, determine the connection probabilities simultaneously at all scales considered. We have checked that the conditions in (3.15) are verified along the entire coarse-graining flow.

Capturing structural and spectral properties

First of all, it is worth noting that all the quantities that had been reproduced in the undirected case, referring to global and local properties of the network, are still well captured by the DSIM, thus confirming the desired consistency between the directed ensemble and its undirected

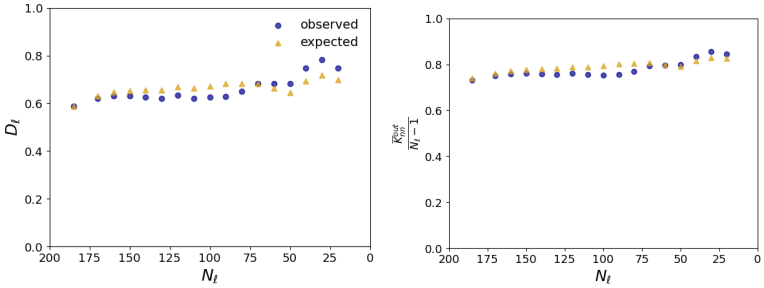

Figure 16: Prediction of global topological properties of the ITN across the full spectrum of geographical aggregation using the DSIM. The panels show the agreements between the empirical and the expected values of the link density D_ℓ (on the left) and the node-averaged rescaled average nearest neighbour out-degree $\overline{k}^{\text{out}}_{\text{nn}}(i_\ell)/(N_\ell - 1)$ as functions of the number N_ℓ of countries at all the 17 hierarchical levels considered, with $N_0 = 185$ and $N_\ell = 180 - \ell\, 10$ for any $\ell = 1, \dots 16$.

projection. To illustrate this, here we consider the direct generalization of the topological quantities considered in Sec. 2.3 (defined in Sec. 3.7.2): the link density D_ℓ, the out-degree $k^{\text{out}}_{i_\ell}$ (the "in-" counterparts being perfectly equivalent) and the average nearest neighbour out-degree $k^{\text{out}}_{\text{nn}\,i_\ell}$. We performed both the global and the local tests and show the outcome in Figures 16 and 17.

Now we focus our attention on quantities that are somehow related to network reciprocity. As anticipated, we find that the strongly reciprocated nature of the ITN cannot be captured without explicitly enforcing it, since the empirical global reciprocity in the network is not consequential to other topological properties. This can be seen by comparing various features of the real ITN with the theoretical expectations provided by our DSIM in two different settings: a *fitted* one, where the parameter δ is tuned to replicate the empirical number of reciprocated links (as explained above and detailed in Sec. 3.7.2), and a *random* one, where $\delta_{\text{rand}} = \frac{\eta(y_i z_j + y_j z_i)}{x_i x_j}$ as in the areciprocal benchmark for our model discussed in 3.2.1. In the symmetric ($y_i = z_i$) setting considered here, this is simply obtained as $\delta_{\text{rand}} = 2\eta$. We consider both topological and spectral properties, for

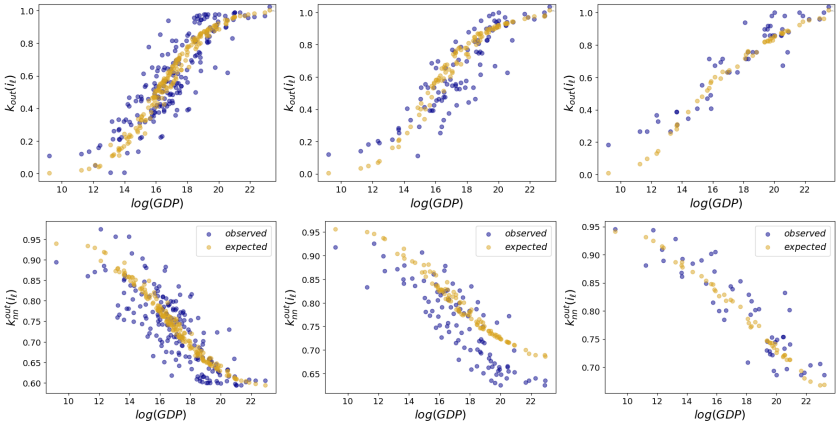

Figure 17: Prediction of local topological properties of the ITN at different levels of aggregation using the DSIM. Top panels: empirical (blue) and expected (goldenrod) out-degree $k_{i_\ell}^{out}$ vs $\ln(GDP_{i_\ell})$ for all N_ℓ nodes, for three representative hierarchical levels ($\ell_1 = 0$, $\ell_2 = 8$, $\ell_3 = 13$) such that $N_0 = 185$ (left), $N_2 = 100$ (centre) and $N_3 = 50$ (right). Bottom panels: empirical (blue) and expected (goldenrod) average nearest-neighbour out-degree $k_{\rm nn}^{\rm out}(i_\ell)$ vs $\ln(GDP_{i_\ell})$ for all N_ℓ nodes, for the same three hierarchical levels.

various hierarchical levels: specifically, the global reciprocity r_ℓ, the reciprocated degree $k_{i_\ell}^{\leftrightarrow} \equiv \sum_{j_\ell \neq i_\ell} a_{i_\ell j_\ell} a_{j_\ell, i_\ell}$ of each node i_ℓ and the spectral density of the adjacency matrix $\mathbf{A}^{(\ell)}$. The results are shown in Fig. 18.

First, we note from the bottom left panel of Fig. 18 that the global reciprocity r_ℓ is always systematically larger in the empirical network than in the areciprocal model with the same link density. By contrast, the reciprocated model replicates r_0 by construction, via the fitted parameter δ. Yet, it is quite remarkable that the agreement between the empirical and the predicted reciprocity remains in place across all subsequent aggregation levels $\ell > 0$, even if no further refitting of the parameters has taken place.

Similar considerations apply to the relationship between the reciprocal degree $k_{i_\ell}^{\leftrightarrow}$ and the undirected degree $k_{i_\ell}^{\rm und}$, which is shown for three representative levels of aggregation (top panels of Fig. 18). Note that

the areciprocal directed model cannot replicate the empirical (approximately linear) relationship between the two quantities, the reciprocated degree being systematically underestimated. By contrast, the full model replicates the empirical replationship remarkably well: even if only the expected global reciprocity at level $\ell = 0$ is fitted through the parameter δ, the model successfully predicts the local reciprocated degree for each node separately, and keeps doing so (again, without refitting) across aggregation levels. Finally, we consider the spectral properties of the network, in analogy with the analysis in Sec. 3.3, but here for heterogeneous fitness values. We compare the spectral distribution of the ITN with the one of samples drawn from the DSIM ensemble. We generated $N_g = 1000$ synthetic realizations of the ITN for each of the two models considered (with δ fitted and δ_{rand}) and then juxtaposed their spectra with the empirical one. The outcome is shown on the bottom right of Fig. 18, where the strongly eccentric (along the real axis) elliptical distribution of the ITN is closely replicated by the model with reciprocation, while the areciprocal model produces the usual circular distribution.

3.5 Annealed fitness

As in the undirected case, the DSIM can be framed in an annealed scenario, where not only the graph structure but also the fitness are treated as random variables. Then, the requirement of invariance upon aggregation is to be extended to the functional form of the probability density function (pdf) describing these random variables. In the directed case, each node is in general endowed with four random fitness variables, so at each level ℓ there are four N_ℓ-dimensional random vectors ($\vec{x}^{(\ell)}, \vec{y}^{(\ell)}, \vec{z}^{(\ell)}$ and $\vec{w}^{(\ell)}$).

For simplicity, we disregard the dyadic factors, i.e. we posit $f \equiv 1$ in Eqs. (3.11), (3.12), (3.13), (3.14), and leave to chance the formation of self-loops by setting $\vec{w}^{(0)} \equiv \vec{0}$. The additivity of w upon aggregation then also implies $\vec{w}^{(\ell)} = \vec{0}$ deterministically for all higher levels $\ell > 0$. We are then left with three random vectors ($\vec{x}^{(\ell)}, \vec{y}^{(\ell)}$ and $\vec{z}^{(\ell)}$) which can in general be mutually dependent in complicated ways, both across nodes (e.g. x_{i_ℓ} and x_{j_ℓ} can be correlated) and across fitness types (e.g. x_{i_ℓ}, y_{i_ℓ}

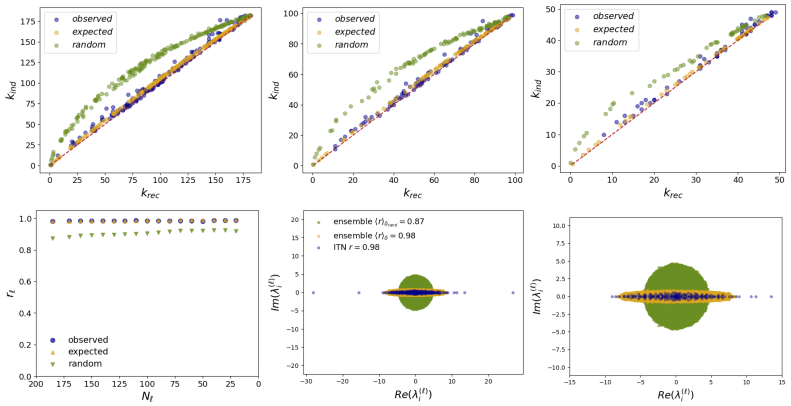

Figure 18: Topological and spectral properties of the ITN across different levels of aggregation. Various features of the empirical ITN (in blue) are compared with the expectations provided by the DSIM (in goldenrod) and by the *areciprocal DSIM* - where $\delta = \delta_{\text{rand}}$ (in green). TOP PANELS: scatter-plot of the undirected projection of the degree $k_{i\ell}^{\text{und}}$ vs the reciprocal degree $k_{i\ell}^{\leftrightarrow}$ of each node, at three representative levels of aggregation: starting from $\ell = 0$ with $N_0 = 185$ nodes, corresponding to the original ITN (on the left), to $\ell = 8$ with $N_8 = 100$ nodes (at the centre) to $\ell = 13$ with $N_{13} = 50$ nodes (on the right). BOTTOM PANELS: in the figure on the left, the observed and expected values of global reciprocity r_ℓ are shown as functions of the number N_ℓ of nodes along the 17 considered levels of aggregation of the original network, with $N_0 = 185$ and $N_\ell = 180 - \ell\, 10$ for any $\ell = 1, \dots 16$. In the two figures on the right, we focus on the spectral density of the ITN and show real and imaginary parts of the eigenvalues induced by the empirical ITN (in blue), by a sample of $N_g = 1000$ synthetic networks drawn from the DSIM ensemble (with the same density and same reciprocity of the empirical network of trades) (in goldenrod) and by a sample of $N_g = 1000$ synthetic networks drawn from the random DSIM (with the same density of the empirical network of trades but random reciprocity) (in green). For visualization purposes, in all cases the spectrum of the centred matrix $\mathbf{A}_c$ (with elements $a_{ij}^{(c)} = a_{ij} - p_{ij}$) is considered instead of the one of the adjacency matrix $\mathbf{A}$. The picture on the right offers a soft zoom on the bulk of the spectrum shown on the left.

and $z_{i\ell}$ can be correlated). While it is reasonable to assume statistical independence between different nodes (in the general spirit of fitness models or inhomogeneous random graphs considered in the literature),

doing the same between different fitness types would lead to an unrealistic scenario where, for instance, there would be no correlation between the in-degree and out-degree of the same node (while, as the ITN example considered above illustrates in a particularly strong way, these variables are in general correlated). For the sake of simplicity, we consider the extreme case of perfect correlation by setting $\vec{x}^{(\ell)} \equiv \vec{y}^{(\ell)} \equiv \vec{z}^{(\ell)}$ for all $\ell \geq 0$.

With the above choices, the multivariate problem reduces to a univariate one, where each ℓ-node i_ℓ is endowed with a single random fitness x_{i_ℓ} (the fitness values for all nodes being *i.i.d.*) that impacts all its types of connections through the probabilities defined by Eqs. (3.11)-(3.14), which here become

$$\begin{aligned} p_{i_\ell j_\ell} &= 1 - e^{-\eta x_{i_\ell} x_{j_\ell}}, \\ q_{i_\ell j_\ell} &= 1 - e^{-\delta x_{i_\ell} x_{j_\ell}}, \\ p^{\rightarrow}_{i_\ell j_\ell} &= p^{\leftarrow}_{i_\ell j_\ell} = e^{-\eta x_{i_\ell} x_{j_\ell}} - e^{-\delta x_{i_\ell} x_{j_\ell}}, \\ p^{\leftrightarrow}_{i_\ell j_\ell} &= 1 + e^{-\delta x_{i_\ell} x_{j_\ell}} - 2e^{-\eta x_{i_\ell} x_{j_\ell}}, \end{aligned} \tag{3.27}$$

for any $i_\ell \neq j_\ell$, while the diagonal terms read $p_{i_\ell i_\ell} = q_{i_\ell i_\ell} = p^{\leftrightarrow}_{i_\ell i_\ell} = 1 - e^{-\frac{\delta}{2} x^2_{i_\ell}}$.

As discussed in Sec. 2.4, requiring that the pdf of the (necessarily positive) fitness values is positively supported and invariant upon aggregation (just like the graph probability distribution) leads to the selection of *one-sided α-stable distributions* [83]. In this way, at any level ℓ of observation, a realization $\{x_{i_\ell}\}_{i_\ell=1}^{N_\ell}$ of fitness variables can be obtained in two equivalent ways, either *hierarchically*, from any lower hierarchical level (say $\ell-1$) by summing the finer-grained variables $\{x_{i_{\ell-1}}\}_1^{N_{\ell-1}}$ based on the chosen partition $\Omega_{\ell-1}$, or *directly* by drawing N_ℓ i.i.d. random variables from a pdf with invariant functional form and scale-dependent parameters.

We recall that the aforementioned property is ensured by the *closure under convolution* that characterizes α-stable random variables, whose pdf we denote as $\varphi_\ell(x_{i_\ell}; \alpha, \beta_\ell, \gamma_\ell, \mu_\ell)$, where α is the stability parameter (which, in the one-sided case, ranges in the interval $0 < \alpha < 1$), β_ℓ controls the skew-

ness, γ_ℓ the scale and μ_ℓ the location of the distribution. In the specific case of one-sided α-stable distributions with support over the non-negative real numbers, we have $\mu_0 = 0$ and $\beta_0 = 1$. Upon aggregation, the only parameter that gets modified is γ_ℓ, while α remains unchanged due to the stability property and β_ℓ and μ_ℓ are mapped onto their original values $\beta_0 = 1$ and $\mu_0 = 0$. In order to preserve the *i.i.d.* nature of the fitness across ℓ-nodes, we assume that, to pass from level ℓ to $\ell + 1$, blocks of homogeneous sizes are formed, i.e. all $(\ell + 1)$-nodes contain exactly the same number (say, b_ℓ) of ℓ-nodes. Under this assumption, γ_ℓ transforms as $\gamma^\alpha_{\ell+1} = b_\ell \gamma^\alpha_\ell$.
In the remainder of this section, we narrow the focus on the particular case of the Lévy distribution, corresponding to $\alpha = 1/2$ and representing the only α-stable distribution that can be written in explicit form when $0 < \alpha < 1$. The Lévy pdf reads

$$\varphi(x; \gamma_{i_\ell}) \equiv \varphi_\ell(x; 1/2, 1, \gamma_{i_\ell}, 0) = \sqrt{\frac{\gamma_{i_\ell}}{2\pi}} \frac{e^{-\gamma_{i_\ell}/2x}}{x^{3/2}}. \tag{3.28}$$

In what follows, we explore the topology of network realizations drawn from the DSIM ensemble in the annealed setting described above, i.e. by connecting pairs of ℓ-nodes with the connection probabilities given in Eq. (3.27) and Lévy distributed fitness variables. In particular, we investigate how the parameters η and δ jointly shape the features of such networks at different levels of aggregation.

3.5.1 Emergent positive reciprocity

Before delving into a multiscale analysis, we first note that, under our choice of Lévy-distributed fitness and $\vec{x}^{(\ell)} \equiv \vec{y}^{(\ell)} \equiv \vec{z}^{(\ell)}$, the formation of antireciprocal networks is suppressed. To support the former statement we recall from the discussion of the homogeneous case in Sec. 3.3 that finding the value $\delta_{\max}$ such that $\langle r \rangle_{\delta_{\max}} = 0$ (corresponding to the maximally antireciprocal limit) is possible only as long as the overall link density (i.e. p) remains below the structural limit of $1/2$ (recall Fig. 13). At the same time, as mentioned in Sec. 3.2.1, strong heterogeneity in the fitness distribution makes it very difficult for the resulting network to

exhibit antireciprocal features, even when the overall density of links is very low. Indeed, the upper bound $\delta_{\max}$ in Eq. (3.15) is dictated by the pair of nodes with the lowest fitness as $\delta_{\max} = \min_{i,j} \delta_{i_0 j_0}^{\max}$, where $\delta_{i_0 j_0}^{\max} = -\frac{\ln(2e^{-\eta x_{i_0} x_{j_0}} - 1)}{x_{i_0} x_{j_0}}$ is a strictly increasing function of $x_{i_0} x_{j_0}$. As a consequence, the wider the gap between the minimum and maximum fitness products in a given realization, the broader the range of values of $\delta_{i_0 j_0}^{\max}$ across pairs of nodes and the larger the number of pairs of nodes with positive mutual connection probability $p_{i_0 j_0}^{\leftrightarrow}$ (contributing to an increase of the attainable value for $\langle r \rangle_{\min}$). In our infinite-mean fitness case, typical values of the fitness are spread so broadly over nodes that they effectively hinder the emergence of antireciprocal patterns in the network.

In analogy with our analysis in Sec. 3.3, we can evaluate the expected reciprocity $\langle r \rangle(\eta, \delta)$ in Eq. (3.19) for Lévy-distributed fitness values as a function of the parameter δ in the domain established by the conditions in Eq. (3.15), for different values of the parameter η (i.e. for a varying density of links). In Fig. 19 the outcome of this procedure is illustrated for four different values of the expected link density $\overline{\kappa} = \frac{1}{N(N-1)} \sum_{i,j \neq i} p_{ij}$. As the figure shows, maximal reciprocity ($\langle r \rangle = 1$) can always be achieved (specifically, when $\delta = \delta_{\min} = \eta$, yielding $p_{ij}^{\rightarrow} = 0$ and $p_{ij} = q_{ij}\ \forall\, i, j$). By contrast, $\langle r \rangle$ can never reach antireciprocated values compatibly with the conditions in Eq. (3.15): $\delta_{\max}$ is indeed too small to allow for global values of reciprocity below the areciprocal benchmark, due to the coexistence of nodes with very low fitness (which yield $\delta_{ij}^{\max} \approx 2\eta$) and nodes with very large fitness (which can establish unreciprocated links only for very large δ).

These results indicate that the annealed random networks that remain invariant under aggregations have non-negative reciprocity: they can only be areciprocal or reciprocal.

3.5.2 Expected topological properties

We now calculate the expected degrees (in-, out-, and reciprocated degrees) and their distributions. Note that the expected in-degree $k_{i_\ell}^{\text{in}}$, in this symmetrical setting, is equivalent to the expected out-degree $k_{i_\ell}^{\text{out}}$.

