

Resistance distance distribution in large sparse random graphs

Pawat Akara-pipattana^{a,b}, Thiparat Chotibut^{a,b} and Oleg Evnin^{a,c}

^a *Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand*

^b *Chula Intelligent and Complex Systems, Faculty of Science, Chulalongkorn University, Bangkok, Thailand*

^c *Theoretische Natuurkunde, Vrije Universiteit Brussel and The International Solvay Institutes, Brussels, Belgium*

akarapawat@gmail.com, thiparatc@gmail.com, oleg.evnin@gmail.com

ABSTRACT

We consider an Erdős-Rényi random graph consisting of N vertices connected by randomly and independently drawing an edge between every pair of them with probability c/N so that at $N \rightarrow \infty$ one obtains a graph of finite mean degree c . In this regime, we study the distribution of resistance distances between the vertices of this graph and develop an auxiliary field representation for this quantity in the spirit of statistical field theory. Using this representation, a saddle point evaluation of the resistance distance distribution is possible at $N \rightarrow \infty$ in terms of an $1/c$ expansion. The leading order of this expansion captures the results of numerical simulations very well down to rather small values of c ; for example, it recovers the empirical distribution at $c = 4$ or 6 with an overlap of around 90%. At large values of c , the distribution tends to a Gaussian of mean $2/c$ and standard deviation $\sqrt{2/c^3}$. At small values of c , the distribution is skewed toward larger values, as captured by our saddle point analysis, and many fine features appear in addition to the main peak, including subleading peaks that can be traced back to resistance distances between vertices of specific low degrees and the rest of the graph. We develop a more refined saddle point scheme that extracts the corresponding degree-differentiated resistance distance distributions. We then use this approach to recover analytically the most apparent of the subleading peaks that originates from vertices of degree 1. Rather intuitively, this subleading peak turns out to be a copy of the main peak, shifted by one unit of resistance distance and scaled down by the probability for a vertex to have degree 1. We comment on a possible lack of smoothness in the true $N \rightarrow \infty$ distribution suggested by the numerics.

1 Introduction

Of all the notions of distance that may be defined on graphs [1, 2], perhaps the most evident one is given by the shortest path, or geodesic [3], distance: the minimal number of edges one must traverse in order to transit from vertex i to vertex j . While intuitive and visual, this notion of distance is limited in that it does not fully capture the ease or difficulty of reaching point j from point i by navigating the graph edges. It does not say whether there is only one path of minimal length or many such paths, whether these paths can be straightforwardly located, or whether alternative paths are considerably or only slightly longer.

A complementary viewpoint is provided by the *resistance distance* [4–8] defined by assigning resistances of 1 ohm to the graph edges and measuring the ordinary electric resistance between vertices i and j . This quantity is naturally expressed through the inverse of the *graph Laplacian* and is therefore closely related [9] to diffusion and random walks on graphs [10], measuring the time a random walker starting at point i typically needs to reach point j , which leads to the alternative name *commute distance*. Resistance distances thus in principle capture properties of all possible paths, though there are important qualifications to this statement [11], see below. In view of these appealing properties, resistance distances have surfaced in research on subjects as diverse as theoretical physics [12, 13], chemistry and bioinformatics [14–19], mathematical graph theory [20–22], data analysis and computer science [23–35]. Studying resistance distance can also be useful for understanding nutrient transport in leaf vascular networks, as hydraulic conductance of laminar flows in this setting can be equivalently studied in terms of electrical conductance in resistor networks [36–39].

One important characterization of graph distances is given by how they behave in random graphs. For the shortest path distance, this question has been studied rather extensively [40–45]. By contrast, studies of the distribution of resistance distances have mostly been limited to concentration phenomena, as in [11, 46, 47]. Indeed, under the assumption that the mean degree grows with the number of vertices (this growth may be very modest, for example, logarithmic), in the large graph limit, the resistance distance between vertices i and j of degrees d_i and d_j almost surely equals $1/d_i + 1/d_j$, the relevance of this number pointed out already in [10]. These concentration phenomena inspired the critique of resistance distance developed in [11], as the simple number $1/d_i + 1/d_j$ cannot possibly hold any refined information on the path properties of graphs, and some modifications of the usual resistance distance definition have been proposed [48] to address these shortcomings.¹ At the same time, concentration phenomena do not occur if the mean degree stays finite at $N \rightarrow \infty$ and the resistance distributions are not only nontrivial, but also ornately shaped. This is straightforwardly verified by simulating a finite mean degree Erdős-Rényi graph obtained by randomly and independently connecting N points with probability c/N per edge, where N is large and the mean vertex degree c is fixed.

Our goal in this paper is precisely to develop an analytic picture of resistance distance distributions in finite-mean-degree Erdős-Rényi graphs. (In other words, our focus is on large, sparse graphs with $c \ll N$.) While the motivations for studying resistance distances

¹We mention additionally that studies have been undertaken of systems where the resistance values assigned to the edges, rather than the underlying graph structure, are random [49]. The problem of choosing the edge resistance values on a given graph to minimize the overall average resistance has been considered in [50].

are broad and interdisciplinary, the methodology we employ is decidedly that of theoretical physics, and more specifically statistical field theory [51, 52]. We shall first develop an auxiliary field representation for the resistance distance distribution in the spirit of the field-theoretic Hubbard-Stratonovich transformation, which will convert the resistance distance computation to the analysis of a statistical field theory on a complete graph with N vertices. Then, in a manner common for large N limits [53] in statistical and quantum field theory, and as a particular realization of the ‘mean field’ principles, we shall identify a saddle point that dominates the auxiliary field theory computation, recovering an estimate for the resistance distance distribution. Similar techniques have been applied to a variety of random graph problems in [54–57]. (Earlier work applying more conventional probabilistic and combinatorial methods to related spectral problems for random graphs can be found in [58].) The level of rigor in our treatment will likewise be typical of statistical field theory, and our approach is of empirical and heuristic nature: we shall identify saddle points that plausibly dominate the quantity of interest and develop estimates based on these saddle points, without attempting to rigorously control the accuracy of these estimates by analytic methods. The true judge of our endeavors is their agreement with numerical simulations that we shall systematically report.

