



Universiteit
Leiden
The Netherlands

On the renormalization of random network models

Catanzaro, A.

Citation

Catanzaro, A. (2026, September 10). *On the renormalization of random network models*. Retrieved from <https://hdl.handle.net/1887/4309543>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4309543>

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 1

Introduction

This thesis deals with the broad topic of renormalization of complex networks, with a particular focus on the topological nature of such models of complex systems. It is divided into two main parts.

In the first part, we discuss the characterization of the spectrum and the clustering function of a recently proposed network model with independent edges. This model can be regarded as the 'fixed point' of a topology-preserving aggregation process that allows to deal with multi-scale clustered data. We investigate its properties in the regime where the node features are distributed according to an infinite-mean stable probability density function, making the model scale-invariant both in the functional form of the link probability and in the hidden variable distribution. In particular, in Chapter 2 we analyze the model spectral properties, while in Chapter 3 we discuss its clustering function.

In the second part, we elaborate on a specific renormalization scheme that is equivalent to marginalization over disordered uncertainty, in the context of random network models that might display a correlation (possibly higher-order) among links. In particular in Chapter 4 we introduce the marginalization framework, and show two classes of effectively solvable models exist. In Chapter 5 we analyze in detail the first class of models that become solvable under the marginalization flow, and elaborate on its applications to temporal networks. Finally, in Chapter 6 we tackle the second class of those models, by providing a general solution framework for II order Random Network Models.

Introduction to Part I

The first part of this thesis is concerned with the mathematical analysis of a recently introduced model of random networks, the *Multi-Scale Model* (MSM) [1, 2, 3], and in particular with two of its structural properties: the spectrum of its adjacency matrix and its clustering function behaviour. Before presenting the technical results, we offer here a wider perspective on the motivations for studying this model and on the significance of the findings collected in Chapters 2 and 3.

Why the Multi-Scale Model?

The scientific study of networks, from the Internet and social platforms to biological regulatory systems and financial markets, has produced an increasingly sophisticated catalogue of features that characterise real-world, complex networks [4, 5, 6]. Among the most robust empirical observations, two stand out for their ubiquity across completely different domains. First, real networks are *heterogeneous*: a relatively small number of highly connected nodes (hubs) coexist with a very large number of poorly connected ones, and the probability of observing a node with a given number of connections falls off as a power law with an exponent typically close to two [4, 5, 7]. Second, real networks are *clustered*: two nodes that share a common neighbour are much more likely to be directly connected to each other than two randomly chosen nodes [4, 8, 9]. This second feature, sometimes described by saying that “the friend of my friend is also my friend”, is known as the clustering effect, and it is responsible for the tight-knit community structure observable in, for instance, social or collaboration networks.

Reproducing both of these properties simultaneously within a single, tractable random graph model has proven surprisingly elusive. Classical random graph models, such as the Erdős–Rényi model [10], connect nodes independently and with equal probability, thus capturing neither heterogeneity nor clustering. Hidden-variable models - also called fitness models [11, 12, 13] - address the heterogeneity problem by assigning each node a numerical weight (a fitness, or hidden variable) drawn from a heavy-tailed distribution, and connecting pairs of nodes with a probability that depends on their respective weights. This framework successfully reproduces the power-law degree distribution observed in real networks [5, 7, 11, 12]. However, models in which edges are drawn independently given the node fitnesses generically produce vanishingly small clustering as the network grows, unless nodes are additionally embedded in some underlying geometric space [14, 15, 16, 17, 18, 19]. The price of geometry is conceptual and practical: one must postulate the existence of a latent metric space, choose its dimension and curvature, and accept that many network properties are then determined by that geometric structure rather than emerging from the interaction rules alone [16, 20, 21].