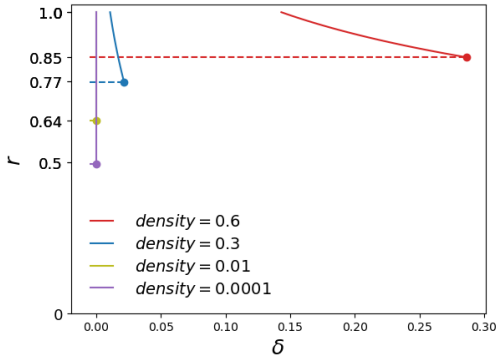

Figure 19: Reciprocity r in the annealed case as a function of the parameter δ. Solid lines illustrate, for four different values of link density ($\overline{\kappa} = 0.0001, 0.01, 0.3, 0.6$, corresponding to $\eta = 2.83\,e^{-10}, 1.83\,e^{-06}, 0.01, 0.14$) the behaviour of the function $\langle r \rangle$ in (3.19), while the full circles indicate the values achieved in the areciprocal case $\delta_{\text{rand}} = 2\eta$ (where $\langle r \rangle_{\text{rand}}$ is given by (3.20)). The dashed lines indicate the antireciprocal limit $\langle r \rangle_{\min} = \langle r \rangle_{\delta_{\max}}$, with $\delta_{\max} = \min_{i,j} \left(-\frac{\ln(2e^{-\eta x_i x_j} - 1)}{x_i x_j} \right)$. In each plot, δ varies in the range of its domain of existence $\delta \in [\eta, \delta_{\max}]$, which is dependent upon the specific realization of fitness values (which is fixed in all plots and amounting to 5000 values) and, clearly, on the parameter η (whose value is found so to enforce the desired density of links $\overline{\kappa}$).

Let us start by deriving the functional form of the expected in-degree distribution. In close resemblance with the analysis provided in Sec. 2.4, we compute, for a typical realization of the fitness values, the distribution $P_\ell(k^{\text{in}})$ from the PDF (3.28) of the fitness and from the connection probability p_{i_ℓ, j_ℓ} written as a function of the fitness of the nodes involved:

$$f(x, y) = 1 - e^{-\eta\, x\, y}. \tag{3.29}$$

For a large number N_ℓ of ℓ-nodes, the expected in-degree

$$\langle k^{\text{in}}_{i_\ell} \rangle = \sum_{j_\ell \neq i_\ell} p_{i_\ell j_\ell} = \sum_{j_\ell \neq i_\ell} f(x_{i_\ell}, x_{j_\ell})$$

can be approximated by an integral over the number $(N_\ell - 1)\varphi(y; \gamma_\ell)$ of ℓ-nodes (except i_ℓ itself) with fitness in a neighbourhood of y. By denoting

with $k_\ell^{\text{in}}(x)$ the expected in-degree of a node with fitness x at level ℓ, we find

$$\begin{aligned} k_\ell^{\text{in}}(x) &= (N_\ell - 1) \int_0^\infty f(x,y)\varphi(y;\gamma_\ell)\mathrm{d}y \\ &= (N_\ell - 1)\left(1 - e^{-\sqrt{2\eta\gamma_\ell x}}\right), \end{aligned} \tag{3.30}$$

a relationship confirmed via numerical simulations in the top left panel of Fig. 20, where the rescaled quantity $\kappa_\ell^{\text{in}}(x) = \frac{k_\ell^{\text{in}}(x)}{N_\ell - 1}$ is considered.

Inverting the above equation allows us to calculate the resulting expected in-degree distribution exactly. Indeed, by monotonicity, it holds:

$$P_\ell\left(k^{\text{in}}\right)\mathrm{d}k^{\text{in}} = \varphi\left(x(k^{\text{in}});\gamma_\ell\right)\mathrm{d}x(k^{\text{in}}). \tag{3.31}$$

We therefore obtain

$$P_\ell(k^{\text{in}}) = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\eta}\gamma_\ell}{N_\ell - 1 - k^{\text{in}}} \frac{e^{-\frac{\eta\gamma_\ell^2}{\log^2(1-\frac{k^{\text{in}}}{N_\ell - 1})}}}{\log^2(1 - \frac{k^{\text{in}}}{N_\ell - 1})}. \tag{3.32}$$

Note that this distribution has a power-law regime proportional to κ^{-2}, followed by a density-dependent cut-off. The top right panel of Fig. 20 illustrates the good agreement between numerical simulations and analytical predictions at four different levels of resolution.

Similarly, for evaluating the distribution $\tilde{P}_\ell(k^{\leftrightarrow})$, we consider the function:

$$f^{\leftrightarrow}(x,y) = 1 + e^{-\delta\, x\, y} - 2e^{-\eta\, x\, y}. \tag{3.33}$$

The expected reciprocal degree $k_\ell^{\leftrightarrow}$ of a node with fitness x, for a large number N_ℓ of nodes, is thus given by:

$$\begin{aligned} k_\ell^{\leftrightarrow}(x) &= (N_\ell - 1) \int_0^\infty f^{\leftrightarrow}(x,y)\varphi(y;\gamma_\ell)\mathrm{d}y \\ &= (N_\ell - 1)\left(1 - 2e^{-\sqrt{2\eta\gamma_\ell x}} + e^{-\sqrt{2\delta\gamma_\ell x}}\right). \end{aligned} \tag{3.34}$$

This is confirmed in the bottom left panel of Fig. 20. In this case, however, the obtained expression cannot be inverted unless a specific relation is

enforced between the parameters η and δ, i.e. $\delta = 4n\eta$, with $n = 1, 2, \ldots$ (this relation might in general be incompatible with the conditions in Eq. (3.15)). Nevertheless, we can estimate the cumulative distribution of rescaled reciprocated degrees by considering the behaviour of $k_\ell^{\leftrightarrow}(x)$ for small values of x, i.e. $k_\ell^{\leftrightarrow}(x) \sim (N_\ell - 1)\sqrt{2\gamma_\ell x}\ (2\sqrt{\eta} - \sqrt{\delta})$. In this regime the inverse function $x(k_\ell^{\leftrightarrow})$ can easily be found:

$$x(k_\ell^{\leftrightarrow}) \approx \frac{1}{2\gamma_\ell}\left(\frac{k}{(N_\ell - 1)(2\sqrt{\eta} - \sqrt{\delta}}\right)^2.$$

Although the above function is not monotone in k, we can still use Eq. (3.31) in our domain of interest, which is the positive real axis. The (approximated) reciprocal degree distribution therefore is given by

$$\tilde{P}_\ell(k^{\leftrightarrow}) \approx \frac{2}{\sqrt{\pi}}(N_\ell - 1)\frac{\gamma_\ell\left(2\sqrt{\eta} - \sqrt{\delta}\right)}{\left(k^{\leftrightarrow}\right)^2} e^{-\left((N_\ell - 1)\frac{\gamma_\ell(2\sqrt{\eta} - \sqrt{\delta})}{k^{\leftrightarrow}}\right)^2}. \tag{3.35}$$

As confirmed in the bottom right panel of Fig. 20, this approximate result reproduces the bulk of the distribution, while failing in the tails as expected (the approximation being valid only for small values of x).

Finally, we look at the expected density of links $\langle \bar{\kappa}_\ell \rangle$ (excluding self-loops) and the expected reciprocity $\langle r \rangle$. An expression for the density is found from

$$\begin{aligned} \langle \bar{\kappa}_\ell \rangle &= \int_0^\infty \mathrm{d}x \int_0^\infty \mathrm{d}y\, f(x, y)\varphi(x; \gamma_\ell)\varphi(y; \gamma_\ell) \\ &= 1 - \int_0^\infty e^{-\sqrt{2\eta\gamma_\ell x}}\varphi(x; \gamma_\ell)\mathrm{d}x, \end{aligned} \tag{3.36}$$

which can be written in terms of the Meijer-G function:

$$\langle \bar{\kappa}_\ell \rangle = 1 - \frac{\sqrt{\eta}\gamma_\ell}{2\pi} G^{3,0}_{0,3}\left(\begin{matrix}\cdot \\ -1/2, 0, 0\end{matrix}\middle| \frac{\eta\gamma_\ell^2}{4}\right). \tag{3.37}$$

Analogously, we can express the expected reciprocity as

$$\begin{aligned} \langle r_\ell \rangle &= \frac{\int_0^\infty \int_0^\infty f^{\leftrightarrow}(x, y)\varphi(x; \gamma_\ell)\varphi(y; \gamma_\ell)\mathrm{d}x\mathrm{d}y}{\langle \bar{\kappa}_\ell \rangle} \\ &= 2 - \frac{1 - \frac{\sqrt{\delta}\gamma_\ell}{2\pi} G^{3,0}_{0,3}\left(\begin{matrix}\cdot \\ -1/2, 0, 0\end{matrix}\middle| \frac{\delta\gamma_\ell^2}{4}\right)}{\langle \bar{\kappa}_\ell \rangle}. \end{aligned} \tag{3.38}$$

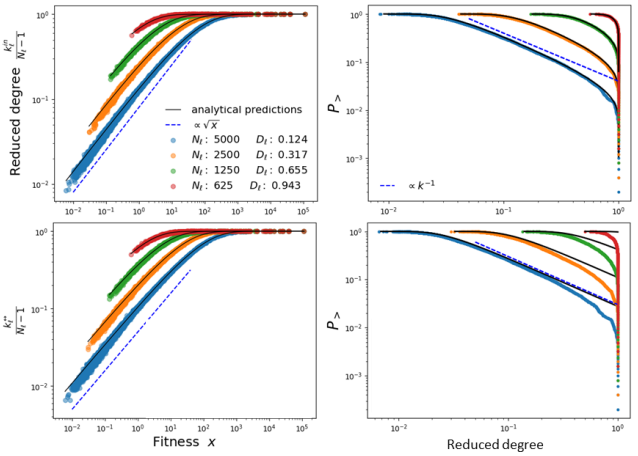

Figure 20: Degrees and reciprocal degrees in the annealed DSIM. We show the outcome of a simulation obtained from one realization of a synthetic network of $N_0 = 5000$ nodes at four different levels of aggregation ($\ell = 0, 1, 2, 3$). This is performed by drawing the network realization from the graph probability in (3.21) and then progressively coarse-graining it via the renormalization rules in (3.1), after having fixed the hierarchy of partitions in the most homogeneous way: at each step ℓ, N_ℓ blocks are created by randomly grouping nodes in pairs so that we progressively halve the number of nodes $N_\ell = N_0 2^{-\ell}$. In the two figures on the left, the reduced in-degree κ_ℓ^{in} (top) and reduced reciprocal degree $\kappa_\ell^{\leftrightarrow}$ (bottom) are shown as functions of the fitness values and compared to the analytical predictions in (3.30) and (3.5.2). In the two figures on the right, the cumulative distributions of the reduced in-degree κ^{in} and reduced reciprocal degree $k^{\leftrightarrow}$ are compared with the analytical expectations shown in equations (3.34) and (3.35). As a reference, the dashed lines in the right panels are proportional to κ^{-1}, corresponding to κ^{-2} for the non-cumulative pdf.

We display the above analytical results in Fig. 21, where the functional dependence, shown in Eqs. (3.37) and (3.38), of $\langle \overline{\kappa}_\ell \rangle$ and $\langle r_\ell \rangle$ on the scale parameter γ_ℓ is illustrated, along with the respective values obtained at four different levels of aggregation of a synthetic network drawn from the graph probability in (3.21) and then progressively coarse-grained by

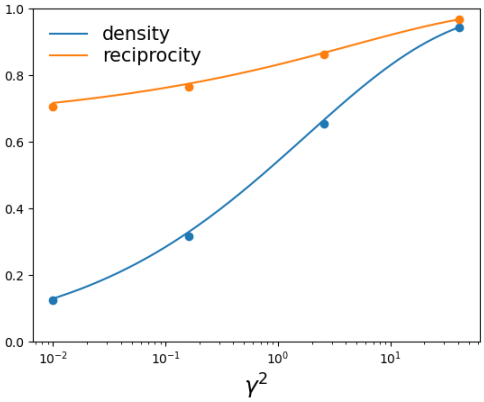

Figure 21: Link density and reciprocity along the coarse-graining flow. The blue and the orange lines correspond to the analytical predictions, respectively, of the link density $\langle \bar{\kappa}_\ell \rangle$ and reciprocity $\langle r_\ell \rangle$ in equations (3.37)-(3.38). The full circles correspond to the realized values of $\langle \bar{\kappa}_\ell \rangle$ and $\langle r_\ell \rangle$ in a simulation where a synthetic network of $N_0 = 5000$ nodes is generated at level $\ell = 0$ and then progressively coarse-grained. At each iteration, nodes are randomly merged in pairs so that the parameter γ_ℓ from level ℓ to $\ell + 1$ scales as $\gamma_{\ell+1} = 2^{2\ell}\gamma_0$. Parameters η and δ are fixed, respectively, to 0.1 and $\delta_{max} = 0.20000037$ (the latter yielding the minimally reciprocated networks, given $\eta = 0.1$ and $\{x_{i_0}\}$).

merging pairs of randomly chosen nodes ($b_\ell = 2$). Again, numerical simulations agree perfectly with the analytical calculations.

3.6 Discussion

In this chapter we have shown that the Scale-Invariant Model, proposed as an approach to consistently model undirected networks at arbitrary (and possibly multi-scale) resolution levels, can be extended coherently to directed networks, leading to the Directed Scale-Invariant Model. Importantly, besides being scale-invariant, the DSIM can account for nontrivial reciprocity.

As in the undirected formulation, the model can generate any (directed) network, equiprobably, either hierarchically (by generating a finer-

grained network and then coarse-graining it through a sequence of nested partitions) or directly (using appropriately renormalized parameters). In the directed formulation, these parameters include three global parameters (meant to tune the overall density of directed links, reciprocated links and self-loops), four sets of fitness variables (each quadruplet being attached to each (block-)node to separately trigger its tendency to establish connections with other nodes - either incoming, outgoing, or reciprocated - and with itself) and, if useful, a set of dyadic factors representing distances or communities. A sufficient condition to ensure that the connection probabilities are indeed scale-invariant even in the presence of dyadic interactions is to consider *metaultrametric* distances compatible with the hierarchy of partitions chosen to define the coarse-graining procedure.

After deriving the renormalization rules for networks with arbitrary reciprocity, we have considered various examples, including a directed multiscale model of the international trade network and an annealed model, where scale-free networks with positive reciprocity emerge spontaneously from coarse-graining. Applying the model to the directed ITN demonstrated its consistency with the undirected formalism (accurately reproducing topological features of the undirected projection of the ITN), while at the same time capturing nontrivial directed topological and spectral properties related to strong reciprocity patterns across hierarchical aggregation levels.

In the annealed scenario, we focused on a specific case, where the connection probabilities are symmetric (in that the different fitness parameters assigned to each node are taken to be identical) and the fitness values are drawn from a Lévy distribution. The high heterogeneity of this distribution suppresses the antireciprocal regime, suggesting that annealed scale-invariant random graphs are intrinsically reciprocal. In such a setting, we have shown that in-degree, out-degree and reciprocal degree distributions feature a power-law decay $\propto k^{-2}$ followed by a density-dependent cutoff. Therefore, the sole requirement of scale-invariance automatically produces directed scale-free graphs with positive reciprocity.

Future research should address further generalizations, such as higher-

dimensional quenched fitness values and multivariate α-stable distributions for the different annealed fitness types.

3.7 Supporting material

3.7.1 GDP, distance and Trade data

Here we specify the procedure used to gather trade, GDP and distance data in our analysis of the International Trade Network (ITN).

For defining the fitness values and the empirical network we relied on the expanded trade dataset developed by K. S. Gleditsch [99], which provides estimates of the bilateral trade flows between independent states between the years from 1948 to 2000, together with estimates of their Gross Domestic Products (GDPs) per capita and population. The GDP estimates are originally based on figures from Penn World Tables [131], produced by the Center for International Comparisons at the University of Pennsylvania, and from reports of Central Intelligence Agency (CIA) [132]. The outcome consists of 185 countries, whose *real GDPs per capita* are given for all years in constant US dollars (with base 1996), along with population figures in units of 1000s. We obtained the real *total GDP* value of each country by multiplying its GDP value per capita by its population. Such quantities are naturally additive (i.e., scale linearly with the population) and are thus suitable to serve as nodes' fitness.
For defining the binary asymmetric matrix $\tilde{\mathbf{A}}^{(0)}$ representing our empirical ITN at level $\ell = 0$, we used the final estimates of the international trade flows provided in Gleditsch database, obtained from a refinement of the Direction of Trade (DOT) of the International Monetary Fund (IMF) data [133] where missing data and suspicious zero had been managed with figures from the World Export Data (WED) [134] or through ad hoc procedures (for which we refer to the original reference [99]). As described in Sec. 3.4, if a positive trade is estimated in the considered year (2000) from country i_0 to country j_0 then we posit $a_{i_0 j_0} = 1$, otherwise $a_{i_0 j_0} = 0$.

In accordance with the analysis of the undirected ITN (reported in Sec. 2.3), we set the dyadic parameter of the model building on the

population-averaged geographic distances provided within the BACI-CEPII GeoDist database [75], extensively described in Sec. 2.6.2 (specifically, they were defined in the formula (2.19)). As explained in Sec. 3.4, in this (directed) case the considered distances are required to be ultrametric (or, more in general, metaultrametric) allowing us to define the coarse-graining flow of the network by progressively cutting the associated dendrogram at always larger heights while keeping invariant the pairwise parameter. To minimize the distortion between the original inter-country distance d_{i_0,j_0} and its metaultrametric approximation $d^{\mathrm{mu}}_{i_0,j_0}$, we consider the subdominant ultrametric distance, i.e., the largest among all ultrametrics that are less than or equal to the original metric. These can be obtained via a single-linkage hierarchical clustering performed by using the original distances $d_{i_0 j_0}$ as measure of dissimilarity. Hierarchical clustering algorithms start from the partition of the dataset into singleton nodes and merge step by step the current pair of mutually closest nodes into a new node until there is one final node left. Several clustering schemes share this procedure as a common definition, but differ in the way in which the measure of inter-cluster dissimilarity is updated after each step. The single-linkage method, which has been already recalled in Sec. 2.3 and 2.6.2, uses the minimum of the distances between all observations of the two sets and is indeed known to produce the *subdominant ultrametric* distance of the original one [56, 135, 136]. The output of this algorithm is perfectly equivalent to the dendrogram shown in Fig. 10, whose leaves correspond to each country (each 0-node) and the subdominant ultrametric distance between each pair is given by the height of the branching point of the corresponding branches. This generates clusters of nodes that are equally distant, in terms of d^{u}, one from each other. For instance, given two clusters I and J, then the subdominant ultrametric distance between each pair of nodes belonging to I and J is given by $d^{\mathrm{u}}_{i_0 j_0} = min_{i_0 \in I, j_0 \in J} d_{i_0 j_0}$. Starting from the ultrametric distances $d^{\mathrm{u}}_{i_0 j_0}$, we define the metaultrametric distances (related to a metametric obeying the ultrametricity condition) as

$$d^{\mathrm{mu}}_{i_0 j_0} = \begin{cases} d^{\mathrm{u}}_{i_0 j_0} & \text{if } i_0 \neq j_0, \\ d_{i_0 j_0} & \text{if } i_0 = j_0, \end{cases} \tag{3.39}$$

where $d_{i_0 j_0}$ is the population averaged distance between countries i_0 and j_0 in Eq. (2.60).