Our study is organized as follows: first, in section 2, we shall provide the basic setup for the resistance distance distribution of Erdős-Rényi graphs and develop our auxiliary field representation. Then, in section 3, we shall describe a saddle point treatment of this auxiliary field representation at leading order in $1/c$ in a simplified setup that ignores the fluctuations of the graph Laplacian determinant, and in section 4, give a full justification for the simplified analysis of section 3. In section 5, we shall focus on minor features of the resistance distance distribution not captured by our saddle point analysis and develop an analytic description of the simplest of these features. We shall conclude with a discussion and some tentative remarks on the roughness of the true resistance distance distribution at $N \rightarrow \infty$.

2 Resistance distance distribution and its auxiliary field representation

As customary in random graph theory, we shall characterize graphs with N vertices by their $N \times N$ *adjacency matrices* A_{ij} , which are taken to be symmetric with a vanishing diagonal. One assigns 1 to A_{ij} if there is an edge connecting vertices i and j , and 0 otherwise.

To describe a large sparse Erdős-Rényi random graph in this language, we treat A_{ij} with $i < j$ as independent random variables taking value 1 with probability c/N and 0 with probability $1 - c/N$. The expectation values of the vertex degrees

$$d_i \equiv \sum_{j=1}^N A_{ij} \tag{1}$$

are then $c(N - 1)/N \approx c$ in the large N limit. It is convenient for our purposes to recast this ensemble in the exponential random graph language [59] better adapted for statistical physics considerations. To each configuration A_{ij} one assigns a Boltzmann weight $\exp[\ln(\frac{c}{N}) \sum_{i < j} A_{ij}] = (c/N)^{\sum_{i < j} A_{ij}}$, so that $\ln(N/c)$ plays the role of inverse temperature

conjugate to the total number of graph edges. Then, the expectation value of any observable $F(\mathbf{A})$ is given by

$$\langle F \rangle = \frac{1}{Z} \sum_{\{\mathbf{A}\}} F(\mathbf{A}) \left(\frac{c}{N} \right)^{\sum_{i < j} A_{ij}}, \quad (2)$$

with the partition function $Z \equiv \sum_{\{\mathbf{A}\}} \left(\frac{c}{N} \right)^{\sum_{i < j} A_{ij}} \approx e^{c(N-1)/2}$, and $\sum_{\{\mathbf{A}\}}$ defined as summing over $A_{ij} \in \{0, 1\}$ for each $i < j$.

For any graph, we can define its degree matrix $\mathbf{D}$ which is diagonal with its diagonal entries $D_{ii} \equiv d_i$ as given by (1), and the graph Laplacian matrix

$$\mathbf{L} = \mathbf{D} - \mathbf{A}. \quad (3)$$

The graph Laplacian controls diffusion and random walks on graphs. Resistance distance Ω_{ij} between vertices i and j can be obtained as [8]

$$\Omega_{ij} = L_{ii}^{\text{inv}} + L_{jj}^{\text{inv}} - 2L_{ij}^{\text{inv}}, \quad (4)$$

where $\mathbf{L}^{\text{inv}}$ is the Moore-Penrose (pseudo)inverse of $\mathbf{L}$, defined so that $\mathbf{L}^{\text{inv}}\mathbf{L}$ is the projector on the subspace spanned by all non-null eigenvectors of $\mathbf{L}$. If the graph is connected, $\mathbf{L}$ has only one null eigenvector proportional to $(1, 1, \dots, 1)^T$ and hence $\mathbf{L}^{\text{inv}}\mathbf{L} = \mathbf{I} - \mathbf{1}/N$, where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is the matrix all of whose entries equal 1. We provide a derivation of (4) in Appendix A.

Our main object of study is the probability density for the resistance measured between vertices 1 and 2 to be equal ρ :

$$P(\Omega_{12} = \rho) \equiv \langle \delta(L_{11}^{\text{inv}} + L_{22}^{\text{inv}} - 2L_{12}^{\text{inv}} - \rho) \rangle = \frac{1}{Z} \sum_{\{\mathbf{A}\}} \delta(L_{11}^{\text{inv}} + L_{22}^{\text{inv}} - 2L_{12}^{\text{inv}} - \rho) \left(\frac{c}{N} \right)^{\sum_{i < j} A_{ij}}. \quad (5)$$

The corresponding distribution for any other pair of vertices would evidently be the same as the Erdős-Rényi ensemble enjoys complete vertex permutation symmetry. This function can be straightforwardly sampled numerically, and such numerical experiments suggest that the distribution tends to a definite curve at large N and displays rather sophisticated shapes. Our purpose is to develop some analytic theory for this distribution.