The Multi-Scale Model and its renormalization origin

The model studied in this first part of the thesis takes a completely different starting point. Rather than assuming geometry, it asks: which random graph model is *self-similar under aggregation*? More concretely, if one groups nodes into “supernodes” - each supernode inheriting the combined fitnesses of its members - and declares two supernodes to be connected whenever any of their constituent nodes are connected, which model remains, in functional form, unchanged by this coarse-graining operation? This is a renormalization group question, phrased in the language of random networks, and the non-trivial answer is essentially unique for independent edges: it is the MSM [1, 2, 3], in which the connection probability between two nodes is a specific function of their fitnesses, and the fitness distribution belongs to the class of stable probability distributions with infinite mean [22].

The model can be regarded as a *fixed point* of a topology-preserving aggregation flow: iterating the coarse-graining procedure leaves the model connection probability invariant. This is a property shared by critical systems in statistical physics and it suggests that the MSM occupies a special, structurally stable position in the space of all possible random graph models.

From this perspective, understanding the MSM is not just an exercise in the analysis of a particular model; it is a step towards understanding what kinds of network structure are robust under changes of resolution or observation scale.

Spectral properties and discrete scale invariance

Chapter 2 is devoted to the spectral analysis of the MSM’s adjacency matrix. The spectrum of a random graph - that is, the collection of eigenvalues and eigenvectors of its adjacency matrix - encodes an enormous amount of information about the network’s structure [25]. Eigenvalues and eigenvectors are central to spectral clustering algorithms [26, 27, 28], to the study of diffusion and random walks on graphs, to the analysis of epidemic spreading, and to many methods of network inference.

For most random graph models with finite-mean fitnesses, the spectrum decomposes cleanly into two regions: a bulk, whose limiting distribution is described by results from classical random matrix theory [29, 30, 31], and a small number of isolated outlier eigenvalues that detach from the bulk and carry structural information about communities, hubs, and other macroscopic features [32, 33, 34]. The number of such outliers is typically small, independent of network size.

The MSM, by contrast, behaves in a radically different way. Because the fitness distribution has infinite mean, standard random matrix techniques break down entirely. What one finds instead is that an extensive number of eigenvalues - a number that increases (albeit slowly) with the size of the network - emerge from the bulk. These outlier eigenvalues are not scattered randomly in the complex plane: they actually lie at the intersections of the real axis with a logarithmic spiral, whose equation we give in full detail. The corresponding eigenvectors display complex-valued scaling exponents and log-periodic oscillations: their entries, when plotted as a function of the node fitness, do not follow a simple power law but instead oscillate (to preserve

orthogonality) in a pattern whose period is set by the logarithm of the network size.

This log-periodicity is the network signature of *discrete scale invariance*: the property that the system is invariant not under continuous rescaling, but only under rescaling by a specific factor (or its integer powers). Discrete scale invariance is known to arise in a variety of physical and biological systems displaying log-periodic corrections to scaling laws [35, 36, 37, 38], and its appearance here connects the network renormalization programme to a broader body of research on complex scaling phenomena.

Beyond the outlier eigenvalues, the bulk of the MSM spectrum is also unusual. Its edge is of the same order as the outlier eigenvalues - both scale as the square root of the number of nodes - unlike in models with finite-mean fitnesses, where the bulk is parametrically smaller than the largest outlier. This “bulk catching up” phenomenon makes it extremely hard to separate structural information from noise using a simple threshold on eigenvalue magnitude, and calls for new analytical strategies.

Clustering without geometry

Chapter 3 turns to the clustering properties of the MSM. As mentioned above, one of the long-standing criticisms of independent-edge network models is their inability to reproduce meaningful clustering [4, 14, 15]. The main result of Chapter 3 is a rigorous proof that the MSM, in the infinite-mean regime, is an exception to this rule: it naturally generates a non-trivial clustering function without any appeal to an underlying geometric space.