3.7.2 (Directed) network properties: empirical and expected values

In this section, the key topological properties considered in our analysis of the directed ITN are defined. By keeping the notation adopted in the undirected case (see Sec. 2.6.3), each such property is a function $Y(\mathbf{A}^{(\ell)})$ of the $N_\ell \times N_\ell$ adjacency matrix $\mathbf{A}^{(\ell)}$ (with entries $a^{(\ell)}_{i_\ell,j_\ell} = 0, 1$) of the generic ℓ-graph. This matrix is in general *asymmetric* and can in principle contain non-zero entries along the diagonal, representing self-loops.
As made explicit in Sec. 3.4, $\tilde{\mathbf{A}}^{(\ell)}$ represents the empirical matrix obtained at the hierarchical level ℓ from the Gleditsch data in year 2000. The observed value of each considered topological property Y of the ITN will be denoted as $\tilde{Y} \equiv Y(\tilde{\mathbf{A}}^{(\ell)})$.
When considering the multiscale model, $\mathbf{A}^{(\ell)}$ is instead a random matrix whose entries $\{a^{(\ell)}_{i_\ell,j_\ell}\}$ are Bernoulli random variables with expected value

$$\langle a^{(\ell)}_{i_\ell,j_\ell} \rangle \equiv p_{i_\ell j_\ell} = \begin{cases} 1 - e^{-\eta \frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d^{\mathrm{mu}}_{i_\ell,j_\ell}}} & \text{if} \quad i_\ell \neq j_\ell \\ 1 - e^{-\frac{\delta}{2}\frac{\mathrm{GDP}^2_{i_\ell}}{d^{\mathrm{mu}}_{i_\ell,i_\ell}} - \tilde{\sigma}\tilde{w}_{i_\ell}} & \text{if} \quad i_\ell = j_\ell \end{cases} \tag{3.40}$$

where the parameters $(\tilde{\sigma}, \tilde{w}_{i_0})$ are chosen so to ensure $p_{i_0 i_0}(\eta, \delta, \tilde{\sigma}) = 0 \quad \forall\, i_0 = 1, N_0$ (in accordance with the adopted convention to fix $\tilde{a}^{(0)}_{i_0 i_0} = 0\, \forall\, i_0$). As explained in Sec. 3.4, this is obtained by imposing $\tilde{\sigma} = -\frac{\delta}{2}$ and $\tilde{w}_{i_0} = \frac{\mathrm{GDP}^2_{i_0}}{d^{\mathrm{mu}}_{i_0 i_0}}\ \forall\, i_0 = 1, \ldots, N_0$, which implies:

$$\tilde{w}_{i_\ell} = \sum_{i_0 \in i_\ell} \frac{\mathrm{GDP}^2_{i_0}}{d^{\mathrm{mu}}_{i_0 i_0}} \quad \forall \ell.$$

Note that even though at level $\ell = 0$ either the empirical network and the multiscale model are by our choice exhibiting null self-loops, as we proceed with the coarse-graining, these will emerge in both the former, as a consequence of the coarse-graining rule $a_{i_\ell j_\ell} = 1 -$

$\prod_{i_{\ell-1}\in i_\ell}\prod_{j_{\ell-1}\in j_\ell}\left(1-a_{i_{\ell-1}j_{\ell-1}}\right)$, and in the latter, in that the parameters $\tilde{\sigma}\tilde{w}_{i_\ell}$ at aggregate levels do not counterbalance anymore the first term at the exponent, which cannot equal 0 for $\ell > 0$:

$$\begin{aligned}-\frac{\delta}{2}\frac{\mathrm{GDP}_{i_\ell}^2}{d_{i_\ell i_\ell}^{up}}-\tilde{\sigma}\tilde{w}_{i_\ell} &= -\frac{\delta}{2}\sum_{i_0\in i_\ell}\sum_{j_0\in i_\ell}\frac{\mathrm{GDP}_{i_0}\mathrm{GDP}_{j_0}}{d_{i_0 j_0}^{\mathrm{mu}}}-\tilde{\sigma}\tilde{w}_{i_\ell}\\ &= \tilde{\sigma}\sum_{i_0\in i_\ell}\sum_{j_0\in i_\ell, j_0\neq i_0}\frac{\mathrm{GDP}_{i_0}\mathrm{GDP}_{j_0}}{d_{i_0 j_0}^{\mathrm{mu}}}\\ &\neq 0.\end{aligned}$$

Then, we consider the undirected projection $\mathbf{B}^{(\ell)}$ of the matrix $\mathbf{A}^{(\ell)}$, whose elements $b_{i_\ell,j_\ell}^{(\ell)} \equiv a_{i_\ell,j_\ell}^{(\ell)} + a_{j_\ell,i_\ell}^{(\ell)} - a_{i_\ell,j_\ell}^{(\ell)}a_{j_\ell,i_\ell}^{(\ell)}$ are Bernoulli random variables with expected values:

$$\langle b_{i_\ell,j_\ell}^{(\ell)}\rangle \equiv q_{i_\ell j_\ell} = \begin{cases} 1-e^{-\delta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}} & i_\ell \neq j_\ell,\\ 1-e^{-\frac{\delta}{2}\frac{\mathrm{GDP}_{i_\ell}^2}{d_{i_\ell,i_\ell}^{\mathrm{mu}}}-\tilde{\sigma}\tilde{w}_{i_\ell}} & i_\ell = j_\ell.\end{cases} \tag{3.41}$$

The probabilities of the four possible dyads between each pair of (block-)countries i_ℓ and j_ℓ (namely an unidirectional connection, a mutual connection or a missing connection) are given by:

$$\begin{aligned}
p_{i_\ell j_\ell}^{\rightarrow} &= \begin{cases} e^{-\eta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}}-e^{-\delta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}} & i_\ell \neq j_\ell,\\ 0 & i_\ell = j_\ell,\end{cases}\\
p_{i_\ell j_\ell}^{\leftarrow} &= \begin{cases} e^{-\eta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}}-e^{-\delta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}} & i_\ell \neq j_\ell,\\ 0 & i_\ell = j_\ell,\end{cases}\\
p_{i_\ell j_\ell}^{\leftrightarrow} &= \begin{cases} 1-2e^{-\eta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}}+e^{-\delta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}} & i_\ell \neq j_\ell,\\ 1-e^{-\frac{\delta}{2}\frac{\mathrm{GDP}_{i_\ell}^2}{d_{i_\ell,i_\ell}^{\mathrm{mu}}}-\tilde{\sigma}\tilde{w}_{i_\ell}} & i_\ell = j_\ell,\end{cases}\\
p_{i_\ell j_\ell}^{\nleftrightarrow} &= \begin{cases} e^{-\delta\frac{\mathrm{GDP}_{i_\ell}\mathrm{GDP}_{j_\ell}}{d_{i_\ell,j_\ell}^{\mathrm{mu}}}} & i_\ell \neq j_\ell,\\ e^{-\frac{\delta}{2}\frac{\mathrm{GDP}_{i_\ell}^2}{d_{i_\ell,i_\ell}^{\mathrm{mu}}}-\tilde{\sigma}\tilde{w}_{i_\ell}} & i_\ell = j_\ell.\end{cases}
\end{aligned} \tag{3.42}$$

are calibrated by taking their unique values $\tilde{\eta}$ and $\tilde{\delta}$ that produce, respectively, the same number $L_\ell^{\text{nl}} = \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell=1, j_\ell \neq i_\ell}^{N_\ell} a_{i_\ell j_\ell}^{(\ell)}$ of links and the same number $L_\ell^{\leftrightarrow} = \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell=1, j_\ell \neq i_\ell}^{N_\ell} a_{i_\ell j_\ell}^{\leftrightarrow\,(\ell)}$ of reciprocated links (or, equivalently, the number $L_\ell^{\text{und,nl}} = \sum_{i_\ell=1}^{N_\ell} \sum_{j_\ell<i_\ell}^{N_\ell} b_{i_\ell j_\ell}^{(\ell)} = L_\ell^{\text{nl}} - 1/2\, L_\ell^{\leftrightarrow}$ of undirected links) as the real ITN (self-loops excluded). In principle, this can be done at an arbitrary level ℓ. Here, we are reporting the results obtained by constraining the above quantities at the native level, i.e. $\ell = 0$, which reads:

$$\begin{aligned} \sum_{i_0=1}^{N_0} \sum_{j_0=1, j_0 \neq i_0}^{N_0} p_{i_0 j_0}(\tilde{\eta}) &= \tilde{L}_0^{\text{nl}}, \\ \sum_{i_0=1}^{N_0} \sum_{j_0=1, j_0 \neq i_0}^{i_0} q_{i_0 j_0}(\tilde{\delta}) &= \tilde{L}_0^{\text{und,nl}}, \end{aligned} \tag{3.43}$$

where the terms on the left side identify the expected values of links $\langle L_0^{\text{nl}} \rangle(\eta)$ and undirected links $\langle L_0^{\text{und,nl}} \rangle(\delta)$ (*not* accounting for self-loops), while the terms on the right correspond to their empirical values. In the considered dataset, we find $\tilde{L}_0^{\text{nl}} = 20101$ and $\tilde{L}_0^{\text{und,nl}} = 10250$, yielding $\tilde{\eta} = 1.211\,10^{-11}(USD)^{-2}$ and $\tilde{\delta} = 1.344\,10^{-11}(USD)^{-2}$.

Once the parameters η and δ are fixed to $\tilde{\eta}$ and $\tilde{\delta}$, we can generate realizations $\{\mathbf{A}^{(\ell)}\}$ of the ℓ-graphs from the multiscale model at any desired hierarchical level ℓ by sampling ℓ-links independently with probabilities $p_{i_\ell j_\ell}^{\rightarrow}(\tilde{\eta}, \tilde{\delta}), p_{i_\ell j_\ell}^{\leftarrow}(\tilde{\eta}, \tilde{\delta})$ and $p_{i_\ell j_\ell}^{\leftrightarrow}(\tilde{\eta}, \tilde{\delta}, \tilde{\sigma})$. By averaging the value $Y(\mathbf{A}^{(\ell)})$ of a topological property over such realizations, we can efficiently estimate the corresponding expected value

$$\langle Y \rangle \equiv \sum_{\mathbf{A}^{(\ell)} \in \mathcal{G}_{N_\ell}} P\big(\mathbf{A}^{(\ell)}\big) Y(\mathbf{A}^{(\ell)}), \tag{3.44}$$

where $P\big(\mathbf{A}^{(\ell)}\big)$ is given by Eq. (3.21). If $Y(\mathbf{A}^{(\ell)})$ is linear in $\mathbf{A}^{(\ell)}$, we can calculate $\langle Y \rangle$ exactly by directly replacing, in the definition of $Y(\mathbf{A}^{(\ell)})$, $a_{i_\ell, j_\ell}^{\rightarrow}$ with $\tilde{p}_{i_\ell, j_\ell}^{\rightarrow}$, $a_{i_\ell, j_\ell}^{\leftarrow}$ with $\tilde{p}_{i_\ell, j_\ell}^{\leftarrow}$ and $a_{i_\ell, j_\ell}^{\leftrightarrow}$ with $\tilde{p}_{i_\ell, j_\ell}^{\leftrightarrow}$, without sampling any graph at all. This is indeed the case for the number of total and reciprocal links in Eq. (3.43).

Given any ℓ-graph $\mathbf{A}^{(\ell)}$ (be it the empirical ℓ-graph or a random realization from the ensemble), the main topological properties of interest to us are listed in the following.

- The *link density*, representing the ratio of realized to the maximum number of links, including possible self-loops:

$$D_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{2L_\ell(\mathbf{A}^{(\ell)})}{N_\ell(N_\ell+1)} = \frac{2\sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1}^{i_\ell} a_{i_\ell,j_\ell}^{(\ell)}}{N_\ell(N_\ell+1)}; \tag{3.45}$$

- the *reciprocity*, quantifying the proportion of mutual links with respect to the total number of links (excluding self-loops to ensure that $r_\ell(\mathbf{A}^{(\ell)}) \in [0,1]$):

$$r_\ell(\mathbf{A}^{(\ell)}) \equiv \frac{L_\ell^{\leftrightarrow}(\mathbf{A}^{(\ell)})}{L_\ell^{nl}(\mathbf{A}^{(\ell)})} = \frac{\sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1,j_\ell\neq i_\ell}^{N_\ell} a_{i_\ell,j_\ell}^{\leftrightarrow\ (\ell)}}{\sum_{i_\ell=1}^{N_\ell}\sum_{j_\ell=1,j_\ell\neq i_\ell}^{N_\ell} a_{i_\ell,j_\ell}{}^{(\ell)}}; \tag{3.46}$$

- the *in-degree* (or, equivalently, the *out-degree* upon exchange of the indices), counting the number of incoming (or outgoing) links to the ℓ-node i_ℓ (excluding self-loops):

$$k_{i_\ell}^{\text{in}}(\mathbf{A}^{(\ell)}) \equiv \sum_{j_\ell\neq i_\ell} a_{i_\ell,j_\ell}^{(\ell)}; \tag{3.47}$$

- the *undirected degree*, counting the number of connections in the undirected projection $\mathbf{B}^{(\ell)}$:

$$k_{i_\ell}^{\text{und}}(\mathbf{B}^{(\ell)}) = \sum_{j_\ell\neq} b_{i_\ell j_\ell}; \tag{3.48}$$

- the *average nearest neighbour out-degree*, representing the average out-degree of the neighbours of i_ℓ:

$$k_{\text{nn}\ i_\ell}^{\text{out}}(\mathbf{A}^{(\ell)}) \equiv \frac{\sum_{j_\ell\neq i_\ell}\sum_{k_\ell\neq j_\ell} a_{i_\ell,j_\ell}^{(\ell)} a_{j_\ell,k_\ell}^{(\ell)}}{\sum_{j_\ell\neq i_\ell} a_{i_\ell,j_\ell}^{(\ell)}}. \tag{3.49}$$

Chapter 4

SIM as an inhomogeneous random graph with infinite-mean fitness variables

This chapter is based on the work [2] by M. Lalli, R. Hazra, D. Garlaschelli, and L. Avena, which provides an investigation of the annealed SI model with a mathematically rigorous perspective.
In what follows, we consider a slightly different model from the one introduced in Chap. 2, as will be highlighted through a different notation. To facilitate the reading, we will start by briefly outlining the new setting and framing it within the context of Probability Theory's evolving literature, which offered the prompt for our analysis. The relationship with the original model and implications of our findings on it will be discussed in Sec. 4.2, while a complete definition of the new setting together with the achieved results is contained in Sec. 4.3. Concluding results are in Sec. 4.6.

4.1 Introduction

Consider a random graph with n nodes, where the vertices are assigned independent weights $(W_i)_{i\in[n]}$ (or fitness variables), which determine the connections between pairs of nodes. The values of the weights are drawn from a common distribution $F_W(\cdot)$, with $1 - F_W(x) \sim x^{-\alpha}$ for some $\alpha \in (0, 1)$. Therefore the weights have infinite mean. Conditioned on the weights, an edge between two distinct vertices i and j is drawn independently with probability

$$p_{ij} = 1 - e^{-\varepsilon W_i W_j}, \tag{4.1}$$

where ε is a positive n-dependent parameter tuning the overall density of edges in the graph and playing a crucial role in the analysis of the model that will come in the present chapter.

The above model corresponds to a particular class of inhomogeneous Erdős - Rényi random graphs, so called in continuation with the foundational Erdős - Rényi random graph [7] which, on the opposite, is homogeneous as all nodes contribute equivalently to the network's connectivity, resulting in a *non-heterogeneous* degree distribution. More in general, Eq. (4.1) represents a special example of connection probability between vertices i and j defined as $\kappa_n(W_i, W_j)$, where $\kappa_n : [0, \infty)^2 \to [0, 1]$ is a well-behaved function, and the weights are drawn independently from a certain distribution. These class of random graphs, which in the physics literature are usually denoted as *fitness* or *hidden variables* models ([20, 137]) find a well-known instance, in the mathematical literature, in the Generalized Random Graph [5]. In most of the cases considered in the literature so far, due to the integrability conditions on κ_n and moment properties of F_W, these models have a locally tree-like structure. The interested reader may find an accessible review of related results in Chapter 6 of [5] for the properties of the degree distribution, and in [55, 138] for further geometric structures. Models based on connection probabilities that resemble almost exactly Eq. (4.1) had been proposed, but with relevant differences regarding either the scaling of the prefactor at the exponent or the distribution of the fitness variables, which most often has finite mean

([58, 139, 140]). Our precise setting was revealed to be non-standard as the analysis of its topological properties defies usual techniques, due to the combination of the specific form of the connection probability and these heavy-tailed weights produced by the choice $\alpha \in (0, 1)$. We believe that certain mathematical features of an *ultra-small world* network, where the degrees exhibit infinite variance, can be captured by this model. In this case, due to the absence of a finite mean, the typical distances may be much slower than the doubly-logarithmic behaviour (in relation to the graph's size) observed in ultra-small networks [141]. Here, we show that the average degree in the considered model grows like $\log n$ if we choose the specific scaling $\varepsilon \sim n^{-\frac{1}{\alpha}}$. In this case, the cumulative degree distribution roughly behaves like a power law with exponent -1. In the literature for random graphs with degree sequences having power-law distribution with exponent τ, this falls in the critical case of exponent $\tau = 1$ [5]. In particular, the configuration model with degree distribution being a power law with exponent $\tau \in (0, 1)$ was studied in [142], where it was shown that the typical distance between two randomly chosen nodes in the graph is either 2 or 3. In addition, they found that in the critical case $\tau = 1$ a similar ultra-small world behaviour holds true. Note that our model differs from this one in that it naturally gives rise to degree distribution with power law exponent -1 rather than imposing it.
Additionally, we investigate certain dependencies between the degrees of different vertices and the asymptotic density of wedges and triangles.

The rest of the chapter is organized as follows: in Sec. 4.3 we state our main results, in Sec. 4.2 we discuss the connection to the original model and, ultimately, in Sections 4.4 and 4.5 we prove our results.

4.2 Connection with the original SIM

Despite the close resemblance between the model studied in this chapter and the original annealed version of the SIM introduced in Sec. 2.4 (in particular, between Eqs. (2.17) and (4.1)), some differences require further discussion.
First, to make the calculations less irksome at the first stage, we have

here considered weights drawn from a Pareto distribution with tail exponent α and fixed scale parameter 1, rather than a one-sided α-stable distribution with scale parameter γ_ℓ (which is ℓ-dependent and therefore n-dependent); second, to approach the asymptotic analysis we have here considered an n-dependent density-parameter ε, while in the original model the density-parameter δ was scale-invariant; third, here we have not exploited the scale-invariant nature of the SIM under coarse-graining. We now clarify the close relationship between the two variants of the model, these apparent differences notwithstanding.