One can start by expressing the δ -function in (5) in terms of its Fourier representation:

$$P(\rho) \propto \int_{-\infty}^{\infty} d\xi \sum_{\{\mathbf{A}\}} e^{i\xi(L_{11}^{\text{inv}} + L_{22}^{\text{inv}} - 2L_{12}^{\text{inv}} - \rho)} \left(\frac{c}{N} \right)^{\sum_{i < j} A_{ij}}. \quad (6)$$

From this point on, we shall be ignoring the overall prefactor of $P(\rho)$ keeping in mind that it can always be recovered at the end by normalizing P as a probability density,

$$\int_0^{\infty} d\rho P(\rho) = 1. \quad (7)$$

Summation over $\mathbf{A}$ still cannot be performed directly in (6), but it will become possible after we introduce a set of auxiliary fields. We start with an N -dimensional real-valued

vector $\boldsymbol{\phi}$ whose components are ϕ_i and write, using the standard multidimensional Gaussian integration,

$$e^{i\xi(L_{11}^{\text{inv}}+L_{22}^{\text{inv}}-2L_{12}^{\text{inv}})} = \left(\frac{i\xi}{\pi}\right)^{\frac{N-1}{2}} \sqrt{\det'\mathbf{L}} \int d\boldsymbol{\phi} \delta(\sum_i \phi_i) e^{-i\xi \sum_{kl} L_{kl} \phi_k \phi_l + 2i\xi(\phi_1 - \phi_2)}. \quad (8)$$

Such Gaussian inversion formulas are frequently employed in random matrix literature relying on statistical physics methods [55, 60–62]; some brief pedagogical comments clarifying the structure of (8) are provided in Appendix B. We have inserted $\delta(\sum_i \phi_i)$ inside the integral in (8) to control the null direction of $\mathbf{L}$ corresponding to the eigenvector $(1, 1, \dots, 1)^T$, and $\det'\mathbf{L}$ denotes the pseudodeterminant of $\mathbf{L}$, that is the product of its nonzero eigenvalues. Then, using $\sum_{ij} L_{ij} \phi_i \phi_j = \frac{1}{2} \sum_{ij} A_{ij} (\phi_i - \phi_j)^2 = \sum_{i < j} A_{ij} (\phi_i - \phi_j)^2$, we arrive at

$$P(\rho) \propto \int d\xi \xi^{\frac{N-1}{2}} e^{-i\xi\rho} \int d\boldsymbol{\phi} \delta(\sum_i \phi_i) e^{2i\xi(\phi_1 - \phi_2)} \sum_{\{\mathbf{A}\}} \sqrt{\det'\mathbf{L}} e^{\sum_{k < l} A_{kl} [\ln(c/N) - i\xi(\phi_k - \phi_l)^2]}. \quad (9)$$

A few comments are in order before we proceed. First, (8) is only literally correct if the graph is connected and $(1, 1, \dots, 1)^T$ is the only null eigenvector of its Laplacian, which is, strictly speaking, not true in our case. Nonetheless, this inaccuracy will not impede our subsequent derivations in terms of $1/c$ expansions at $N \rightarrow \infty$, as we shall now explain. At $N \rightarrow \infty$, an Erdős-Rényi graph with a fixed $c > 1$ consists almost surely [59] of a single giant connected component whose size is of order N and further small components whose size is at most logarithmic in N . The probability u for a given vertex not to belong to the giant component satisfies [59] the equation

$$u = e^{-c(1-u)}. \quad (10)$$

At large c , this implies $u \approx e^{-c}$. Therefore, the number of vertices outside the giant connected component of the Erdős-Rényi graph, and certainly the number of disconnected components, and hence the number of distinct null eigenvectors of the Laplacian are all suppressed by e^{-c} . Thus, one may legitimately expect that their effect will be simply invisible within an $1/c$ expansion. Our derivations will indeed show no pathologies due to these neglected null eigenvectors, while the results will be in agreement with numerical simulations, which validates the intuition given above.

Second, the summation over $\mathbf{A}$ still cannot be performed directly in (9) because of the presence of $\sqrt{\det'\mathbf{L}}$. The determinant of the graph Laplacian is known to equal the number of spanning trees of the graph, and it has been studied for random graphs in [63, 64]. There are two possible approaches to handling that determinant, both of which we shall explore. At the most naive level, one may expect some concentration behavior for this determinant at $N \rightarrow \infty$ and large c , which would turn it into an irrelevant numerical factor that can be taken outside the sum. The sum over $\mathbf{A}$ can then be immediately evaluated, leaving behind an integral over ξ and $\boldsymbol{\phi}$ accessible to statistical field theory methods. At a more refined level, one may simply write an exact representation for $\sqrt{\det'\mathbf{L}}$ by introducing a few further Gaussian integrals over extra auxiliary fields, a technique often employed in the physics of disordered systems [65–67]. After that, the summation over $\mathbf{A}$ can again be performed, but now without any approximations or guesses for $\sqrt{\det'\mathbf{L}}$, and the result is still

tractable by statistical field theory methods. As an outcome of implementing the more refined approach, one in fact observes that the naive approach of discarding the determinant had been completely justified at leading order in $1/c$, and receives explicit computable corrections at subleading orders, where the determinant should not be neglected.

Whether working with the naive approach, as we shall do in section 3, or with the exact approach, as we shall do in section 4, evaluating the sum over $\mathbf{A}$ leaves behind an integral over variables defined on the graph vertices and interacting in a pair-wise manner. This integral may be viewed as a rather peculiar statistical field theory on a complete graph. Such field-theoretic representations have been successfully applied in the past to a variety of random graph problems, as in [54], see also [55–57]. The essence of these constructions is that the resulting field-theoretic integral admits a saddle-point evaluation, in our case, as an expansion in powers of $1/c$.