The mechanism is subtle. In the infinite-mean regime, the fitness of the nodes spans many orders of magnitude. Nodes with very high fitness - the hubs - are connected to virtually everyone else, so their neighbours form a group among which further connections are relatively rare: the hub’s local clustering coefficient is close to zero. Nodes with low or moderate fitness, on the other hand, are connected almost exclusively to hubs; since leaf nodes with degree small but greater than one prevalently connect to hubs, and hubs are connected in a tight clique, the leaf local clustering coefficient approaches one. The overall result is a network that is, in a precise asymptotic sense, “almost everywhere” clustered, with only the vanishingly small fraction of nodes that are extreme hubs remaining unclustered. This is strikingly consistent with empirical observations in social and technological networks, where high-degree nodes tend to bridge otherwise weakly connected communities rather than being embedded in tightly knit cliques [8, 39, 40].

The analysis also reveals an interesting lack of self-averaging: the number of low degree nodes, when computed across different realisations of the same random graph model, exhibits significant fluctuations that do not vanish as the network grows. This is a consequence of the infinite-mean fitness distribution, which allows individual realisations to differ dramatically in the presence or absence of extreme hubs. The lack of self-averaging has practical implications for the statistical inference of network parameters from a single observed network.

Introduction to Part II

The second part of this thesis shifts the focus from the characterisation of a specific fixed-point model to the development of a general renormalization theory applicable to a broad class of random network models. The unifying thread is the idea of *marginalization*: when parts of a network are hidden, unresolved, or simply unobserved, the appropriate way to account for this ignorance is to average the model over the missing degrees of freedom, rather than to discard them or to treat them as fixed at some nominal value. This operation turns out to be precisely equivalent to a renormalization group (RG) transformation in the context of exponential random graph models. Part II develops this equivalence systematically and exploits it to derive exact and approximate solutions for several families of network models in which edges are not drawn independently.

Why correlations matter

The independent-edge assumption is a powerful simplification, and it underlies the tractability of many of the most widely used random graph models. However, real networks frequently display correlations among edges that cannot be captured by independent-edge models. In a social network, the probability that two individuals are connected is influenced not only by their individual attributes but also by the presence or absence of mutual acquaintances - a genuinely higher-order effect. In a financial network, the probability that two institutions share an exposure depends on the broader structure of counterparty relationships, creating systemic-risk profiles that independent-edge models systematically underestimate [17, 41, 42, 43]. In a biological network of gene regulation, the formation of feedback loops and motifs creates statistical dependencies among edges that are functionally significant [44, 45, 46].

Exponential random graph models (ERGMs) provide the most natural language for encoding such dependencies [47, 48]: they assign a probability to each network configuration in proportion to the exponential of a linear combination of graph statistics (edge counts, triangle counts, degree sequences, and so on), exactly as the Boltzmann weight in statistical mechanics assigns probabilities to microscopic configurations. Within this framework, independent-edge models correspond to Hamiltonians that contain only terms linear in individual edges, while correlated models involve Hamiltonians with higher-order terms coupling different edges.

The challenge with correlated models is their intractability: the normalisation constant (the partition function) is generically computationally hard to evaluate, and closed-form expressions for observables are available only in very special cases [49, 50]. The renormalization approach developed in Part II provides a systematic strategy to identify and exploit these special cases.

Marginalization as renormalization

The starting point of Chapter 4 is an observation that connects two apparently unrelated concepts. On one hand, in the theory of disordered spin systems (such as spin glasses), a standard technique for studying models on irregular or partially observed graphs is *decimation*: one integrates out a subset of the degrees of freedom, obtaining an effective model on the remaining sites [51, 52, 53]. On the other hand, in network science, the standard probabilistic approach is to write a probability distribution over the observable part of the network alone, and run the inference machinery on the reduced complete statistics.

The central insight of Chapter 4 is that these two operations are, in the context of exponential random graph models, complementary. Marginalising over unobserved network components is the same as performing a decimation step in the renormalization group sense. This equivalence is not merely formal: it allows one to import the entire conceptual machinery of the RG - fixed points, universality classes, flow equations, relevant and irrelevant directions - into the study of network models [54, 55, 56].