Let us stat by recalling that α-stable distributions belongs to the broader family of regularly varying random variables with power law exponent $-\alpha$, i.e. random variables with tail such that $\mathbf{P}(W_i > w) = w^{-\alpha} L(w)$ where $L(\cdot)$ is a slowly varying function, that is, for any $w > 0$,

$$\lim_{t\to\infty} \frac{L(wt)}{L(t)} = 1.$$

More precisely, for large values of the argument x, a one-sided α-stable distribution $\mathbf{P}(X > x)$ with scale parameter γ is well approximated by a pure power law (Pareto) distribution $\mathbf{P}(X > x) \sim C_{\alpha,\gamma}\, x^{-\alpha}$ with a prefactor $C_{\alpha,\gamma}$ that depends on the parameters of the stable law [143]:

$$C_{\alpha,\gamma} \equiv \gamma^\alpha c_\alpha, \quad \text{with} \quad c_\alpha \equiv \frac{2\Gamma(\alpha)}{\pi} \sin\frac{\pi\alpha}{2}. \tag{4.2}$$

It is our belief that most of the results stated in this chapter for pure Pareto will go through also in presence of a slowly varying function, even if the analysis will be more involved. However, we haven't gone through this technical side yet.

Concerning the shift in moving from a Pareto with scale $c_\alpha \gamma^\alpha$ to a Pareto with scale 1, this implies a shift for the global prefactor at the exponent of the connection probability which immediately relates the results obtained here to those obtained in the original model at a given level of aggregation of nodes. Indeed we note that, besides n_0 and b, the three remaining parameters of the original annealed SIM are $\alpha \in (0,1)$, $\gamma_0 \in (0,\infty)$ and $\delta \in (0,\infty)$. However, among these three parameters, only

α and the combination $\delta\gamma_0^2$ are independent. Indeed, it is easy to realize that rescaling γ_0 to γ_0/λ (which is equivalent to rescaling X_{i_ℓ} to X_{i_ℓ}/λ, for some $\lambda > 0$) while simultaneously rescaling δ to $\lambda^2\delta$ leaves the connection probability unchanged. In combination with the scale-invariant nature of the SIM, this property can be exploited to map the quantities $(\{X_{i_\ell}\}_{i_\ell=1}^{n_\ell}, \delta)$ introduced in the original model to the quantities $(\{W_i\}_{i=1}^{n}, \varepsilon)$ used here by choosing a level ℓ such that the number n_ℓ of vertices in the SIM equals the one desired here, i.e. $n = n_\ell$, and defining the weight of vertex i as $W_i \equiv c_\alpha^{-1/\alpha}\gamma_\ell^{-1}X_{i_\ell}$ for $i = 1, n$. From this redefinition, it follows:

$$\lim_{x\to\infty} \mathbf{P}(W_i > x)\, x^\alpha = 1 \qquad i = 1, n,$$

irrespective of ℓ. This implies that, while the distribution of X_{i_ℓ} in Eq. (4.2) depends on ℓ through the parameter γ_ℓ, the distribution of W_i is actually ℓ-independent in the tail, which is the reason why we could drop the subscript ℓ in redefining W_i. In this way, we can keep the connection probability unchanged (i.e. $1 - e^{-\delta X_{i_\ell} X_{j_\ell}} \equiv 1 - e^{-\varepsilon_{n_\ell} W_i W_j}$) while moving the scale-dependence from $\{X_{i_\ell}\}_{i_\ell=1}^{n_\ell}$ to ε_{n_ℓ} by redefining the latter in one of the following equivalent ways:

$$\varepsilon_{n_\ell} \equiv c_\alpha^{2/\alpha}\gamma_\ell^2\delta = c_\alpha^{2/\alpha}\left(\frac{n_0}{n_\ell}\right)^{2/\alpha}\gamma_0^2\delta = c_\alpha^{2/\alpha}b^{2\ell/\alpha}\gamma_0^2\delta = b^{2\ell/\alpha}\varepsilon_{n_0} \qquad (4.3)$$

where $\varepsilon_{n_0} \equiv c_\alpha^{2/\alpha}\gamma_0^2\delta$.

In other words, our formulation here can be thought of as deriving from an equivalent SIM where, rather than having a scale-dependent fitness distribution and a scale-independent global parameter δ, we have a scale-independent fitness distribution (with asymptotically the same tail as the Pareto in Eq. (4.2)) and a scale-dependent global parameter $\varepsilon_n = \varepsilon_{n_\ell}$, for an implied hierarchical level ℓ. According to Eq. (4.3), since δ, α and b are finite constants, achieving a certain scaling of ε_n with n in the model considered here corresponds to achieving a corresponding scaling of γ_ℓ with n_ℓ (or equivalently of γ_0 with n_0), or ultimately to finding an appropriate ℓ, in the original model. *Results that we obtain for a specific range of values of ε can therefore be thought of as applying to a*

corresponding specific range of hierarchical levels in the original model.

4.3 Model and main results

The formal definition of the model considered here reads as follows. Let the vertex set be given by $[n] = \{1, 2, \ldots, n\}$ and let $\varepsilon = \varepsilon_n > 0$ be a parameter which will depend on n. The random graph with law $\mathbf{P}$ is constructed in the following way:

(a) Sample n independent weights (W_i), under $\mathbf{P}$, according to a Pareto distribution with parameter $\alpha \in (0, 1)$, that is,

$$1 - F_W(w) = \mathbf{P}(W_i > w) = \begin{cases} w^{-\alpha}, & w > 1, \\ 1, & 0 < w \leq 1. \end{cases} . \tag{4.4}$$

(b) For all $n \geq 1$, given the weights $(W_i)_{i \in [n]}$, construct the random graph G_n by joining edges independently with probability given by (4.1). That is,

$$p_{ij} := \mathbf{P}(i \leftrightarrow j \mid W_i, W_j) = 1 - e^{-\varepsilon_n W_i W_j}, \tag{4.5}$$

where the event $\{i \leftrightarrow j\}$ means that vertices i and j are connected by an edge in the graph.

The above random graph will be denoted by $\mathbf{G}_n(\alpha, \varepsilon)$ as it depends on the parameters α and ε. Self-loops and multi-edges are not allowed and hence the final graph is given by a simple graph on n vertices.

Notation Convergence in distribution and convergence in probability will be denoted respectively by $\xrightarrow{d}$ and $\xrightarrow{P}$. $\mathbf{E}[\cdot]$ is the expectation with respect to $\mathbf{P}$ and the conditional expectation with respect to the weight W of a typical vertex is denoted by $\mathbf{E}_W[\cdot] = \mathbf{E}[\cdot|W]$. We write $X|W$ to denote the distribution of the random variable, conditioned on the variable W.

Let $(a_{ij})_{1\leq i,j\leq n}$ be the indicator variables $(\mathbb{1}_{i\leftrightarrow j})_{1\leq i,j\leq n}$.
As standard, as $n\to\infty$, then $f(n)\sim g(n)$ stands for $f(n)/g(n)\to 1$, while $f(n)=o\left(g(n)\right)$ is equivalent to $f(n)/g(n)\to 0$, and $f(n)=O\left(g(n)\right)$ to $f(n)/g(n)\leq C$ for some $C>0$, for n large enough. Lastly, $f(n)\asymp g(n)$ denote that there exists positive constants c_1 and C_2 such that

$$c_1\leq \liminf_{n\to\infty} f(n)/g(n)\leq \limsup_{n\to\infty} f(n)/g(n)\leq C_2.$$

4.3.1 Degrees

Our first theorem characterizes the behaviour of a typical degree and the joint distribution of the degrees. Consider the degree of vertex $i\in[n]$ defined as

$$D_n(i)=\sum_{j\neq i} a_{ij}, \tag{4.6}$$

where a_{ij} denotes the entries of the adjacency matrix of the graph, that is,

$$a_{ij}=\begin{cases}1 & \text{if } i\leftrightarrow j\\ 0 & \text{otherwise.}\end{cases}$$

Theorem 1. Scaling and asymptotic of the degrees. *Consider the graph $\mathbf{G}_n(\alpha,\varepsilon)$ and let $D_n(i)$ be the degree of the vertex $i\in[n]$.*

(i) **Expected degree.** *The expected degree of a vertex i scales as follows*

$$\mathbf{E}\left[D_n(i)\right]\sim -(n-1)\Gamma(1-\alpha)\varepsilon^{\alpha}\log\varepsilon^{\alpha},\ \textit{as } \varepsilon\downarrow 0.$$

In particular, if $\varepsilon_n=n^{-1/\alpha}$ then we have

$$\mathbf{E}[D_n(i)]\sim\Gamma(1-\alpha)\log n \textit{ as } n\to\infty. \tag{4.7}$$

(ii) **Asymptotic degree distribution.** *Let $\varepsilon_n=n^{-\frac{1}{\alpha}}$, then for all $i\in[n]$*

$$D_n(i)\xrightarrow{d} D_\infty \textit{ as } n\to\infty,$$

where D_∞ is a mixed Poisson random variable with parameter $\Lambda=\Gamma(1-\alpha)W^{\alpha}$, where W has distribution (4.4)*. Additionally, we have*

$$\mathbf{P}(D_\infty>x)\sim\Gamma(1-\alpha)x^{-1} \textit{ as } x\to\infty. \tag{4.8}$$

(iii) **Asymptotic joint degree behaviour.** *Let $D_\infty(i)$ and $D_\infty(j)$ be the asymptotic degrees of two arbitrary distinct vertices $i, j \in \mathbb{N}$. Then*

$$\mathbf{E}\left[t^{D_\infty(i)} s^{D_\infty(j)}\right] \neq \mathbf{E}\left[t^{D_\infty(i)}\right] \mathbf{E}\left[s^{D_\infty(j)}\right], \tag{4.9}$$

for fixed $t, s \in (0,1)$. For t, s sufficiently close to 1 we have,

$$\begin{aligned}
&\left|\mathbf{E}\left[t^{D_\infty(i)} s^{D_\infty(j)}\right] - \mathbf{E}\left[t^{D_\infty(i)}\right] \mathbf{E}\left[s^{D_\infty(j)}\right]\right| \\
&\leq \mathrm{O}\left((1-s)(1-t)\log\left(\left(1+\frac{1}{\Gamma(1-\alpha)(1-s)}\right)\left(1+\frac{1}{\Gamma(1-\alpha)(1-t)}\right)\right)\right) \\
&+ C((1-t)+(1-s)),
\end{aligned} \tag{4.10}$$

for some constant $C \in (0, \infty)$.

Theorem 1 is proved in Section 4.4. The first part of the result deals with the expectation of the typical degree of the graph for small values of parameter ϵ. The particular choice $\epsilon = n^{-1/\alpha}$ results in a logarithmic divergence of $D_n(i)$ with n, corresponding to a semi-sparse regime where the density of links behave as $\sim \frac{log(n)}{n}$. In this regime, asymptotically the degrees exhibit a cumulative power-law exponent -1, for any $\alpha \in (0,1)$. The third part of the result deserves further comments. Indeed Eq. (4.9) shows that $D_\infty(i)$ and $D_\infty(j)$ are *not independent*, and this is a breaking point with respect to previous research. For example, if one considers a Generalized Random Graph, with weights defined as in (4.4) and

$$\widetilde{p}_{ij} = \frac{W_i W_j}{n^{1/\alpha} + W_i W_j}$$

then it follows from Theorem 6.14 of [5] that the asymptotic degree distribution has the same behaviour as our case while the degrees are in fact asymptotically independent. In our case, although there is no independence as (4.9) shows, we still believe that the following will be true

$$\begin{aligned}
\Big|\mathbf{P}(D_\infty(i) > x, D_\infty(j) > x) - \mathbf{P}(D_\infty(i) > x)\mathbf{P}(D_\infty(j) > x)\Big| \\
= \mathrm{o}\left(\mathbf{P}(D_\infty(i) > x)\mathbf{P}(D_\infty(j) > x)\right),
\end{aligned} \tag{4.11}$$

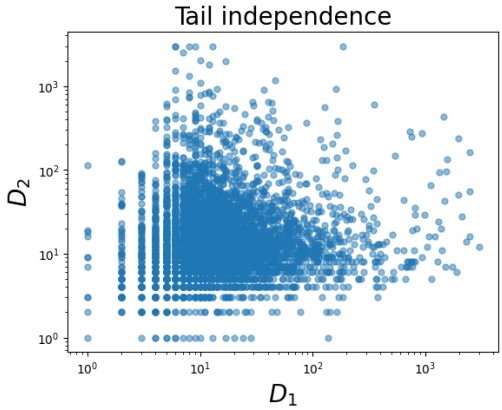

Figure 22: Asymptotic tail independence between degrees. Scatter-plot of the degrees of the vertices with labels 1 and 2 (assigned randomly but fixed for every realisation in the ensemble). Each point in the plot corresponds to one of 5000 realizations of a network of $N = 3000$ vertices, each generated as described at the beginning of Section 4.3 (see Eqs. (4.4) and (4.5)).

and hence that the limiting vector will be *asymptotically tail-independent*. Although not provided with a rigorous proof yet, this conjecture is supported by numerical simulations (see Fig.22). Such a property of limiting degree was observed and proved for Preferential attachment models by using a multivariate version of Karamata's Tauberian theorem [144]. In our case, (4.11) would be valid, given an explicit characterization of the complete joint distribution of the asymptotic degrees. Currently, we have not been able to verify the conditions outlined in [144] for the application of their general multivariate Tauberian theorem. We hope to address this question in the future.

4.3.2 Wedges, triangles and clustering

Our second result concerns the number of wedges and triangles associated to a typical vertex $i \in [n]$, defined respectively as follows:

$$\mathbb{W}_n(i) := \frac{1}{2} \sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{ik}, \qquad \Delta_n(i) := \frac{1}{6} \sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{ik} a_{jk}. \quad (4.12)$$

Theorem 2. [Triangles and Wedges of typical vertices.] *Consider the graph* $\mathbf{G}_n(\alpha, \varepsilon)$ *and let* $\mathbb{W}_n(i)$ *and* $\Delta_n(i)$ *be the number of wedges and triangles at vertex* $i \in [n]$*. Then:*

(i) ***[Average number of wedges.]***

$$\mathbf{E}\left[\mathbb{W}_n(i)\right] \sim \frac{\Gamma^2(-\frac{\alpha}{2})\alpha^2}{2}\varepsilon^{\alpha} n^2 \text{ as } \varepsilon \downarrow 0.$$

In particular, when $\varepsilon_n = n^{-1/\alpha}$*, then*

$$\mathbf{E}\left[\mathbb{W}_n(i)\right] \sim \frac{\Gamma^2(-\frac{\alpha}{2})\alpha^2}{2} n.$$

(ii) ***[Asymptotic distribution of wedges.]*** *Let* $\varepsilon_n = n^{-1/\alpha}$*, then*

$$\mathbb{W}_n(i) \xrightarrow{d} \mathbb{W}_\infty(i)$$

where $\mathbb{W}_\infty(i) = D_\infty(i)(D_\infty(i) - 1)$ *with* $D_\infty(i)$ *as in Theorem 1. Also, we have*

$$\mathbf{P}(\mathbb{W}_\infty(i) > x) \sim \Gamma(1-\alpha)x^{-1/2} \text{ as } x \to \infty.$$

(iii) ***[Average number of triangles.]*** *Let* $i \in [n]$*, the average number of triangles grows as follows:*

$$\mathbf{E}\left[\Delta_n(i)\right] \sim -\frac{\alpha^3}{12}\Gamma^3\left(-\frac{\alpha}{2}\right)\varepsilon^{\frac{3}{2}\alpha} n^2 \text{ as } \varepsilon \downarrow 0.$$

In particular, when $\varepsilon_n = n^{-1/\alpha}$ *we have*

$$\mathbf{E}\left[\Delta_n(i)\right] \sim -\frac{\alpha^3}{12}\Gamma^3\left(-\frac{\alpha}{2}\right)\sqrt{n} \text{ as } n \to \infty.$$

(iv) ***[Convergence in probability for the total number of triangles.]*** *Let* $\varepsilon_n = n^{-1/\alpha}$ *and* $\Delta_n = \sum_{i\in[n]} \Delta_n(i)$ *be the total number of triangles, then*

$$\frac{12\Delta_n}{\alpha^3 n^{3/2}} \xrightarrow{\mathbf{P}} -\Gamma^3\left(-\frac{\alpha}{2}\right) \qquad \text{and} \qquad \frac{12\Delta_n(i)}{\alpha^3 n^{1/2}} \xrightarrow{\mathbf{P}} -\Gamma^3\left(-\frac{\alpha}{2}\right).$$

Remark 3. [Global and local clustering.] *Let $\mathbb{W}_n = \sum_{i\in[n]} \mathbb{W}_n(i)$ be the total number of wedges. We see from the above result that*

$$\frac{\mathbf{E}[\Delta_n]}{\mathbf{E}[\mathbb{W}_n]} \asymp \varepsilon^{\alpha/2}, \text{ as } \varepsilon \to 0.$$

This shows in a quantitative form that the graph is not highly clustered from the point of view of the global *count of triangles. In particular, in the scale of $\varepsilon_n = n^{-1/\alpha}$, the above ratio goes to zero like $n^{-1/2}$. However, this does not mean that the graph is not highly clustered from the point of view of the* local *count of triangles around individual vertices. In particular, simulations from Chap. 2 of this thesis concerning the average local clustering coefficient suggest that the graph is in fact locally clustered (see Fig. 8). A dissimilarity in the behaviour of local and global clustering coefficients has also been observed in different inhomogeneous random graph models, see for example [97, 145, 146].*

4.4 Proof of Theorem 1: typical degrees

Since the Karamata's Tauberian theorem is used here as a key tool in the analysis of the degree distribution and later analysis, it is first worth recalling those results.

Theorem 4 (Karamata's Tauberian theorem [147, Theorem 8.1.6]). *Let X be a non-negative random variable with distribution F and Laplace transform*

$$\widehat{F}(s) = \mathbf{E}\left[e^{-sX}\right], s \geq 0.$$

Let L be a slowly varying function and $\alpha \in (0, 1)$, then the following are equivalent

$$\begin{aligned} &\text{(a) } 1 - \widehat{F}(s) \sim \Gamma(1-\alpha) s^{\alpha} L\left(\frac{1}{s}\right), \text{ as } s \downarrow 0, \\ &\text{(b) } 1 - F(x) \sim x^{-\alpha} L(x), \text{ as } x \to \infty. \end{aligned} \tag{4.13}$$

Then, another property of the tails of products of regularly varying distributions will be needed. A general statement about the product of n iid random variables with Pareto tails can be found in [148, Lemma 4.1 (4)]. For completeness, a proof for two random variables is given here, which is useful in our analysis.