In view of our subsequent comparisons of the asymptotic evaluation of (5) with numerical simulations, we conclude this section with a remark on how such comparisons are made. One can easily generate an adjacency matrix of an Erdős-Rényi graph and then invert it to obtain the resistance distance matrix according to (4). It may be additionally wise to restrict the adjacency matrix to the giant component of the Erdős-Rényi graph, though it is not essential, at least at large c , since all the other components contain a small number of vertices and cannot significantly affect the statistics. Once the resistance distance matrix has been obtained from (4), it would be wasteful to only use its Ω_{12} entry for constructing the distribution of the resistance distance between vertices 1 and 2 and discard the rest. Indeed, resistance distances between different pairs of vertices must be identically distributed due to the permutation symmetry of the Erdős-Rényi ensemble, and one also expects them to be rather weakly correlated since the entries of the adjacency matrix are completely uncorrelated. It is then natural to sample the whole set of values of Ω_{ij} corresponding to all pairs of vertices, and plot histograms of this sample. It is such samples that we use for testing our analytics, and find convincing agreement, normally for $N = 25000$ (this provides a giant number of entries, given by $N(N - 1)/2$, for constructing the histograms).

3 Saddle point analysis

We shall now proceed with a basic saddle point analysis of (9), assuming that $\sqrt{\det' \mathbf{L}}$ concentrates in an appropriate sense and can be treated as $\mathbf{A}$ -independent within the Erdős-Rényi ensemble. An accurate justification will be given to this assumption in the next section. The saddle point analysis we present gives a concrete mathematical realization to the ‘mean field’ principles common in statistical physics, and the idea that a large number of nearest neighbors given by c simplifies the behavior of the system due to ‘averaging over neighbors’ is very much in line with these principles.

3.1 Summation over $\mathbf{A}$

Assuming that $\sqrt{\det' \mathbf{L}}$ in (9) is approximately constant, it can be taken outside the $\mathbf{A}$ -sum and merged with the other normalization factors. Thereafter, the sum over $\mathbf{A}$ is immediately evaluated in a manner that closely parallels [54], since the summand turns into a product of

factors each of which depends only on a single entry of the matrix $\mathbf{A}$:

$$\sum_{\{\mathbf{A}\}} \prod_{k < l} e^{A_{kl} [\ln(c/N) - i\xi(\phi_k - \phi_l)^2]} = \prod_{k < l} \left(1 + \frac{c}{N} e^{-i\xi(\phi_k - \phi_l)^2}\right) = \exp \left[\sum_{k < l} \ln \left(1 + \frac{c}{N} e^{-i\xi(\phi_k - \phi_l)^2}\right) \right]. \quad (11)$$

This formula follows from the tautological summation identity $\sum_{\alpha=0}^1 e^{\alpha\beta} = 1 + e^\beta$.

At $N \rightarrow \infty$ and a fixed c , we can furthermore approximate the logarithm as

$$\ln \left(1 + \frac{c}{N} e^{-i\xi(\phi_k - \phi_l)^2}\right) \approx \frac{c}{N} e^{-i\xi(\phi_k - \phi_l)^2}. \quad (12)$$

Hence, (9) turns into

$$P(\rho) \propto \int d\xi \xi^{\frac{N-1}{2}} e^{-i\xi\rho} \int d\phi \delta(\sum_i \phi_i) e^H, \quad (13)$$

where H is defined as

$$H \equiv \frac{c}{N} \sum_{k < l} e^{-i\xi(\phi_k - \phi_l)^2} + 2i\xi(\phi_1 - \phi_2) = \frac{c}{2N} \sum_{k \neq l} e^{-i\xi(\phi_k - \phi_l)^2} + 2i\xi(\phi_1 - \phi_2). \quad (14)$$

One aspect of (13) that will play a crucial role hereafter is that H appears in the exponent and features a term proportional to c . Hence, assuming that c is large makes (13) amenable to conventional saddle point techniques, resulting in a viable approach to evaluating $P(\rho)$ in terms of an $1/c$ expansion. As we shall see below, the leading term of this expansion does a convincing job at capturing the behavior of $P(\rho)$ even at values of c that are not nominally ‘large,’ such as $c = 4$.

3.2 The saddle point equation

Given the apparent saddle point structure of (13) at large c , we must identify the stationary points of H , given by solutions of²

$$0 = \frac{\partial H}{\partial \phi_k} = -\frac{2ic\xi}{N} \sum_{l \neq k} \left[e^{-i\xi(\phi_k - \phi_l)^2} (\phi_k - \phi_l) \right] + 2i\xi(\delta_{1,k} - \delta_{2,k}). \quad (15)$$

²It may appear odd at the first sight that we are including the ξ -term in the saddle point equation, while it comes with no explicit dependence on the saddle point parameter c . It is important to keep in mind, however, that ξ is an integration variable and all possible values enter the game. The relevant scaling of the resistance distance is $\rho \sim 1/c$ (one can immediately see this numerically, and it will also come out of our analysis) and the corresponding scaling of the Fourier-conjugate variable ξ is $\sim c$, making the ξ -term in (14) relevant for the saddle point equation. The test of this approach to generating the saddle point expansion is that a well-formed asymptotic series in $1/c$ will be produced. Namely, if (14) is expanded in Taylor series around the saddle point defined as we have specified, the leading contribution will arise from the Gaussian part of (13), and higher order terms in (14) will generate $1/c$ corrections, as evident from our treatment below and considerations of Appendix C. By contrast, if we had omitted the ξ -term from the saddle point equation, the saddle point solution would have been $\phi = 0$, resulting in a trivial estimate of the resistance distance distribution $P(\rho) = \delta(\rho)$ that conveys no meaningful message beyond the smallness of ρ at large c and cannot be straightforwardly corrected in terms of a perturbative expansion in $1/c$.