A crucial consequence of this equivalence is that marginalization generically introduces effective higher-order interactions among the remaining variables. If the original model has only pairwise edge couplings, integrating out a subset of edges will in general produce a model in which the probabilities of the remaining edges depend on each other in a complex way, through higher-order interaction terms. This is the network analogue of the well-known fact that coarse-graining a lattice model generates effective couplings between spins that were previously non-interacting [51]. Managing the proliferation of these higher-order terms is the central technical challenge of Part II, and Chapters 5 and 6 are devoted to two complementary strategies for dealing with it.

Exactly solvable models and the role of topology

The first strategy, developed in Chapter 5, is to identify network topologies and coupling structures for which the RG flow *closes exactly* within a finite-dimensional family of models. This means that, no matter how many marginalization steps one performs, the resulting effective model remains within the same parametric family, with only the numerical values of the coupling constants changing.

Chapter 5 provides a complete analysis of one such exactly solvable class: random network models defined on a one-dimensional chain. This is the network analogue of the Ising chain [58, 59, 60], and the exact solvability has the same origin: the locality of the interactions, which prevents the generation of long-range effective couplings under decimation. Building on this base case, Chapter 5 systematically analyses the effect of adding disorder - in the form of spatially varying coupling constants or link probabilities - to the chain model. Two types of disorder are considered: sparse disorder (affecting only a finite fraction of sites) and uniform disorder (affecting all sites in a statistically independent way). In both cases, the exact RG equations can be derived and the induced probability flows characterised analytically.

The exact solvability of the chain model is not merely a mathematical curiosity.

Chapter 5 shows that the RG equations for the chain can be mapped onto a variety of network phenomena of practical interest: peer effects (the tendency of individuals to adopt the behaviour of their network neighbours), preferential attachment (the rich-get-richer mechanism underlying the growth of many real networks), susceptibility effects (in which the formation of a connection depends on the current state of a node), and reinforcement learning dynamics. By understanding the RG flow of the chain model, one gains insight into the universality classes that govern all of these phenomena.

Second-order models and the harmonic oscillator of random graphs

The second strategy, developed in Chapter 6, takes a different approach. Rather than restricting to a specific topology for which the RG flow closes exactly, Chapter 6 studies the class of models to which most correlated network models flow under marginalization when higher-order interactions are irrelevant. In this case, the flow identifies a specific class of “second-order” models, characterised by pairwise interactions between *edges* (rather than between nodes). The name “second-order random network model” reflects this pairwise edge coupling structure, which is the simplest form of edge correlation.

The central result of Chapter 6 is that second-order random network models can be solved exactly and in full generality, for arbitrary interaction structures. The key observation is that the problem of computing the partition function of a second-order model can be recast as a problem in algebraic graph theory: specifically, as the diagonalization of the *link-adjacency matrix*, a matrix that encodes the coupling between pairs of edges rather than between nodes. The eigenmodes of this matrix determine the phase structure of the model, and the general single-link probability can be expressed as a function of these eigenmodes.

This structural insight has far-reaching consequences. It provides a unified framework for solving a large family of models that had previously been treated separately and in special cases [49, 50]. It reveals that the phase behaviour of correlated network models can be much richer than that of independent-edge models: depending on the eigenvalue structure of the link-adjacency matrix, a second-order model can exhibit multiple coexisting phases, complex phase diagrams, and transitions analogous to those found in condensed-matter physics. It also establishes a connection between network science and algebraic graph theory that has not previously been exploited in this context.

The title “harmonic oscillator of random graphs” reflects both the central role of the second-order coupling structure - analogous to the quadratic potential of a harmonic oscillator, which is the simplest exactly solvable problem in quantum mechanics - and the expectation that the methods developed in Chapter 6 will serve as a building block for the analysis of more complex, higher-order network models, much as the harmonic oscillator serves as a building block in perturbation theory.