Lemma 5. *Let W_1 and W_2 be independent random variables satisfying the tail assumptions* (4.4)*. Then*

$$\mathbf{P}(W_1 W_2 \geq x) \sim \alpha x^{-\alpha} \log x, \text{ as } x \to \infty. \tag{4.14}$$

Proof. Consider the random variable $\log(W_1)$ which follows an exponential distribution, or alternatively a Gamma distribution with shape parameter $k = 1$ and scale $\theta = 1/\alpha$. Then, the random variable $Z = \log(W_1) + \log(W_2)$ follows a Gamma distribution with shape parameter 2 and scale θ. This means:

$$\mathbf{P}(\log(W_1) + \log(W_2) > x) = \frac{\alpha^2}{\Gamma(2)} \int_x^\infty y e^{-\alpha y} dy.$$

Therefore

$$\begin{aligned} \mathbf{P}(W_1 W_2 > x) &= \mathbf{P}(\log W_1 + \log W_2 > \log x) \\ &= \alpha^2 \int_{\log x}^\infty y e^{-\alpha y} dy \\ &= \alpha^2 \int_x^\infty \log(t) t^{-\alpha - 1} dt. \end{aligned} \tag{4.15}$$

Then applying Karamata's Theorem (see [148, Theorem 12])

$$\mathbf{P}(W_1 W_2 > x) \sim \alpha^2 \, \frac{x^{-\alpha} \log x}{\alpha}, \tag{4.16}$$

which proves the statement.

□

Remark 6. *The above theorem remains true if W_1 and W_2 are not exactly Pareto, but asymptotially tail equivalent to a Pareto distribution, that is under the assumption $\mathbf{P}(W_1 > x) \sim c x^{-\alpha}$ as $x \to \infty$. See Lemma 4.1 of [148] for a proof.*

Proof of Theorem 1. (i) We begin by evaluating the asymptotics of the expected degree, which is an easy consequence of Lemma 5 and Theorem 4. Indeed we can write

$$\mathbf{E}\left[D_n(i)\right] = \sum_{j \neq i} \mathbf{E}\left[1 - \exp\left(-\varepsilon W_i W_j\right)\right] = (n-1)\mathbf{E}\left[1 - \exp\left(-\varepsilon W_1 W_2\right)\right], \tag{4.17}$$

where the last equality is due to the iid nature of the weights.
It follows from Lemma 5 that $\mathbf{P}(W_1 W_2 > x) \sim \alpha x^{-\alpha} \log x$. Therefore, using Theorem 4 we have

$$\mathbf{E}\left[1 - \exp\left(-\epsilon W_1 W_2\right)\right] \sim \Gamma(1-\alpha)\alpha\varepsilon^{\alpha} \log \frac{1}{\varepsilon} \quad \text{as } \varepsilon \downarrow 0, \tag{4.18}$$

which together with (4.17) gives the claim.

(ii) By following the line of the proof of Theorem 6.14 of [5], we can prove our statement by showing that the probability generating function of $D_n(i)$ in the limit $n \to \infty$ corresponds to the probability generating function of a mixed Poisson random variable. Let $t \in (0,1)$, the probability generating function of the degree $D_n(i)$ reads:

$$\mathbf{E}[t^{D_n(i)}] = \mathbf{E}\left[t^{\sum_{j\neq i} a_{ij}}\right] = \mathbf{E}\left[\prod_{j\neq i} t^{a_{ij}}\right], \tag{4.19}$$

where a_{ij} are the entries of the adjacency matrix related to the graph $\mathbf{G}_n(\alpha, \varepsilon)$, *i.e.* Bernoulli random variables with parameter p_{ij} as in (4.5). Conditioned on the weights, these variables are independent. Recall that we denoted by $\mathbf{E}_{W_i}[\cdot]$ the conditional expectation given the weight W_i. Then:

$$\begin{aligned}\mathbf{E}_{W_i}[t^{D_n(i)}] &= \mathbf{E}_{W_i}\left[\prod_{j\neq i}\left((1-t)e^{-\varepsilon_n W_j W_i} + t\right)\right] \\ &= \prod_{j\neq i} \mathbf{E}_{W_i}\left[(1-t)e^{-\varepsilon_n W_j W_i} + t\right] = \prod_{j\neq i} \mathbf{E}_{W_i}\left[\varphi_{W_i}(\varepsilon_n W_j)\right],\end{aligned} \tag{4.20}$$

where we have used the independence of the weights and introduced the function

$$\varphi_{W_i}(x) := (1-t)e^{-W_i x} + t.$$

Let us introduce the following notation to simplify our expression:

$$\psi_n(W_i) := \mathbf{E}_{W_i}\left[\varphi_{W_i}(\varepsilon_n W_j)\right] \text{ for some } j \neq i. \tag{4.21}$$

Using exchangeability, tower property of the conditional expectation, the moment generating function of the $D_n(i)$ can be written as

$$\mathbf{E}[t^{D_n(i)}] = \mathbf{E}\left[\prod_{j\neq i} \mathbf{E}_{W_i}\left[\varphi_{W_i}(\varepsilon_n W_j)\right]\right] = \mathbf{E}\left[\psi_n(W_i)^{n-1}\right]. \tag{4.22}$$

Consider now a differentiable function $h : [0,\infty) \to \mathbb{R}$ such that $h(0) = 0$. Integrating by parts one can show that

$$\mathbf{E}[h(W_j)] = \int_0^\infty h'(x)\mathbf{P}(W_j > x)dx. \tag{4.23}$$

By using (4.23) with $h(w) = \varphi_{W_i}(\varepsilon_n w) - 1$, we have

$$\begin{aligned}
\psi_n(W_i) &= 1 + \mathbf{E}\left[\varphi_{W_i}(\varepsilon_n w) - 1\right] \\
&= 1 + \int_0^\infty \varepsilon_n \varphi'_{W_i}(\varepsilon_n w)(1 - F_W(w))dw \\
&= 1 + \int_0^\infty \varphi'_{W_i}(y)(1 - F_W(\varepsilon_n^{-1} y))dy \\
&= 1 + \int_0^{\varepsilon_n} \varphi'_{W_i}(y)dy + \int_{\varepsilon_n}^\infty \varphi'_{W_i}(y)(1 - F_W(\varepsilon_n^{-1} y))dy \\
&= 1 + \varphi_{W_i}(\varepsilon_n) - \varphi_{W_i}(0) + \varepsilon_n^\alpha \int_{\varepsilon_n}^\infty (t-1) W_i e^{-yW_i} y^{-\alpha} dy.
\end{aligned} \tag{4.24}$$

In particular for $\varepsilon_n = n^{-1/\alpha}$, combining (4.22) and (4.24) gives

$$\begin{aligned}
\mathbf{E}\left[t^{D_n(i)}\right] &= \mathbf{E}\left[\psi_n(W_i)^{n-1}\right] \\
&= \mathbf{E}\left[\left(1 + \varphi_{W_i}(n^{-1/\alpha}) - \varphi_{W_i}(0) + \right.\right. \\
&\qquad \left.\left. + \frac{1}{n}\int_{n^{-1/\alpha}}^\infty (t-1) W_i e^{-yW_i} y^{-\alpha} dy\right)^{n-1}\right].
\end{aligned}$$

Note that for a fixed realization of W_i, using change of variable $z = W_i y$, we have as $n \to \infty$,

$$(1-t)\int_{n^{-1/\alpha}}^\infty W_i e^{-yW_i} y^{-\alpha} dy \to (1-t) W_i^\alpha \Gamma(1-\alpha)$$

and

$$\varphi_{W_i}(n^{-1/\alpha}) \to \varphi_{W_i}(0) = 1.$$

Observe that $\varphi_{W_i}(n^{-1/\alpha}) - \varphi(0) = -(1-t)(1-e^{-W_i n^{1/\alpha}}) \leq 0$ and

$$0 \leq (1-t)\int_{n^{-1/\alpha}}^{\infty} W_i e^{-yW_i} y^{-\alpha} dy \leq (1-t)W_i^{\alpha}\Gamma(1-\alpha).$$

Hence using $(1-x/n)^n \leq e^{-x}$ we have

$$\begin{aligned}\left(1 + \varphi_{W_i}(n^{-1/\alpha}) - \varphi_{W_i}(0) - (1-t)\frac{1}{n}\int_{n^{-1/\alpha}}^{\infty} W_i e^{-yW_i} y^{-\alpha} dy\right)^{n-1} \\ \leq \exp\left(-(1-t)W_i^{\alpha}\Gamma(1-\alpha)\right) \\ \leq 1.\end{aligned}$$

Thus we can apply the Dominated Convergence Theorem to claim that

$$\lim_{n\to\infty} \mathbf{E}\left[t^{D_n(i)}\right] = \mathbf{E}\left[\exp\left(-(1-t)\, W_i^{\alpha}\Gamma(1-\alpha)\right)\right].$$

So the generating function of the graph degree $D_n(i)$ asymptotically corresponds to the generating function of a mixed Poisson random variable with parameter $\Gamma(1-\alpha)W_i^{\alpha}$. Therefore, the variable $D_n(i) \xrightarrow{d} D_{\infty}(i)$ where $D_{\infty}(i) \mid W_i \stackrel{d}{=} \text{Poisson}\left(\Gamma(1-\alpha)W_i^{\alpha}\right)$.

In particular, we have the following tail of the distribution of the random variable $D_{\infty}(i)$.

$$\begin{aligned}\mathbf{P}(D_{\infty}(i) \geq k) &= \int_0^{\infty} \mathbf{P}(\text{Poisson}\left(\Gamma(1-\alpha)w^{\alpha}\right) \geq k | W_i = w)\; F_{W_i}(dw) \\ &= \int_0^{\infty} \sum_{m\geq k} \frac{e^{-\Gamma(1-\alpha)w^{\alpha}}\Gamma(1-\alpha)^m w^{\alpha m}}{m!}\; F_{W_i}(dw) \\ &= \sum_{m\geq k} \frac{1}{m!}\int_1^{\infty} e^{-\Gamma(1-\alpha)w^{\alpha}}\Gamma(1-\alpha)^m w^{\alpha m}\alpha w^{-\alpha-1} dw.\end{aligned} \tag{4.25}$$

Let us introduce the new variable $y = \Gamma(1-\alpha)w^\alpha$, then

$$\int_1^\infty e^{-\Gamma(1-\alpha)w^\alpha}\Gamma(1-\alpha)^m w^{\alpha m}\alpha w^{-\alpha-1}dw \tag{4.26}$$

$$= \int_1^\infty e^{-\Gamma(1-\alpha)w^\alpha}\Gamma(1-\alpha)^{m-1} w^{\alpha m} w^{-2\alpha}\alpha\Gamma(1-\alpha)w^{\alpha-1}dw$$

$$= \Gamma(1-\alpha)\int_{\Gamma(1-\alpha)}^\infty e^{-y}y^{m-2}dy$$

$$= \Gamma(1-\alpha)\Gamma(m-1) - \Gamma(1-\alpha)\int_0^{\Gamma(1-\alpha)} e^{-y}y^{m-2}dy. \tag{4.27}$$

The first integral is dominant with respect to the second one. To show this, we can use a trivial bound:

$$\int_0^{\Gamma(1-\alpha)} e^{-y}y^{m-2}dy \le \Gamma(1-\alpha)^m.$$

Since $m! \ge (m/e)^m$, then the following inequalities hold true

$$\sum_{m\ge k}\frac{\Gamma(1-\alpha)^m}{m!} \le \sum_{m\ge k}\frac{(e\Gamma(1-\alpha))^m}{m^m} \le C(e\Gamma(1-\alpha))^k k^{-k},$$

for some positive constant C. Note that in last step we used k is large enough (it is at least greater that $e\Gamma(1-\alpha)$). Note that

$$k\sum_{m\ge k}\frac{\Gamma(1-\alpha)^m}{m!} \le Ce^{\log k - k\log k + ke\Gamma(1-\alpha)} \to 0, \text{ as } k\to\infty.$$

By using (4.27) we therefore obtain

$$\sum_{m\ge k}\frac{1}{m!}\int_1^\infty e^{-\Gamma(1-\alpha)w^\alpha}\Gamma(1-\alpha)^m w^{\alpha m}\alpha w^{-\alpha-1}dw$$

$$= \Gamma(1-\alpha)\sum_{m\ge k}\frac{\Gamma(m-1)}{m!} + o\left(k^{-1}\right)$$

$$= \Gamma(1-\alpha)\sum_{m\ge k}\frac{(m-2)!}{m!} + o\left(k^{-1}\right)$$

$$= \Gamma(1-\alpha)\sum_{m\ge k}\frac{1}{m(m-1)} + o\left(k^{-1}\right) \sim \frac{\Gamma(1-\alpha)}{k}.$$

This shows that $\mathbf{P}(D_\infty(i) \geq k) \sim \Gamma(1-\alpha)k^{-1}$ as $k \to \infty$.

(iii) Fix $t, s \in (0,1)$. Due to the exchangeability of vertices, without loss of generality, we consider the vertices 1 and 2.

$$\mathbf{E}\left[t^{D_n(1)}s^{D_n(2)}\right] = \mathbf{E}\left[t^{\sum_{j\neq 1} a_{1j}}s^{\sum_{j\neq 2} a_{2j}}\right] = \mathbf{E}\left[\prod_{j\neq 1,2} t^{a_{1j}}s^{a_{2j}}\,(ts)^{a_{12}}\right]$$
$$= \mathbf{E}\left[\left((1-ts)\,e^{-\varepsilon W_1 W_2}+ts\right)\prod_{j\neq 1,2}\left((1-t)\,e^{-\varepsilon W_1 W_j}+t\right)\left((1-s)\,e^{-\varepsilon W_2 W_j}+s\right)\right] \tag{4.28}$$

where we have used the independence of the connection probabilities *given* the weights. In order to simplify the notation, we can introduce the following functions:

$$\begin{aligned}&\phi_a^b(x) := (1-b)\,e^{-\varepsilon_n a x}+b,\\ &\psi_n(W_1,W_2) := \mathbf{E}_{W_1,W_2}\left[\phi^t_{W_1}(W_j)\phi^s_{W_2}(W_j)\right], \quad \text{for some } j\neq 1,2,\end{aligned} \tag{4.29}$$

where $a, b > 0$ and, as customary throughout this chapter, $\mathbf{E}_{W_1,W_2}[\,\cdot\,] := \mathbf{E}[\,\cdot\,|W_1,W_2]$.

Using the tower property of conditional expectation, Eq. (4.28) reads

$$\begin{aligned}\mathbf{E}\left[t^{D_n(1)}s^{D_n(2)}\right] &= \mathbf{E}\left[\phi^{ts}_{W_1}(W_2)\prod_{j\neq 1,2}\phi^t_{W_1}(W_j)\phi^s_{W_2}(W_j)\right]\\ &= \mathbf{E}\left[\phi^{ts}_{w_1}(w_2)\,\mathbf{E}_{W_1,W_2}\left[\prod_{j\neq 1,2}\phi^t_{w_1}(w_j)\phi^s_{w_2}(w_j)\right]\right]\\ &= \mathbf{E}\left[\phi^{ts}_{w_1}(W_2)\prod_{j\neq 1,2}\mathbf{E}_{W_1,W_2}\left[\phi^t_{W_1}(W_j)\phi^s_{W_2}(W_j)\right]\right]\\ &= \mathbf{E}\left[\phi^{ts}_{W_1}(W_2)\psi_n(W_1,W_2)^{n-2}\right],\end{aligned} \tag{4.30}$$

where we used conditional independence in the second last step and exchangeability in the last step. The function ψ_n can be processed as

follows. Just as in the one dimensional case, using $\varepsilon_n = n^{-1/\alpha}$, we get $\mathbf{P}$ a.s.,

$$\begin{aligned}
&\psi_n(W_1, W_2) - 1 = \mathbf{E}_{W_1,W_2}\left[\phi^t_{W_1}(W_3)\phi^s_{W_2}(W_3) - 1\right] \\
&\to -\Gamma(1-\alpha)\left[(1-t)(1-s)(W_1+W_2)^\alpha + (1-t)sW_1^\alpha + t(1-s)W_2^\alpha\right] \\
&= -\Gamma(1-\alpha)\Big\{(1-t)(1-s)\left[(W_1+W_2)^\alpha - W_1^\alpha - W_2^\alpha\right] + (1-t)W_1^\alpha + \\
&\qquad (1-s)W_2^\alpha\Big\}
\end{aligned} \tag{4.31}$$

where in the second step we used (4.23) with $h(x) := \phi^t_{W_1}(x)\phi^s_{W_2}(x) - 1$. Observe that $\phi^{ts}_{w_1}(W_2) \leq 1$ and $1 - \psi_n(W_1, W_2) \geq 0$ and hence we can use Dominated convergence theorem as in the single vertex case. Therefore using $\phi^{ts}_{W_1}(W_2) \to 1$ and (4.31) we get

$$\begin{aligned}
&\mathbf{E}\left[t^{D_\infty(1)}s^{D_\infty(2)}\right] = \lim_{n\to\infty}\mathbf{E}\left[t^{D_n(1)}s^{D_n(2)}\right] \\
&= \mathbf{E}\left[e^{-\Gamma(1-\alpha)\Big\{(1-t)(1-s)[(W_1+W_2)^\alpha - W_1^\alpha - W_2^\alpha] + (1-t)W_1^\alpha + (1-s)W_2^\alpha\Big\}}\right] \\
&= \mathbf{E}\left[e^{-\Gamma(1-\alpha)(1-t)(1-s)[(W_1+W_2)^\alpha - W_1^\alpha - W_2^\alpha]}e^{-\Gamma(1-\alpha)(1-t)W_1^\alpha}e^{-\Gamma(1-\alpha)(1-s)W_2^\alpha}\right].
\end{aligned} \tag{4.32}$$

It is straightforward to note that in the limit $t \to 1$ and for fixed $s \in (0,1)$ we recover the correct moment generating function of $D_\infty(1)$ and the inverse holds true as well. Finally, since $(W_1+W_2)^\alpha \neq W_1^\alpha + W_2^\alpha$ $\mathbf{P}$-a.s., then (4.9) follows.