Once ϕ_0 that satisfies this equation has been found, it can be used to obtain a saddle point approximation of (13) by applying

$$\int d\phi \delta(\sum_i \phi_i) e^{H(\phi)} \approx e^{H(\phi_0)} \int d\tilde{\phi} \delta\left(\sum_i \tilde{\phi}_i\right) e^{-\frac{1}{2} \sum_{ij} (-M_{ij}) \tilde{\phi}_i \tilde{\phi}_j} = \frac{(-2\pi)^{\frac{N-1}{2}}}{\sqrt{\det' \mathbf{M}}} e^{H(\phi_0)}, \quad (16)$$

where $\tilde{\phi} \equiv \phi - \phi_0$, $\det'$ once again denotes the pseudodeterminant, and we have used $\sum_i \phi_{0i} = 0$, as all configurations integrated over in (13) satisfy this relation. The Hessian $\mathbf{M}$ is given by

$$M_{ij} = \frac{\partial^2 H}{\partial \phi_i \partial \phi_j} \Big|_{\phi=\phi_0}. \quad (17)$$

Note that the δ -function in (16) simply controls the null direction $(1, 1, \dots, 1)^T$ of the Hessian $\mathbf{M}$ (some further details on Gaussian integrals with null directions can be found in Appendix B).

We must then identify the solutions of (15). It is natural to start by examining the saddle points that respect the symmetries of H . (An alternative is families of saddle points converted into each other by the symmetries. It is more exotic, though certainly not impossible, and the ultimate validation of our approach, here and in general, comes from comparisons with numerical simulations.) As one can see from (14), H is invariant under arbitrary permutations of $\phi_3, \phi_4, \dots, \phi_N$ as well as the symmetry $(\phi_1, \phi_2) \rightarrow (-\phi_2, -\phi_1)$. Since only configurations with $\sum_i \phi_i = 0$ contribute to (13), the only saddle candidate that respects the symmetries of H is $\phi_0 = (\phi_0, -\phi_0, 0, 0, \dots)$ with a yet-unknown ϕ_0 . For configurations of this form, (15) is rewritten as

$$0 = \begin{cases} -\frac{2i\xi}{N} c e^{-4i\xi\phi_0^2} (2\phi_0) - 2i\xi \frac{N-2}{N} c e^{-i\xi\phi_0^2} \phi_0 + 2i\xi & \text{for } k = 1, \\ \frac{2i\xi}{N} c e^{-4i\xi\phi_0^2} (2\phi_0) + 2i\xi \frac{N-2}{N} c e^{-i\xi\phi_0^2} \phi_0 - 2i\xi & \text{for } k = 2, \\ \frac{2i\xi}{N} \left(c e^{-i\xi\phi_0^2} \phi_0 + c e^{-i\xi\phi_0^2} (-\phi_0) \right) & \text{for } k \geq 3. \end{cases} \quad (18)$$

The last equation is identically satisfied, while the equations for $k = 1$ and $k = 2$ are equivalent to each other, and reduce at $N \rightarrow \infty$ to

$$c\phi_0 = e^{i\xi\phi_0^2}. \quad (19)$$

The saddle point equation (19) is closely related to the equation $W(x)e^{W(x)} = x$ defining the Lambert W function [68]. The solution can correspondingly be expressed through the Lambert W function as

$$\phi_0(\xi) = \sqrt{\frac{W(-2i\xi/c^2)}{-2i\xi}}. \quad (20)$$

In practice, it may often be convenient to construct ϕ_0 by solving the ODE

$$\frac{d\phi_0}{d\xi} = -\frac{\phi_0^3}{i + 2\xi\phi_0^2} \quad (21)$$

that can be readily derived from (19), with the initial condition $\phi_0(0) = 1/c$. We quote the series expansion for ϕ_0 to give an impression of its behavior near the origin:

$$\phi_0(\xi) = \frac{1}{c} + \frac{i\xi}{c^3} - \frac{5\xi^2}{2c^5} - \frac{49i\xi^3}{6c^7} + \frac{243\xi^4}{8c^9} + \dots \quad (22)$$

Evidently $c\phi_0$ only depends on ξ/c^2 .

Finally, the value of H at the saddle point is evaluated as

$$H(\phi_0) = \frac{Nc}{2} + \frac{2}{\phi_0} + 4i\xi\phi_0 \rightarrow \frac{2}{\phi_0} + 4i\xi\phi_0, \quad (23)$$

where the saddle point equation (19) has been used to simplify this expression, and the last ‘arrow’ operation indicates that the ξ -independent term $Nc/2$ can be merged into the normalization of $P(\rho)$ and hence ignored. What remains for completing the saddle point estimate (16) is to evaluate the Hessian determinant $\det' \mathbf{M}$.