We next move to the proof of (4.10), for which we abbreviate $\eta = 1-t$, $\gamma = 1-s$ and show that

$$\lim_{\substack{\eta\to 0\\ \gamma\to 0}}\left|\mathbf{E}\left[(1-\eta)^{D_\infty(1)}(1-\gamma)^{D_\infty(2)}\right] - \mathbf{E}\left[(1-\eta)^{D_\infty(1)}\right]\mathbf{E}\left[(1-\gamma)^{D_\infty(2)}\right]\right| = 0.$$

$$
\begin{aligned}
&\left| \mathbf{E}\left[(1-\eta)^{D_\infty(1)}(1-\gamma)^{D_\infty(2)}\right] - \mathbf{E}\left[(1-\eta)^{D_\infty(1)}\right]\mathbf{E}\left[(1-\gamma)^{D_\infty(1)}\right] \right| \\
&= \Big| \mathbf{E}\left[e^{-\Gamma(1-\alpha)\eta\gamma[(w_1+w_2)^\alpha - w_1^\alpha - w_2^\alpha]} e^{-\Gamma(1-\alpha)\eta w_1^\alpha} e^{-\Gamma(1-\alpha)\gamma w_2^\alpha}\right] + \\
&\qquad\qquad - \mathbf{E}\left[e^{-\Gamma(1-\alpha)\eta w_1^\alpha}\right]\mathbf{E}\left[e^{-\Gamma(1-\alpha)\gamma w_2^\alpha}\right] \Big| \\
&= \left| \mathbf{E}\left[\left(e^{-\Gamma(1-\alpha)\eta\gamma[(w_1+w_2)^\alpha - w_1^\alpha - w_2^\alpha]} - 1\right) e^{-\Gamma(1-\alpha)\eta w_1^\alpha} e^{-\Gamma(1-\alpha)\gamma w_2^\alpha}\right] \right| \\
&= \left| \int_1^\infty \left(\sum_{k\geq 1} \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} \left[-(x+y)^\alpha + x^\alpha + y^\alpha\right]^k \right) \right. \\
&\qquad\qquad \left. e^{-\Gamma(1-\alpha)\eta x^\alpha} e^{-\Gamma(1-\alpha)\gamma y^\alpha} \alpha^2 (xy)^{-\alpha-1} dxdy \right| \\
&\leq \int_1^\infty \sum_{k=1}^\infty \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} \left| -(x+y)^\alpha + x^\alpha + y^\alpha \right|^k \\
&\qquad\qquad e^{-\Gamma(1-\alpha)\eta x^\alpha} e^{-\Gamma(1-\alpha)\gamma y^\alpha} \alpha^2 (xy)^{-\alpha-1} dxdy.
\end{aligned}
\tag{4.33}
$$

Now, since $(x+y)^\alpha \leq (x^\alpha + y^\alpha) \quad \forall\ \alpha \in (0,1)$, we get

$$
\left| x^\alpha + y^\alpha - (x+y)^\alpha \right|^k \leq \left(x^\alpha + y^\alpha\right)^k \leq 2^{k-1}\left(x^{\alpha k} + y^{\alpha k}\right). \tag{4.34}
$$

Therefore, using Fubini we can bring the summation out of the integral and using the inequality (4.34), we are left with the following quantity (which we will show to be converging to zero):

$$
\alpha^2 \sum_{k=1}^\infty \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} 2^{k-1} \int_1^\infty \left(x^{\alpha k} + y^{\alpha k}\right) e^{-\Gamma(1-\alpha)(\eta x^\alpha + \gamma y^\alpha)} \frac{dxdy}{(xy)^{\alpha+1}}. \tag{4.35}
$$

Let us consider the different terms of the sum separately. In the following we will consider the exponential integral $E_1(x) = \int_1^\infty \frac{e^{-tx}}{t} dt$ and the related inequality $E_1(x) < e^{-x}\ln(1+\frac{1}{x})$ for any $x > 0$.

Case 1: $k = 1$

$$
\begin{aligned}
&\alpha^2\Gamma(1-\alpha)\eta\gamma \int_1^\infty \int_1^\infty \left(x^\alpha + y^\alpha\right) e^{-\Gamma(1-\alpha)(\eta x^\alpha + \gamma y^\alpha)} \frac{dxdy}{(xy)^{\alpha+1}} \\
&= \Gamma(1-\alpha)\eta\gamma \left[E_1\left(\Gamma(1-\alpha)\eta\right) \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\gamma z}}{z^2} dz \right. \\
&\qquad \left. + E_1\left(\Gamma(1-\alpha)\gamma\right) \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\eta z}}{z^2} dz \right] \\
&\leq \Gamma(1-\alpha)\eta\gamma \left[E_1\left(\Gamma(1-\alpha)\eta\right) + E_1\left(\Gamma(1-\alpha)\gamma\right) \right] \\
&< \Gamma(1-\alpha) \left[\eta\gamma e^{-\Gamma(1-\alpha)\,\eta} \log\left(1 + \frac{1}{\Gamma(1-\alpha)\,\eta}\right) \right. \\
&\qquad \left. + \eta\gamma\, e^{-\Gamma(1-\alpha)\,\gamma} \log\left(1 + \frac{1}{\Gamma(1-\alpha)\,\gamma}\right) \right]
\end{aligned}
$$

Case 2: $k = 2$

$$
\begin{aligned}
&\left(\alpha\,\Gamma(1-\alpha)\eta\gamma\right)^2 \int_1^\infty \int_1^\infty \left(x^{2\alpha} + y^{2\alpha}\right) e^{-\,\Gamma(1-\alpha)\eta x^\alpha} e^{-\,\Gamma(1-\alpha)\gamma y^\alpha} (xy)^{-\alpha-1} dxdy \\
=&\Gamma(1-\alpha)\left(\eta\gamma\right)^2 \left[\frac{e^{-\Gamma(1-\alpha)\eta}}{\eta} \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\gamma z}}{z^2} dz + \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\eta z}}{z^2} dz\; \frac{e^{-\Gamma(1-\alpha)\gamma}}{\gamma} \right] \\
\leq&\Gamma(1-\alpha)\left(\eta\gamma\right)^2 \left[\frac{e^{-\Gamma(1-\alpha)\eta}}{\eta} + \frac{e^{-\Gamma(1-\alpha)\gamma}}{\gamma} \right]
\end{aligned}
$$

Case 3: $k > 2$

$$
\begin{aligned}
&\alpha^2 \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} 2^{k-1} \int_1^\infty \left(x^{\alpha k} + y^{\alpha k}\right) e^{-\Gamma(1-\alpha)\eta x^\alpha} e^{-\Gamma(1-\alpha)\gamma y^\alpha} (xy)^{-\alpha-1} dxdy \\
=& \alpha^2 \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} 2^{k-1} \left[\int_1^\infty \frac{e^{-\Gamma(1-\alpha)\eta x^\alpha}}{x^{\alpha(1-k)+1}} dx \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\gamma y^\alpha}}{y^{\alpha+1}} dy \right. \\
&\qquad \left. + \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\eta x^\alpha}}{x^{\alpha+1}} dx \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\gamma y^\alpha}}{y^{\alpha(1-k)+1}} dy \right] \\
=& \alpha^2 \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} 2^{k-1} \left[\int_{\Gamma(1-\alpha)\eta}^\infty \left(\frac{z}{\Gamma(1-\alpha)\eta}\right)^{k-2} \frac{e^{-z} dz}{\alpha\Gamma(1-\alpha)\eta} \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\gamma z}}{z^2} \frac{dz}{\alpha} \right. \\
&\qquad \left. + \int_1^\infty \frac{e^{-\Gamma(1-\alpha)\eta z}}{z^2} dz \frac{1}{\alpha} \int_{\Gamma(1-\alpha)\gamma}^\infty \left(\frac{z}{\Gamma(1-\alpha)\gamma}\right)^{k-2} \frac{e^{-z} dz}{\alpha\Gamma(1-\alpha)\gamma} \right] \\
\leq& \alpha^2 \frac{(\Gamma(1-\alpha)\eta\gamma)^k}{k!} 2^{k-1} \left[\left(\frac{1}{\Gamma(1-\alpha)\eta}\right)^{k-1} \frac{1}{\alpha^2} \int_{\Gamma(1-\alpha)\eta}^\infty z^{k-2} e^{-z} dz \right. \\
&\qquad \left. + \left(\frac{1}{\Gamma(1-\alpha)\gamma}\right)^{k-1} \frac{1}{\alpha^2} \int_{\Gamma(1-\alpha)\gamma}^\infty z^{k-2} e^{-z} dz \right] \\
\leq& \frac{\Gamma(1-\alpha)}{k!} 2^{k-1} \Gamma(k-1) (\eta\gamma)^k \left[\left(\frac{1}{\eta}\right)^{k-1} + \left(\frac{1}{\gamma}\right)^{k-1} \right] \\
=& \Gamma(1-\alpha) \frac{2^{k-1}}{k(k-1)} \left(\eta\gamma^k + \eta^k\gamma\right)
\end{aligned}
$$

So, combining together all the bounds, we have

$$
\begin{aligned}
&\left| \mathbf{E}\left[(1-\eta)^{D_\infty(1)}(1-\gamma)^{D_\infty(2)}\right] - \mathbf{E}\left[(1-\eta)^{D_\infty(1)}\right] \mathbf{E}\left[(1-\gamma)^{D_\infty(2)}\right] \right| \\
&< \Gamma(1-\alpha) \left[\eta\gamma \log\left(1 + \frac{1}{\Gamma(1-\alpha)\,\eta}\right) + \eta\gamma \log\left(1 + \frac{1}{\Gamma(1-\alpha)\,\gamma}\right) \right. \\
&\qquad \left. + \frac{1}{2} \sum_{k=2}^\infty \frac{2^k}{k(k-1)} (\eta\gamma^k + \eta^k\gamma) \right].
\end{aligned} \tag{4.36}
$$

Since $x \log(1 + \frac{1}{x}) \to 0$ as $x \to 0$, the above quantity goes to 0 as $\eta, \gamma \to 0$. This completes the proof of Theorem 1. □

4.5 Proof of Theorem 2: wedges & triangles

Proof of Theorem 2. **(i)** Start from the equality

$$\begin{aligned} 2\mathbf{E}\left[\mathbb{W}_n(i)\right] &= \mathbf{E}\left[\sum_{j\neq i}\sum_{k\neq i,j} a_{ij}a_{ik}\right] \\ &= (n-1)(n-2)\alpha^3\int_1^\infty\int_1^\infty\int_1^\infty \frac{(1-e^{-\varepsilon xy})(1-e^{-\varepsilon xz})}{(xyz)^{\alpha+1}}dxdydz. \end{aligned} \tag{4.37}$$

The latter integral can be solved exactly by using the substitutions $A = xy$, $B = xz$ and $C = yz$. Thus:

$$\begin{aligned} \mathbf{E}\left[\mathbb{W}_n(i)\right] &= \frac{(n-1)(n-2)}{4}\alpha^3\int_1^\infty \frac{1-e^{-\varepsilon A}}{A^{\alpha/2+1}}dA\int_1^\infty \frac{1-e^{-\varepsilon B}}{B^{\alpha/2+1}}dB\int_1^\infty C^{-\alpha/2-1}dC \\ &= \frac{(n-1)(n-2)}{4}\alpha^3\left[\frac{2}{\alpha}-\varepsilon^{\alpha/2}\Gamma\left(-\frac{\alpha}{2};\varepsilon\right)\right]^2\frac{2}{\alpha} \end{aligned} \tag{4.38}$$

where $\Gamma\left(-\frac{\alpha}{2};\varepsilon\right)$ is the incomplete Gamma function. When ε is small, the following expansion ([149]) can be used:

$$\Gamma(s;\varepsilon) \sim \Gamma(s) - \sum_{k=0}^{\infty}(-1)^k\frac{\varepsilon^{s+k}}{k!(s+k)}, \quad \text{as } \varepsilon\to 0^+ \text{ and } s\neq 0,-1,-2,-3,\ldots \tag{4.39}$$

So we see that in our case,

$$\Gamma\left(-\frac{\alpha}{2},\varepsilon\right) \sim \Gamma\left(-\frac{\alpha}{2}\right) + \frac{2}{\alpha}\varepsilon^{-\alpha/2} + O\left(\varepsilon^{1-\alpha/2}\right), \text{ as } \varepsilon\to 0.$$

Therefore, at the first order:

$$\mathbf{E}\left[\mathbb{W}_n(i)\right] \overset{\varepsilon\to 0}{\sim} \frac{\alpha^2n^2}{2}\left[-\epsilon^{\alpha/2}\Gamma\left(-\frac{\alpha}{2}\right)\right]^2 = \frac{\alpha^2n^2}{2}\varepsilon^{\alpha}\Gamma\left(-\frac{\alpha}{2}\right)^2. \tag{4.40}$$

(ii) Assume now that $\varepsilon_n = n^{-1/\alpha}$. From Theorem 1 we know that $D_n(i) \xrightarrow{d} D_\infty(i)$. Using the continuous mapping $x\mapsto x(x-1)$ we have,

by the Continuous Mapping Theorem, convergence in distribution of the number of wedges $\mathbb{W}_n(i)$:

$$\mathbb{W}_n(i) = D_n(i)(D_n(i)-1) \xrightarrow{d} D_\infty(i)(D_\infty(i)-1) \equiv \mathbb{W}_\infty(i). \tag{4.41}$$

Let us now show that as $x \to \infty$, the tail satisfies

$$\mathbf{P}(\mathbb{W}_\infty(i) > x) \sim \Gamma(1-\alpha)x^{-1/2} \text{ as } x \to \infty.$$

Indeed, first notice that by (4.8) we have

$$\mathbf{P}(D_\infty(i)^2 > x) = \mathbf{P}(D_\infty(i) > \sqrt{x}) \sim \Gamma(1-\alpha)x^{-1/2}. \tag{4.42}$$

We first find the upper bound. Let $\delta > 0$,

$$\begin{aligned}
\mathbf{P}(D_\infty(i)^2 - D_\infty(i) > x) &= \mathbf{P}(D_\infty(i)^2 > x + D_\infty(i), D_\infty(i) > x+\delta) \\
&\quad + \mathbf{P}(D_\infty(i)^2 > x + D_\infty(i), D_\infty(i) \le x+\delta) \\
&\le \mathbf{P}(D_\infty(i)^2 > x + D_\infty(i), D_\infty(i) > x+\delta) + \mathbf{P}(D_\infty(i)^2 > \omega) \\
&\le \mathbf{P}(D_\infty(i) > x+\delta) + \mathbf{P}(D_\infty(i)^2 > x).
\end{aligned}$$

Then for any $\delta > 0$

$$\begin{aligned}
\frac{\mathbf{P}(\mathbb{W}_\infty(i) > x)}{\mathbf{P}(D_\infty(i)^2 > x)} &= \frac{\mathbf{P}(D_\infty(i)^2 - D_\infty(i) > x)}{P(D_\infty(i)^2 > x)} \le \frac{\mathbf{P}(D_\infty(i) > x+\delta)}{\mathbf{P}(D_\infty(i)^2 > x)} + 1 \\
&= \frac{\Gamma(1-\alpha)(x+\delta)^{-1}}{\Gamma(1-\alpha)x^{-1/2}} + 1 \sim x^{-1/2} + 1 \to 1.
\end{aligned}$$

This shows that

$$\limsup_{x\to\infty} \frac{\mathbf{P}(\mathbb{W}_\infty(i) > x)}{\Gamma(1-\alpha)x^{-1/2}} \le 1.$$

We do a similar break-up for the lower bound. Let $\delta > 0$,

$$\begin{aligned}
\mathbf{P}(D_\infty(i)^2 - D_\infty(i) > x) &\ge \mathbf{P}(D_\infty(i)^2 > (1+\delta)x, D_\infty(i) \le \delta x) \\
&\ge \mathbf{P}(D_\infty(i)^2 > (1+\delta)x) - \mathbf{P}(D_\infty(i) > \delta x).
\end{aligned}$$

Then

$$\begin{aligned}\frac{\mathbf{P}(D_\infty(i)^2 - D_\infty(i) > x)}{\mathbf{P}(D_\infty(i)^2 > x)} &\geq \frac{((1+\delta)x)^{-1/2} - (\delta x)^{-1}}{x^{-1/2}} \\ &\sim \sqrt{\frac{1}{(1+\delta)}} - \frac{\sqrt{x}}{\delta x} \quad (\text{as } x \to \infty) \\ &\sim \frac{1}{\sqrt{(1+\delta)}} - \delta^{-1}x^{-1/2} \to_{x\to\infty} \frac{1}{\sqrt{1+\delta}}.\end{aligned} \tag{4.43}$$

So we have

$$\liminf_{x\to\infty} \frac{\mathbf{P}(\mathbb{W}_\infty(i) > x)}{\Gamma(1-\alpha)x^{-1/2}} \geq \frac{1}{\sqrt{(1+\delta)}}.$$

The result follows by taking $\delta \to 0$ and it shows that $\mathbf{P}(\mathbb{W}_\infty(i) > x) \sim \Gamma(1-\alpha)x^{-1/2}$.

(iii) We will here focus on the average number of triangles, whose evaluation will require integral asymptotics similar to the ones used for the wedges.

$$\begin{aligned}6\,\mathbf{E}\left[\Delta_n(i)\right] &= \mathbf{E}\left[\sum_{j\neq i}\sum_{k\neq i,j} a_{ij}a_{ik}a_{jk}\right] = \sum_{j\neq i}\sum_{k\neq i,j}\mathbf{E}\left[p_{ij}p_{ik}p_{jk}\right] \\ &= (n-1)(n-2)\alpha^3\int_1^\infty\int_1^\infty\int_1^\infty \frac{(1-e^{-\varepsilon xy})(1-e^{-\varepsilon xz})(1-e^{-\varepsilon yz})}{(xyz)^{\alpha+1}}\,dx\,dy\,dz.\end{aligned} \tag{4.44}$$

Analogously to what has been done before, the latter integral can be

solved exactly by using the substitutions $A = xy, B = xz, C = yz$. Thus:

$$\begin{aligned}
6\,\mathbf{E}\left[\Delta_n(i)\right] &= \frac{(n-1)(n-2)\alpha^3}{2}\int_1^\infty \frac{1-e^{-\epsilon A}}{A^{\alpha/2+1}}dA\int_1^\infty \frac{1-e^{-\epsilon B}}{B^{\alpha/2+1}}dB\int_1^\infty \frac{1-e^{-\epsilon C}}{C^{\alpha/2+1}}dC \\
&= \frac{(n-1)(n-2)\alpha^3}{2}\left[\frac{2}{\alpha} - \epsilon^{\alpha/2}\Gamma\left(-\frac{\alpha}{2};\epsilon\right)\right]^3 \\
&= \frac{n^2\alpha^3}{2}\left[-\epsilon^{\alpha/2}\Gamma\left(-\frac{\alpha}{2}\right) + O\left(\varepsilon\right)\right]^3 \\
&= -\frac{n^2\alpha^3}{2}\varepsilon^{\frac{3\alpha}{2}}\Gamma\left(-\frac{\alpha}{2}\right)^3 + O\left(n^2\varepsilon^3\right)
\end{aligned} \tag{4.45}$$

where in the last step we used the expansion approximating the incomplete Gamma function (4.39). By our assumption, since $\alpha < 2$ and $\varepsilon \to 0$, we have $\varepsilon^3 = o\left(\varepsilon^{\frac{3\alpha}{2}}\right)$ and hence the result follows.

(iv) Let $\varepsilon_n = n^{-1/\alpha}$ then $\mathbf{E}[\Delta_n(i)] \sim -\frac{\alpha^3}{12}n^{1/2}\Gamma\left(-\frac{\alpha}{2}\right)^3$. The above computations also shows that $\Delta_n = \sum_{i=1}^n \Delta_n(i)$ behaves as

$$\mathbf{E}[\Delta_n] \sim -\frac{\alpha^3 n^{3/2}}{12}\Gamma\left(-\frac{\alpha}{2}\right)^3. \tag{4.46}$$

For studying the concentration of the latter quantity, we start by evaluating the second moment:

$$\mathbf{E}\left[\Delta_n^2\right] = \mathbf{E}\left[\sum_{i,j,k}{}' \sum_{u,v,w}{}' a_{ij}a_{ik}a_{jk}a_{uv}a_{uw}a_{vw}\right] = A + B + C + D \tag{4.47}$$

where A represents the term in which there is no intersection between the triples of indices of the two summations ($(u,v,w) \neq (i,j,k)$), that is, $|\{u,v,w\} \cap \{i,j,k\}| = 0$, B is the term in which there is an intersection of 1 index, that is, $|\{u,v,w\} \cap \{i,j,k\}| = 1$ C an intersection of 2 indices, that is, $|\{u,v,w\} \cap \{i,j,k\}| = 2$ and D is the term in which all the indices coincide $|\{u,v,w\} \cap \{i,j,k\}| = 3$. Above, $\sum_{i,j,k}{}'$ means sum over distinct indices.

(A) No common indices:

$$\begin{aligned} A &= \mathbf{E}\left[\sum_{i,j,k}{}' \sum_{(u,v,w)\neq(i,j,k)}{}' a_{ij}a_{ik}a_{jk}a_{uv}a_{uw}a_{vw}\right] \\ &= \mathbf{E}\left[\sum_{i,j,k}{}' a_{ij}a_{ik}a_{jk}\right]\mathbf{E}\left[\sum_{u,v,w}{}' a_{uv}a_{uw}a_{vw}\right] \\ &= \mathbf{E}\left[\Delta_n\right]^2. \end{aligned} \tag{4.48}$$

(B) One common index:

$$\begin{aligned} B &= \mathbf{E}\left[\sum_{i,j,k}{}' \sum_{1\,intersection}{}' a_{ij}a_{ik}a_{jk}a_{uv}a_{uw}a_{vw}\right] \\ &= \mathbf{E}\left[\sum_{i,j,k} a_{ij}a_{ik}a_{jk}\,3\sum_{v,w} a_{vw}(a_{iv}a_{iw}+a_{jv}a_{jw}+a_{kv}a_{kw})\right] \\ &\leq \mathbf{E}\left[\sum_{i,j,k} a_{ij}a_{ik}a_{jk}\,9\sum_{v,w} a_{vw}\right] = 9\,n\mathbf{E}\left[\Delta_n\right]\,\mathbf{E}\left[D_n(i)\right]. \end{aligned} \tag{4.49}$$

(C) Exactly two common indices:

$$\begin{aligned} C &= \mathbf{E}\left[\sum_{i,j,k}{}' \sum_{2\,intersections}{}' a_{ij}a_{ik}a_{jk}a_{uv}a_{uw}a_{vw}\right] \\ &\leq 6\,n\,\mathbf{E}\left[\sum_{i,j,k}{}' a_{ij}a_{ik}a_{jk}\right] = 6\,n\,\mathbf{E}\left[\Delta_n\right]. \end{aligned} \tag{4.50}$$

(D) All indices match:

$$D = \mathbf{E}\left[\sum_{i,j,k}{}' a_{ij}a_{ik}a_{jk}a_{ij}a_{ik}a_{jk}\right] = \mathbf{E}\left[\sum_{i,j,k}{}' a_{ij}a_{ik}a_{jk}\right] = \mathbf{E}\left[\Delta_n\right]. \tag{4.51}$$

Therefore:

$$\begin{aligned}\frac{\mathrm{Var}\left(\Delta_n\right)}{\mathbf{E}\left[\Delta_n\right]^2} &= \frac{\mathbf{E}\left[\Delta_n^2\right]-\mathbf{E}\left[\Delta_n\right]^2}{\mathbf{E}\left[\Delta_n\right]^2} = \frac{B+C+D}{\mathbf{E}\left[\Delta_n\right]^2} \\ &\leq \frac{9\,n\mathbf{E}\left[\Delta_n\right]\,\mathbf{E}\left[D_n(i)\right]+6\,n\,\mathbf{E}\left[\Delta_n\right]+\mathbf{E}\left[\Delta_n\right]}{\mathbf{E}\left[\Delta_n\right]^2} \\ &\sim \frac{\alpha 9\,c_6\,c_7\,n^5 n^{-5/2}\log 1/n+6\,c_6\,n^4 n^{-3/2}+c_6\,n^3 n^{-3/2}}{c_6^2 n^6 n^{-3}} \\ &= O\left(\frac{\log n}{n^{1/2}}\right),\end{aligned} \tag{4.52}$$

where $c_6 = \frac{\alpha^3}{12}\Gamma\left(-\frac{\alpha}{2}\right)^3$ and $c_7 = \Gamma(1-\alpha)$ are taken respectively from equations (4.46) and (4.7). Now using Chebyshev's inquality it follows that

$$\mathbf{P}\left(\left|\frac{\Delta_n}{\mathbf{E}[\Delta_n]}-1\right|\geq\varepsilon\right)\leq\frac{\mathrm{Var}\left(\Delta_n\right)}{\varepsilon^2\mathbf{E}\left[\Delta_n\right]^2}=O\left(\frac{\log n}{n^{1/2}}\right).$$

This completes the proof of the first statement in Part (iv). The second one follows from the very same computations.

□

4.6 Discussion

Along this chapter, we have focused on the mathematical properties of the annealed model introduced in Sec. 2.4 of Chap. 2, building upon the research line in probability theory that investigates the features of inhomogeneous random graphs with infinite-mean weights.

Already in Chap. 2, some topological properties of the original SIM were investigated numerically, either analytically (for $\alpha = 1/2$) or semi-analytically (for generic $\alpha \in (0,1)$). Notably, it was found that networks sampled from the SIM feature an empirical degree distribution $P(k)$ exhibiting a scale-free region, characterized by a universal power-law decay $\propto k^{-2}$ (corresponding to a cumulative distribution with decay $\propto k^{-1}$) irrespective of the value of $\alpha \in (0,1)$, followed by a density-dependent cut-off. The results obtained here provide significant additional insights. In particular, we have identified the specific scaling

(or equivalently, as explained in Sec. 4.2 above, the specific hierarchical level) for which the density-dependent cut-off disappears and the tail of the cumulative degree distribution can be rigorously proven (through an independent proof) to become a pure power law with exponent -1, for any $\alpha \in (0, 1)$. Secondly, we have provided a rigorous evaluation of the overall number of triangles and wedges, valid for any α, that supports the outcome of simulations shown in Sec. 2.4, which illustrated the vanishing of the global clustering coefficient in the sparse limit (as opposed to the local clustering coefficient, which remains bounded away from zero as recalled above).

Chapter 5

Conclusions

Motivated by the problem of multiscale modeling, in this thesis we proposed a random graph model built on the principles of renormalization group theory as defined in the configuration space. We refrained from formally developing a renormalization group theory for complex networks in a strict sense, but rather we leveraged its formalism for designing a random graph model meant to harbour the hierarchical organization of complex systems while resting on minimal assumptions.

The model lies on three essential ingredients: *functional scale-invariance of the graph probability*, *statistical independence between edges*, and the dependence of connection probabilities on *additive features* attached to each node (denoted as fitness). This is sufficient for defining a Scale-Invariant random graph Model (SIM), where the functional form of the connection probability among each pair of nodes remains invariant across different levels of observation. The model allows for the generation of any network, at any given level of resolution, in two equivalent ways: hierarchically, by first generating the finest-grained network and then coarse-graining it through progressive partitions, or directly, using appropriately rescaled parameters. The SIM allows for arbitrarily heterogeneous aggregations of nodes, it remains self-consistent irrespective of the topological properties of the networks and does not

require the presence of communities, Laplacian properties, or (geo)metric coordinates. The model has the *option* to integrate dependencies on additional parameters, granting the incorporation of dyadic interactions and global parameters. All these peculiarities are expounded in Chap. 2, where we formally derived the model for binary and undirected graphs. In addition, the SIM can be generalized to encompass *directed* (binary) networks, as we showed in Chap. 3, where we introduced a Directed Scale-Invariant Model (DSIM). The DSIM relies on the same ingredients as the undirected case, with the caveat, however, of considering the connections in opposite directions between each pair of nodes as distinct events that are also *not independent*. Such prescription allowed us to take into account network reciprocity, which plays a key role in the characterization of directed graphs.

Along the three chapters that constitute this thesis, we explored two different applications of the model, that interweave different relevant topics of network science.

In a quenched scenario, the model can comply with empirical networks to make predictions at arbitrary levels of aggregation. In this case, the fitness are quenched since they are measurable quantities gathered from the particular system under study. We used one of the most studied real-world networks as a testbed: the International Trade Network, first (in Sec. 2.3) by focusing on its undirected and binary projection and later (in Sec. 3.4) by considering its directed, binary representation. Notwithstanding the assumption of edge-independence at the basis of the model, we found that the model can effectively replicate the degrees of the ITN and also higher-order topological properties. In particular, we found that a one-parameter fit of the model to the observed network density (and a two-parameter fit to adjust the global reciprocity in the directed case) is enough to capture the profile of node-specific properties as the local clustering coefficient or the average nearest neighbour degree *of each node*, even across several hierarchical levels of resolution. These results illustrate the distinction between scale-free networks (i.e. characterized by power-law degree distributions, which are absent in the ITN) and scale-

invariant networks (i.e. endowed with network formation mechanisms consistent across scales, as found in the ITN). At the same time, using the model to analyze the directed ITN confirmed the importance of taking into account the correlations among reciprocated connections between each pair of nodes to capture the non-trivial emergence of reciprocity.

To explore the asymptotic properties of the SIM, we defined as well an annealed scenario, that allowed us to achieve the thermodynamic limit upon the definition of a proper generative scheme for the fitness variables. In this case, the model naturally led to one-sided Lévy-stable fitness distributions, whose cumulative function is characterized by a tail exponent $-\alpha \in (-1, 0)$ and, therefore, by diverging moments. Along Secs. 2.4 and 3.5 and Chap. 4, we studied the architecture of network realizations drawn from the scale-invariant ensemble, which turned out to exhibit several features observed in real-world systems. We saw either analytically or numerically that the sole requirement of scale-invariance spontaneously leads to scale-free networks with degree distribution featuring a universal power-law decay $P(k_\ell) \propto k^{-2}$, endowed with a density-dependent cut-off, and with realistic assortativity and clustering properties. In particular, in the sparse regime (where the link density is of the order $\sim \frac{1}{N}$), the model appears to be simultaneously scale-free and locally clustered, without resorting to metric distances to produce clustering as a result of triangular inequalities. Contrarily to the local clustering coefficient, the global clustering has been observed to vanish in numerical simulations, and in Chap. 4 an educated estimate of such behaviour has been assessed. In the same chapter, a special regime has been identified where the link density scales as $\sim \frac{log(N)}{N}$. In such a regime, the tail of the degree distribution drops the cut-off due to finite sizes, and the global clustering coefficient converges to 0 as $\propto N^{-1/2}$.

Contrarily to other approaches to network renormalization, the model we propose doesn't rely on any structural assumption or underlying geometry regarding the network topology, and requires only node-specific parameters representing their "importance" in network organization, belonging indeed to the class of fitness models. In particular, this is a fitness

model with the unique peculiarity of remaining a fitness model also upon aggregation of nodes into blocks, as it doesn't require the knowledge of the fitness parameters of the original nodes to provide predictions at higher levels. Being partition-independent, the SIM does not provide a specific hierarchy of nodes' partitions and in this sense is to be considered complementary to other approaches to network renormalization which specifically address such task. In fact, given any such hierarchy, the SIM can guide the renormalization of real networks by providing the node-specific parameters and the *exact* form of the renormalized connection probabilities at the coarse-grained level. Given the margin of freedom in the choice of the nested partitions, the possible presence of scale-invariance (i.e., the compliance with model's predictions) in a given empirical network is to be assessed concurrently with a proper definition of the dyadic factor. With this respect, an important research direction would include a systematic detection of such factors and eventually the identification of classes of real systems featuring scale-invariance based on the same notion of dyadic interactions.

On the practical side, achieving a proper description of a real system through the lens of the SIM model would enable its *multiscale reconstruction*, where one can conveniently move across different scales for example for fitting the parameters at the most convenient resolution (i.e. the finest or the most populated scale) and then provide predictions for any finer or coarser level of aggregation of nodes, or for focusing the attention on a subregion of the system while incorporating the contribution to the overall connectivity in a rest-of-the-network node. Such possibility can also have an impact on the study of dynamical processes taking place on networks and specifically on their critical behavior [150]. As in every fitness model, any process with phase transitions whose critical points depend on the average connectivity of the graph (e.g. epidemic spreading, percolation processes or Ising phase transitions to a large scale ordered state) can be studied in terms of the fitness distribution among different vertices. This model, allowing for a multiscale description, further enables a simoultaneous analysis of the process across hierachical levels where the large ordered state will only depend on the new fitness distribution

induced by the chosen partition at the coarse-grained level, given a proper renormalization rule for the state of block-nodes (akin to the majority rule often employed for defining the renormalized spin of a block-spin). Different choices for the hierarchy of partitions will in general lead to very different scenarios and their analysis can be informative to further characterize phase transitions in multiscale or self-similar systems [151]. This can be crucial for example in the management of infectous diseases [152], allowing to discriminate the main contribution to the spreading of the long- and short-range modes and paving the way to quantitative approximation schemes that calibrate the level of data resolution and the needed computational resources with respect to the accuracy in the description of the epidemics.

The exploited ansätze point at the limitations of the SIM, which is restricted to model binary networks with pairwise interactions and additive node features. Concerning the latter point, different renormalization rules of nodes' features could still be made additive by a proper reparametrization, for instance a multiplicative process reduces to an additive rule for the logarithm of nodes features, while a sublinear or superlinear growth is made linear by defining the fitness as a power of such feature.
More in general, while preliminary results discussed in this thesis look promising, the full spectrum of theoretical implications and practical implementations of the SIM is still far from being thoroughly explored. We refer, for instance, to the demand of a robust node-embedding scheme to infer fitness values when they are not immediately deducible from first principles (as was the case for the international trade network); this would enhance the feasibility of the model for the multiscale analysis of a wider range of empirical systems. A major refinement would also consist in a deeper understanding of the higher-order topological properties of the generated graphs. On the tail of the analysis provided in Chap. 4, ongoing efforts are currently focused on obtaining a precise estimate of the local clustering coefficient as well as on investigating the connectivity properties of the ensemble, specifically drawing a threshold for the presence of dust (i.e. isolated nodes) and for the emergence of a giant connected component.

We hope that this contribution can stimulate further discussions both on the theoretical side, to understand the essential ingredients behind the formation and evolution of real networks, and on the practical side, to design a flexible model that can provide reliable and meaningful multiscale predictions.

Bibliography

[1] E. Garuccio, M. Lalli, and D. Garlaschelli, "Multiscale network renormalization: Scale-invariance without geometry," *Physical Review Research*, vol. 5, no. 4, p. 043 101, 2023.

[2] L. Avena, D. Garlaschelli, R. S. Hazra, and M. Lalli, "Inhomogeneous random graphs with infinite-mean fitness variables," *arXiv preprint arXiv:2212.08462*, 2022.

[3] M. Lalli and D. Garlaschelli, "Geometry-free renormalization of directed networks: Scale-invariance and reciprocity," *arXiv preprint arXiv:2403.00235*, 2024.

[4] M. Newman, *Networks*. Oxford university press, 2018.

[5] R. van der Hofstad, *Random graphs and complex networks*. Cambridge university press, 2016, vol. 43.

[6] P. Erdős and A. Rényi, "On random graphs i," *Publicationes mathematicae*, vol. 6, no. 1, pp. 290–297, 1959.

[7] P. Erdős, A. Rényi, *et al.*, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

[8] P. Erdős and A. Rényi, "On the strength of connectedness of a random graph," *Acta Mathematica Hungarica*, vol. 12, no. 1, pp. 261–267, 1961.

[9] S. Janson, A. Rucinski, and T. Luczak, *Random graphs*. John Wiley & Sons, 2011.

[10] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[11] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[12] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world-wide web," *nature*, vol. 401, no. 6749, pp. 130–131, 1999.

[13] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[14] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the national academy of sciences*, vol. 97, no. 21, pp. 11 149–11 152, 2000.

[15] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *ACM SIGCOMM computer communication review*, vol. 29, no. 4, pp. 251–262, 1999.

[16] A. Broder *et al.*, "Graph structure in the web," *Computer networks*, vol. 33, no. 1-6, pp. 309–320, 2000.

[17] M. Boguná, R. Pastor-Satorras, and A. Vespignani, "Absence of epidemic threshold in scale-free networks with degree correlations," *Physical review letters*, vol. 90, no. 2, p. 028 701, 2003.

[18] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[19] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Structure of growing networks with preferential linking," *Physical review letters*, vol. 85, no. 21, p. 4633, 2000.

[20] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz, "Scale-free networks from varying vertex intrinsic fitness," *Physical review letters*, vol. 89, no. 25, p. 258 702, 2002.

[21] D. V. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, "Hyperbolic geometry of complex networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 82 3 Pt 2, p. 036 106, 2010.

[22] S. Wasserman and K. Faust, "Social network analysis: Methods and applications," 1994.

[23] J. Park and M. E. Newman, "Statistical mechanics of networks," *Physical Review E*, vol. 70, no. 6, p. 066 117, 2004.

[24] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, "The statistical physics of real-world networks," *Nature Reviews Physics*, vol. 1, no. 1, pp. 58–71, 2019.

[25] P. W. Anderson, "More is different: Broken symmetry and the nature of the hierarchical structure of science.," *Science*, vol. 177, no. 4047, pp. 393–396, 1972.

[26] C. Adami, "What is complexity?" *BioEssays*, vol. 24, no. 12, pp. 1085–1094, 2002.

[27] M. Sales-Pardo, R. Guimera, A. A. Moreira, and L. A. N. Amaral, "Extracting the hierarchical organization of complex systems," *Proceedings of the National Academy of Sciences*, vol. 104, no. 39, pp. 15 224–15 229, 2007.

[28] P. Ronhovde *et al.*, "Detection of hidden structures for arbitrary scales in complex physical systems," *Scientific reports*, vol. 2, no. 1, p. 329, 2012.

[29] R. Lambiotte, J.-C. Delvenne, and M. Barahona, "Random walks, markov processes and the multiscale modular organization of complex networks," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 2, pp. 76–90, 2014.

[30] L. P. Kadanoff, "Scaling laws for ising models near T_c," *Physics Physique Fizika*, vol. 2, pp. 263–272, 6 1966. DOI: 10.1103/PhysicsPhysiqueFizika.2.263. [Online]. Available: https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.2.263.

[31] L. P. Kadanoff, *Statistical physics: statics, dynamics and renormalization*. World Scientific Publishing Company, 2000.

[32] K. G. Wilson, "The renormalization group and critical phenomena," *Reviews of Modern Physics*, vol. 55, no. 3, p. 583, 1983.

[33] N. Goldenfeld, *Lectures on phase transitions and the renormalization group*. CRC Press, 2018.

[34] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, no. 7024, pp. 392–395, 2005.

[35] C. Song, S. Havlin, and H. A. Makse, "Origins of fractality in the growth of complex networks," *Nature physics*, vol. 2, no. 4, pp. 275–281, 2006.

[36] P. T. Kovács, M. Nagy, and R. Molontay, "Comparative analysis of box-covering algorithms for fractal networks," *Applied Network Science*, vol. 6, no. 1, pp. 1–37, 2021.

[37] R. Cohen and S. Havlin, "Scale-free networks are ultrasmall," *Physical review letters*, vol. 90, no. 5, p. 058 701, 2003.

[38] G. García-Pérez, M. Boguñá, and M. Á. Serrano, "Multiscale unfolding of real networks by geometric renormalization," *Nature Physics*, vol. 14, no. 6, pp. 583–589, 2018.

[39] M. Á. Serrano, D. Krioukov, and M. Boguná, "Self-similarity of complex networks and hidden metric spaces," *Physical review letters*, vol. 100, no. 7, p. 078 701, 2008.