3.3 The Hessian determinant

The second derivatives of H can be expressed as

$$\frac{\partial^2 H}{\partial \phi_l \partial \phi_k} = \begin{cases} \frac{2i\xi c}{N} \sum_{i \neq k} e^{-i\xi(\phi_k - \phi_i)^2} [2i\xi(\phi_k - \phi_i)^2 - 1] & \text{for } l = k, \\ -\frac{2i\xi c}{N} e^{-i\xi(\phi_k - \phi_l)^2} [2i\xi(\phi_k - \phi_l)^2 - 1] & \text{for } l \neq k. \end{cases} \quad (24)$$

At the saddle point (19), this becomes

$$M_{lk} = \left. \frac{\partial^2 H}{\partial \phi_l \partial \phi_k} \right|_{\phi_0} = \begin{cases} -\frac{2\xi}{N} \left[\frac{1}{\phi_0^4 c^3} (i + 8\xi\phi_0^2) + (N-3) \frac{1}{\phi_0} (i + 2\xi\phi_0^2) \right] \equiv a & \text{for } l = k \leq 2, \\ -\frac{2\xi}{N} \left[\frac{2}{\phi_0} (i + 2\xi\phi_0^2) + (N-3)c \right] \equiv b & \text{for } l = k > 2, \\ \frac{2}{N\xi^2\phi_0^4} (i + 8\xi\phi_0^2) \equiv g & \text{for } l = 1, k = 2, \\ \frac{2\xi}{N\phi_0} (i + 2\xi\phi_0^2) \equiv f & \text{for } l = 1, 2; k > 2, \\ \frac{2ic\xi}{N} \equiv d & \text{for } l \neq k, l, k > 2, \end{cases} \quad (25)$$

where we have specified the values for $l \leq k$, and the remaining values can be recovered using the symmetry of M . The structure of this matrix can be visualized as follows:

$$\mathbf{M} = \begin{pmatrix} a & g & f & f & f & \dots \\ g & a & f & f & f & \dots \\ f & f & b & d & d & \dots \\ f & f & d & b & d & \dots \\ f & f & d & d & b & \vdots \\ \vdots & \vdots & \vdots & \vdots & \dots & \ddots \end{pmatrix} \quad (26)$$

where a, b, d, f and g are defined in (25). Note that each row (and column) sums to 0 since H is not affected by a simultaneous common shift of all ϕ_k . In other words,

$$a + g + (N - 2)f = 0, \quad 2f + b + (N - 3)d = 0. \quad (27)$$

The eigenvectors of $\mathbf{M}$ can be worked out explicitly. The vector $(1, 1, \dots)^T$ is annihilated by $\mathbf{M}$, as follows from (27). This null eigenvector and its zero eigenvalue will evidently not contribute to the computation of the pseudodeterminant $\det' \mathbf{M}$. The second obvious eigenvector is $(1, -1, 0, 0, \dots)^T$ and its eigenvalue is $a - g$. Finally, there are $N - 3$ eigenvectors of the form $(0, \dots, 0, 1, -1, 0, 0, \dots)^T$ where the first “1” must occur in position 3 or higher. Their eigenvalues are $b - d$. We have explicitly recovered $N - 1$ eigenvectors, and since $\mathbf{M}$ is a Hermitian $N \times N$ matrix, the unique remaining eigenvector must be orthogonal to all of them and hence proportional to $(-1, -1, \frac{2}{N-2}, \frac{2}{N-2}, \frac{2}{N-2}, \dots)^T$. Its eigenvalue can be computed by direct application of $\mathbf{M}$ and then simplified using (27) to give $-Nf$. With all the eigenvalues at hand, we can write down the desired pseudodeterminant explicitly as

$$\det' \mathbf{M} = -Nf(a - g)(b - d)^{N-3}. \quad (28)$$

Now, we can substitute (25) back in (28) and take the $N \rightarrow \infty$ limit. This requires some care in the last factor since $(1 + x/N)^N \rightarrow e^x$; elsewhere, we can simply discard all subleading contributions in $1/N$. Altogether,

$$\det' \mathbf{M} \propto \xi^{N-1} \left(\frac{1}{\phi_0} - 2i\xi\phi_0 \right)^2 e^{\frac{2}{c\phi_0} - \frac{4i\xi\phi_0}{c}}, \quad (29)$$

where we have omitted all the irrelevant ξ -independent factors.

3.4 The leading saddle point estimate

Putting together (13), (16), (23) and (29), we obtain the leading order saddle point evaluation of (5) in the form

$$P(\rho) \propto \int d\xi \left(\frac{1}{\phi_0} - 2i\xi\phi_0 \right)^{-1} e^{-i\xi\rho} e^{\frac{2}{\phi_0}(1 - \frac{1}{2c}) + 4i\xi\phi_0(1 + \frac{1}{2c})}, \quad (30)$$

with ϕ_0 defined by (19) and conveniently reconstructed by solving (21). Normalization (7) should be applied. This expression for $P(\rho)$ is the main result of this paper. We shall now recast it in a few alternative forms, and provide further simplified formulas valid at large c .

A useful observation is that the pre-exponential factor in (30) can be expressed in terms of $d\phi_0/d\xi$ using (21), which gives

$$P(\rho) \propto \int_{-\infty}^{\infty} d\xi \frac{d\phi_0}{d\xi} \frac{1}{\phi_0^2} e^{\frac{2}{\phi_0}(1 - \frac{1}{2c}) + 4i\xi\phi_0(1 + \frac{1}{2c})}. \quad (31)$$

As a result, one can choose ϕ_0 to be the new integration variable, which is convenient since, while $\phi_0(\xi)$ is expressed through the Lambert W function, its inverse $\xi(\phi_0) = -i \ln(c\phi_0)/\phi_0^2$