[40] M. Boguná, F. Papadopoulos, and D. Krioukov, "Sustaining the internet with hyperbolic mapping," *Nature communications*, vol. 1, no. 1, p. 62, 2010.

[41] C. Stegehuis, R. van Der Hofstad, and J. S. Van Leeuwaarden, "Scale-free network clustering in hyperbolic and other random graphs," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 29, p. 295 101, 2019.

[42] P. Villegas, T. Gili, G. Caldarelli, and A. Gabrielli, "Laplacian renormalization group for heterogeneous networks," *Nature Physics*, pp. 1–6, 2023.

[43] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, "Economic networks: The new challenges," *science*, vol. 325, no. 5939, pp. 422–425, 2009.

[44] S. Battiston *et al.*, "Complexity theory and financial regulation," *Science*, vol. 351, no. 6275, pp. 818–819, 2016.

[45] G. Caldarelli, S. Wolf, and Y. Moreno, "Physics of humans, physics for society," *Nature Physics*, vol. 14, no. 9, pp. 870–870, 2018.

[46] L. K. Gallos, C. Song, and H. A. Makse, "A review of fractality and self-similarity in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 386, no. 2, pp. 686–691, 2007.

[47] B. J. Kim, "Geographical coarse graining of complex networks," *Physical review letters*, vol. 93, no. 16, p. 168 701, 2004.

[48] D. Gfeller and P. De Los Rios, "Spectral coarse graining of complex networks," *Physical review letters*, vol. 99, no. 3, p. 038 701, 2007.

[49] L. Cardelli, M. Tribastone, M. Tschaikowski, and A. Vandin, "Maximal aggregation of polynomial dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 114, no. 38, pp. 10 029–10 034, 2017.

[50] F. Radicchi, A. Barrat, S. Fortunato, and J. J. Ramasco, "Renormalization flows in complex networks," *Physical Review E*, vol. 79, no. 2, p. 026 104, 2009.

[51] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, "Skeleton and fractal scaling in complex networks," *Physical review letters*, vol. 96, no. 1, p. 018 701, 2006.

[52] S. Itzkovitz, R. Levitt, N. Kashtan, R. Milo, M. Itzkovitz, and U. Alon, "Coarse-graining and self-dissimilarity of complex networks," *Physical Review E*, vol. 71, no. 1, p. 016 127, 2005.

[53] P. Fronczak, A. Fronczak, and M. Bujok, "Exponential random graph models for networks with community structure," *Physical Review E*, vol. 88, no. 3, p. 032 810, 2013.

[54] T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics*. Springer, 2017.

[55] B. Bollobás, S. Janson, and O. Riordan, "The phase transition in inhomogeneous random graphs," *Random Structures & Algorithms*, vol. 31, no. 1, pp. 3–122, 2007.

[56] R. Rammal, G. Toulouse, and M. A. Virasoro, "Ultrametricity for physicists," *Reviews of Modern Physics*, vol. 58, no. 3, p. 765, 1986.

[57] S. Janson, "Asymptotic equivalence and contiguity of some random graphs," *Random Structures & Algorithms*, vol. 36, no. 1, pp. 26–45, 2010.

[58] I. Norros and H. Reittu, "On a conditionally poissonian graph process," *Advances in Applied Probability*, vol. 38, no. 1, pp. 59–75, 2006.

[59] F. Caron and E. B. Fox, "Sparse graphs using exchangeable random measures," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, no. 5, pp. 1295–1366, 2017.

[60] F. Chung and L. Lu, "Connected components in random graphs with given expected degree sequences," *Annals of combinatorics*, vol. 6, no. 2, pp. 125–145, 2002.

[61] B. Karrer and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical review E*, vol. 83, no. 1, p. 016 107, 2011.

[62] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[63] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, and D. Garlaschelli, "Reconstruction methods for networks: The case of economic and financial systems," *Physics Reports*, vol. 757, pp. 1–47, 2018.

[64] F. Radicchi, D. Krioukov, H. Hartle, and G. Bianconi, "Classical information theory of networks," *Journal of Physics: Complexity*, vol. 1, no. 2, p. 025 001, 2020.

[65] J. Park and M. E. Newman, "Origin of degree correlations in the internet and other networks," *Physical review e*, vol. 68, no. 2, p. 026 112, 2003.

[66] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[67] G. Gaulier and S. Zignago, "Baci: International trade database at the product-level (the 1994-2007 version)," 2010.

[68] D. Garlaschelli and M. I. Loffredo, "Fitness-dependent topological properties of the world trade web," *Physical review letters*, vol. 93, no. 18, p. 188 701, 2004.

[69] A. Almog, T. Squartini, and D. Garlaschelli, "The double role of gdp in shaping the structure of the international trade network," *International Journal of Computational Economics and Econometrics*, vol. 7, no. 4, pp. 381–398, 2017.

[70] A. Almog, T. Squartini, and D. Garlaschelli, "A gdp-driven model for the binary and weighted structure of the international trade network," *New Journal of Physics*, vol. 17, no. 1, p. 013 009, 2015.

[71] A. Almog, R. Bird, and D. Garlaschelli, "Enhanced gravity model of trade: Reconciling macroeconomic and network models," *Frontiers in Physics*, vol. 7, p. 55, 2019.

[72] J. Tinbergen, "Shaping the world economy; suggestions for an international economic policy," 1962.

[73] M. Duenas and G. Fagiolo, "Modeling the international-trade network: A gravity approach," *Journal of Economic Interaction and Coordination*, vol. 8, no. 1, pp. 155–178, 2013.

[74] `https://data.worldbank.org/`.

[75] T. Mayer and S. Zignago, "Notes on cepii's distances measures: The geodist database," 2011.

[76] K. Head, T. Mayer, *et al.*, *Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade*. Citeseer, 2002, vol. 1.

[77] P. Lévy, *Calcul des probabilités*. Gauthier-Villars, 1925.

[78] W. Feller, "An introduction to probability theory and its applications," *1, 2nd*, 1967.

[79] V. V. Uchaikin and V. M. Zolotarev, "Chance and stability," in *Chance and Stability*, de Gruyter, 2011.

[80] I. Schneider, "Jakob bernoulli, ars conjectandi (1713)," in *Landmark Writings in Western Mathematics 1640-1940*, Elsevier, 2005, pp. 88–104.

[81] P.-S. Laplace, *Pierre-Simon Laplace philosophical essay on probabilities: translated from the fifth french edition of 1825 with notes by the translator*. Springer Science & Business Media, 1998, vol. 13.

[82] P. Lévy, "L'addition des variables aléatoires définies sur une circonférence," *Bulletin de la Société mathématique de France*, vol. 67, pp. 1–41, 1939.

[83] K. Penson and K. Górska, "Exact and explicit probability densities for one-sided lévy stable distributions," *Physical review letters*, vol. 105, no. 21, p. 210 604, 2010.

[84] K. Górska and K. Penson, "Lévy stable distributions via associated integral transform," *Journal of mathematical physics*, vol. 53, no. 5, p. 053 302, 2012.

[85] A. Wintner, "Stable distributions and laplace transforms," *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, vol. 10, no. 3-4, pp. 127–134, 1956.

[86] H. Pollard, "The representation of $e^{-x^{\lambda}}$ as a laplace integral," *Bulletin of the American Mathematical Society*, vol. 52, no. 10, pp. 908–910, 1946.

[87] M. Zheng, G. García-Pérez, M. Boguñá, and M. Á. Serrano, "Scaling up real networks by geometric branching growth," *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, e2018994118, 2021.

[88] D. J. Aldous, "Representations for partially exchangeable arrays of random variables," *Journal of Multivariate Analysis*, vol. 11, no. 4, pp. 581–598, 1981.

[89] P. Diaconis and S. Janson, "Graph limits and exchangeable random graphs," *arXiv preprint arXiv:0712.2749*, 2007.

[90] P. van der Hoorn, G. Lippner, and D. Krioukov, "Sparse maximum-entropy random graphs with a given power-law degree distribution," *Journal of Statistical Physics*, vol. 173, no. 3, pp. 806–844, 2018.

[91] A. P. Kartun-Giles, D. Krioukov, J. P. Gleeson, Y. Moreno, and G. Bianconi, "Sparse power-law network model for reliable statistical predictions based on sampled data," *Entropy*, vol. 20, no. 4, p. 257, 2018.

[92] D. Krioukov, M. Kitsak, R. S. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá, "Network cosmology," *Scientific reports*, vol. 2, no. 1, pp. 1–6, 2012.

[93] A. Barrat and M. Weigt, "On the properties of small-world network models," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 13, no. 3, pp. 547–560, 2000.

[94] L. Ostroumova Prokhorenkova and E. Samosvat, "Global clustering coefficient in scale-free networks," in *International Workshop on Algorithms and Models for the Web-Graph*, Springer, 2014, pp. 47–58.

[95] E. Estrada, "When local and global clustering of networks diverge," *Linear Algebra and its Applications*, vol. 488, pp. 249–263, 2016.

[96] E. Candellero and N. Fountoulakis, "Clustering and the hyperbolic geometry of complex networks," in *International Workshop on Algorithms and Models for the Web-Graph*, Springer, 2014, pp. 1–12.

[97] R. van der Hofstad, P. Van der Hoorn, N. Litvak, and C. Stegehuis, "Limit theorems for assortativity and clustering in null models for scale-free networks," *Advances in applied probability*, vol. 52, no. 4, pp. 1035–1084, 2020.

[98] P. Colomer-de-Simón and M. Boguñá, "Double percolation phase transition in clustered complex networks," *Physical Review X*, vol. 4, no. 4, p. 041 020, 2014.

[99] K. S. Gleditsch, "Expanded trade and gdp data," *Journal of Conflict Resolution*, vol. 46, no. 5, pp. 712–724, 2002.

[100] D. Garlaschelli and M. I. Loffredo, "Structure and evolution of the world trade network," *Physica A: Statistical Mechanics and its Applications*, vol. 355, no. 1, pp. 138–144, 2005.

[101] M. Bardoscia *et al.*, "The physics of financial networks," *Nature Reviews Physics*, vol. 3, no. 7, pp. 490–507, 2021.

[102] L. N. Ialongo *et al.*, "Reconstructing firm-level interactions in the dutch input–output network from production constraints," *Scientific Reports*, vol. 12, no. 1, pp. 1–12, 2022.

[103] J. Scott, *What is social network analysis?* Bloomsbury Academic, 2012.

[104] M. E. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Physical Review E*, vol. 66, no. 3, p. 035 101, 2002.

[105] D. Garlaschelli and M. I. Loffredo, "Patterns of link reciprocity in directed networks.," *Physical review letters*, vol. 93 26 Pt 1, p. 268 701, 2004.

[106] J. H. Martínez, S. Romero, J. J. Ramasco, and E. Estrada, "The world-wide waste web," *Nature communications*, vol. 13, no. 1, p. 1615, 2022.

[107] D. Garlaschelli and M. I. Loffredo, "Multispecies grand-canonical models for networks with reciprocity," *Physical Review E*, vol. 73, no. 1, p. 015 101, 2006.

[108] T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, "Reciprocity of weighted networks," *Scientific reports*, vol. 3, no. 1, pp. 1–9, 2013.

[109] C. Seshadhri, A. Pinar, N. Durak, and T. G. Kolda, "Directed closure measures for networks with reciprocity," *Journal of Complex Networks*, vol. 5, no. 1, pp. 32–47, Apr. 2016, ISSN: 2051-1310. DOI: 10.1093/comnet/cnv032. eprint: https://academic.oup.com/comnet/article-pdf/5/1/32/10253477/cnv032.pdf. [Online]. Available: https://doi.org/10.1093/comnet/cnv032.

[110] T. Squartini and D. Garlaschelli, "Triadic motifs and dyadic self-organization in the world trade network," in *International Workshop on Self-Organizing Systems*, Springer, 2012, pp. 24–35.

[111] T. Squartini, I. Van Lelyveld, and D. Garlaschelli, "Early-warning signals of topological collapse in interbank networks," *Scientific reports*, vol. 3, no. 1, p. 3357, 2013.

[112] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, "Spectrum of large random asymmetric matrices," *Physical review letters*, vol. 60, no. 19, p. 1895, 1988.

[113] M. Boguná and M. Á. Serrano, "Generalized percolation in random directed networks," *Physical Review E*, vol. 72, no. 1, p. 016 106, 2005.

[114] M. A. Nowak and K. Sigmund, "Evolution of indirect reciprocity," *Nature*, vol. 437, no. 7063, pp. 1291–1298, 2005.

[115] V. Zlatić and H. Štefančić, "Influence of reciprocal edges on degree distribution and degree correlations," *Physical Review E*, vol. 80, no. 1, p. 016 117, 2009.

[116] L. D. Molm, "The structure of reciprocity," *Social psychology quarterly*, vol. 73, no. 2, pp. 119–131, 2010.

[117] L. K. Gallos, D. Rybski, F. Liljeros, S. Havlin, and H. A. Makse, "How people interact in evolving online affiliation networks," *Physical Review X*, vol. 2, no. 3, p. 031 014, 2012.

[118] A. Allard, M. Á. Serrano, and M. Boguñá, "Geometric description of clustering in directed networks," *Nature Physics*, pp. 1–7, 2023.

[119] B. Kovács and G. Palla, "Model-independent embedding of directed networks into euclidean and hyperbolic spaces," *Communications Physics*, vol. 6, no. 1, p. 28, 2023.

[120] T. Squartini and D. Garlaschelli, "Analytical maximum-likelihood method to detect patterns in real networks," *New Journal of Physics*, vol. 13, no. 8, p. 083 001, 2011. DOI: 10.1088/1367-2630/13/8/083001. [Online]. Available: https://doi.org/10.1088/1367-2630/13/8/083001.

[121] V. Gemmetto, T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, "Multiplexity and multireciprocity in directed multiplexes," *Physical Review E*, vol. 94, no. 4, p. 042 316, 2016.

[122] J. Väisälä, "Gromov hyperbolic spaces," *Expositiones Mathematicae*, vol. 23, no. 3, pp. 187–231, 2005, ISSN: 0723-0869. DOI: https://doi.org/10.1016/j.exmath.2005.01.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0723086905000113.

[123] T. Das, D. Simmons, and M. Urbański, *Geometry and dynamics in Gromov hyperbolic metric spaces*. American Mathematical Soc., 2017, vol. 218.

[124] M. Bukatin, R. Kopperman, S. Matthews, and H. Pajoohesh, "Partial metric spaces," *The American Mathematical Monthly*, vol. 116, no. 8, pp. 708–718, 2009.

[125] P. Hitzler and A. K. Seda, "Dislocated topologies," *J. Electr. Eng*, vol. 51, no. 12, pp. 3–7, 2000.

[126] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs," *Journal of the american Statistical association*, vol. 76, no. 373, pp. 33–50, 1981.

[127] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annals of Mathematics*, pp. 325–327, 1958.

[128] V. Girko, "Elliptic law," *Theory of Probability & Its Applications*, vol. 30, no. 4, pp. 677–690, 1986.

[129] M. Di Vece, D. Garlaschelli, and T. Squartini, "Gravity models of networks: Integrating maximum-entropy and econometric approaches," *Physical Review Research*, vol. 4, no. 3, p. 033 105, 2022.

[130] M. Di Vece, D. Garlaschelli, and T. Squartini, "Reconciling econometrics with continuous maximum-entropy network models," *Chaos, Solitons & Fractals*, vol. 166, p. 112 958, 2023.

[131] R. Summers and A. Heston, "The penn world table (mark 5): An expanded set of international comparisons, 1950–1988," *The Quarterly Journal of Economics*, vol. 106, no. 2, pp. 327–368, 1991.

[132] C. I. Agency, *World factbook*. [Online]. Available: `http://www.cia.gov/cia/publications/%20factbook/index.html`.

[133] I. M. F. G. S. Division, *Direction of trade statistics*. International Monetary Fund, 1993.

[134] J. Faber and T. Nierop, "World export data, 1948-1983 (icpsr no. 9116). amsterdam: University of amsterdam (producers)," *Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributors)*, 1989.

[135] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011.

[136] G. E. Carlsson, F. Mémoli, *et al.*, "Characterization, stability and convergence of hierarchical clustering methods.," *J. Mach. Learn. Res.*, vol. 11, no. Apr, pp. 1425–1470, 2010.

[137] M. Boguná and R. Pastor-Satorras, "Class of correlated random networks with hidden variables," *Physical Review E*, vol. 68, no. 3, p. 036 112, 2003.

[138] R. van der Hofstad, *Random graphs and complex networks, Volume 2*. https://www.win.tue.nl/ rhofstad/NotesRGCNII.pdf, 2022.

[139] G. J. Rodgers, K. Austin, B. Kahng, and D. Kim, "Eigenvalue spectra of complex networks," *Journal of Physics A: Mathematical and General*, vol. 38, no. 43, p. 9431, 2005. DOI: 10.1088/0305-4470/38/43/003. [Online]. Available: https://dx.doi.org/10.1088/0305-4470/38/43/003.

[140] S. Bhamidi, R. van der Hofstad, and S. Sen, "The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs," *Probability Theory and Related Fields*, vol. 170, pp. 387–474, 2018.

[141] R. van der Hofstad and J. Komjáthy, "When is a scale-free graph ultra-small?" *Journal of Statistical Physics*, vol. 169, pp. 223–264, 2017.

[142] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski, "Distances in random graphs with infinite mean degrees," *Extremes*, vol. 8, no. 3, pp. 111–141, 2005.

[143] G. S.
bibinitperiod M. Taqqu, "Stable non-gaussian random processes," *Stochastic models with infinite variance. Stochastic Modeling. Chapman & Hall, New York*, 1994.

[144] S. Resnick and G. Samorodnitsky, "Tauberian theory for multivariate regularly varying distributions with application to preferential attachment networks," *Extremes*, vol. 18, no. 3, pp. 349–367, 2015.

[145] R. van der Hofstad, A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and C. Stegehuis, "Local clustering in scale-free networks with hidden variables," *Phys. Rev. E*, vol. 95, p. 022 307, 2 2017. DOI: 10.1103/PhysRevE.95.022307. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.95.022307.

[146] R. Michielan, N. Litvak, and C. Stegehuis, "Detecting hyperbolic geometry in networks: Why triangles are not enough," *Physical Review E*, vol. 106, no. 5, p. 054 303, 2022.

[147] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*. Cambridge university press, 1989.

[148] T. M. Anders Hedegaard Jessen, "Regularly varying functions," eng, *Publications de l'Institut Mathématique*, vol. 80(94), no. 100, pp. 171–192, 2006. [Online]. Available: http://eudml.org/doc/258304.

[149] C. M. Bender, S. Orszag, and S. A. Orszag, *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 1999, vol. 1.

[150] S. N. Dorogovtsev, A. V. Goltsev, and J. F. Mendes, "Critical phenomena in complex networks," *Reviews of Modern Physics*, vol. 80, no. 4, p. 1275, 2008.

[151] M. Á. Serrano, D. Krioukov, and M. Boguñá, "Percolation in self-similar networks," *Physical review letters*, vol. 106, no. 4, p. 048 701, 2011.

[152] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the national academy of sciences*, vol. 106, no. 51, pp. 21 484–21 489, 2009.