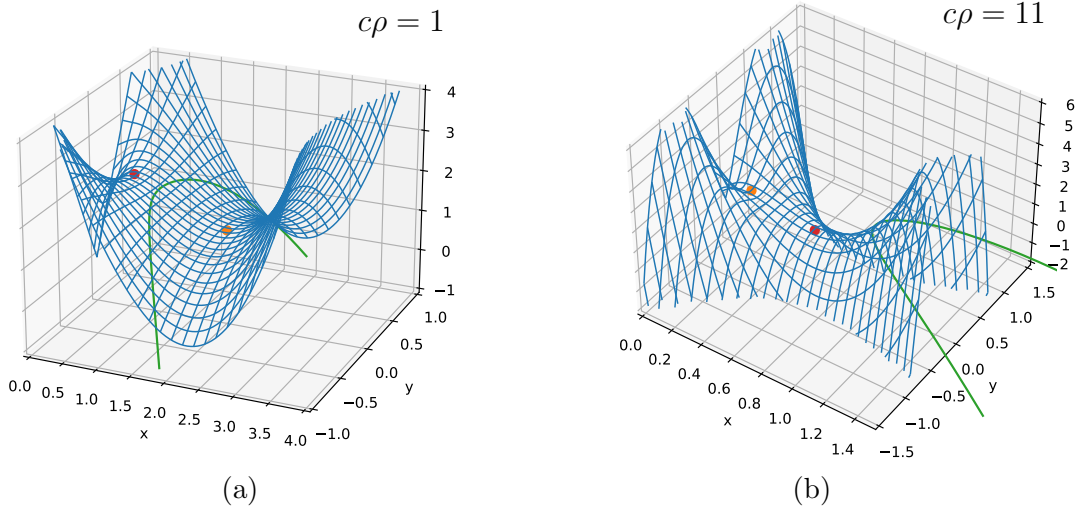


Figure 1: The real part of the saddle point function $2z + (\rho cz^2 - 4z) \ln z$ of (32), evaluated at $z = x + iy$. The orange dots represent the saddle at $z = 2/c\rho$, and the red dots represent the saddle at $z = 1/\sqrt{e}$. The green curves visualize the integration contour in (32). The two plots demonstrate how the two saddle point locations change as $c\rho$ varies from $c\rho = 1$ (a) to $c\rho = 11$ (b).

is an elementary function. A particularly convenient choice for the integration variable is $z = 1/(c\phi_0)$, which gives

$$P(\rho) \propto \int_{1/c\phi_0(-\infty)}^{1/c\phi_0(\infty)} dz e^{2cz(1-\frac{1}{2c})} z^{\rho c^2 z^2 - (4c+2)z}. \quad (32)$$

The integration contour, as inherited from (31), comes from infinity within the first quadrant ($\text{Re } z > 0, \text{Im } z > 0$), passes through $z = 1$ and exists to infinity via the fourth quadrant ($\text{Re } z > 0, \text{Im } z < 0$). The contour can evidently be deformed freely, as with any complex plane integral, as long as it does not touch the singular branching point at $z = 0$.

As the integrand features c in the exponent, and we are treating c as a large parameter, it is natural to apply further saddle point evaluation to (32). To this end, we note that the logarithm of the integrand is $-2z \ln z - z + c[2z + (\rho cz^2 - 4z) \ln z]$. Demanding that the leading part of this expression (given in the square brackets) is stationary at z_0 yields the saddle point equation for the z -integral:

$$(c\rho z_0 - 2)(1 + 2 \ln z_0) = 0. \quad (33)$$

This equation evidently has two solutions, $z_0 = 2/c\rho$ and $z_0 = 1/\sqrt{e}$. The topography of the integrand is such that the contour can always be deformed to pass through $\max(2/c\rho, 1/\sqrt{e})$, while the second saddle lies between the contour and the singularity at $z = 0$ and is thus irrelevant for the saddle point evaluation. The location of the contour relative to the saddles is depicted in Fig. 1. Extracting the leading saddle point estimate due to the appropriate saddles thus yields

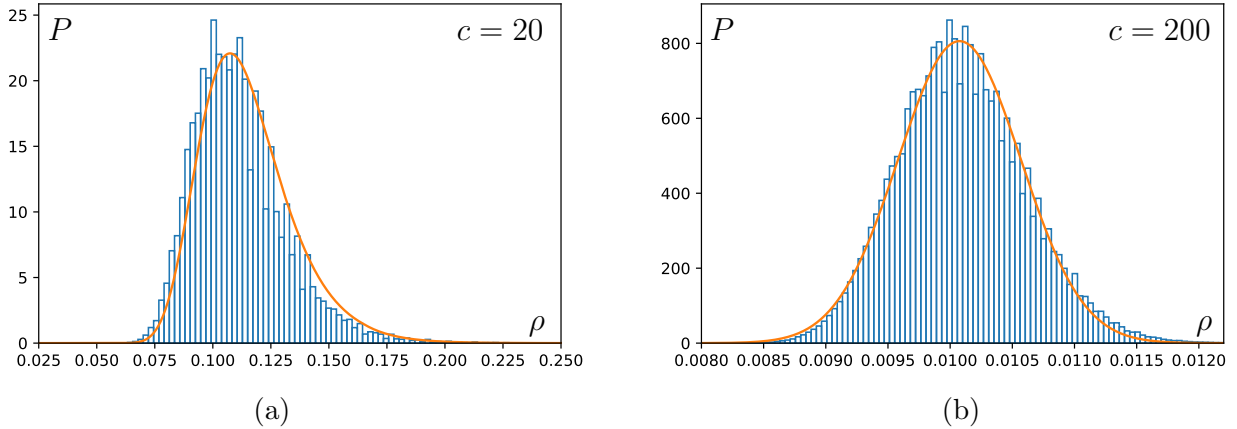


Figure 2: Resistance distance distribution of the largest connected component of an Erdős-Rényi graph with 25000 vertices at mean degree 20 (a) and 200 (b). Numerical simulation results are plotted as blue histograms, while the saddle point estimate (34) for (a) and the large c Gaussian approximation (35) for (b) are plotted as orange curves.

$$P(\rho) \propto \begin{cases} \left(\frac{c\rho}{2}\right)^{\frac{4}{c\rho}} e^{-\frac{2}{c\rho}} \left(\frac{ec\rho}{2}\right)^{\frac{4}{\rho}} & \text{for } \rho \leq 2\sqrt{e}/c, \\ \exp\left(\frac{4c}{\sqrt{e}} - \frac{\rho c^2}{2e}\right) & \text{for } \rho > 2\sqrt{e}/c. \end{cases} \quad (34)$$

We have not attempted recovering the pre-exponential factors in this saddle point estimate, which is slightly subtle due to the coalescing saddles [69] at $\rho = 2\sqrt{e}/c$, and would not have improved the accuracy significantly in the regime of interest. We compare this curve to the actual numerical simulations at the moderate value $c = 20$ in Fig. 2a, showing a good agreement.

Finally, when c is very large, it makes sense to expand (34) around its maximum at $\rho \approx 2/c$ which yields a Gaussian approximation of the form

$$P(\rho) \propto \exp\left[-\frac{c^3}{4(1-4/c)}\left(\rho - \frac{2}{c} - \frac{3}{c^2}\right)^2\right]. \quad (35)$$

This distribution is again in an excellent agreement with the numerics at large c , see Fig. 2b. Asymptotically, the distribution tends to a Gaussian of mean $2/c$ and standard deviation $\sqrt{2/c^3}$.

The agreement between our saddle point estimate and numerical simulations continues to hold even for such low values of c as 6 or 4. In this case, (30) gives a slightly better match than (34), and that is what we plot in Fig. 3. To quantify the agreement of the analytic and empirical distributions, we use the *overlapping coefficient* [70], which is defined, for any two normalized probability distributions $p_1(x)$ and $p_2(x)$, as $\int \min(p_1(x), p_2(x)) dx$. By definition, the overlapping coefficient is a number between 0 and 1 that equals 1 if and only if the two distributions are identical. For our analytic curve as compared to the numerics, this gives approximately 92% for $c = 6$ and approximately 90% for $c = 4$. We emphasize that the saddle point estimate (30) only captures the main peak of the distribution without

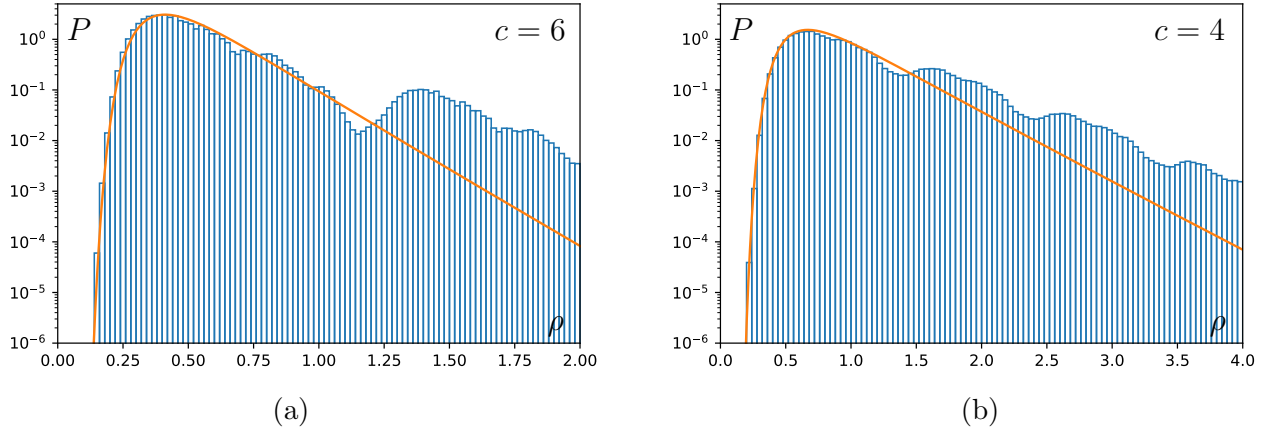


Figure 3: Resistance distance distribution of the largest connected component of an Erdős-Rényi graph with 25000 vertices at mean degree 6 (a) and 4 (b). Numerical simulation results are plotted as blue histograms, while the saddle point estimate (30) is plotted as orange curves. We have used the log-linear scale to emphasize that, while the main peak is captured adequately, there is excess probability in the right tail compared to our saddle point estimate. We shall revisit the contributions at larger values of ρ in section 5.

reproducing accurately its right tail. We shall return to the excess probability observed to the right of the main peak in section 5.

4 Fluctuations of $\sqrt{\det' L}$ in (9)

This section can be comfortably skipped by practically-minded readers that are interested in obtaining useful approximations to the resistance distance distribution, rather than in justifying the corresponding derivations. Our purpose here is to close an essential gap in the reasoning of the previous section, though after this has been accomplished, the result of the previous section will remain unchanged.

We would like to return to (9) and provide an accurate treatment of the $\sqrt{\det' L}$ factor that has been neglected in the previous section. This is done by introducing a few more auxiliary fields and produces a more elaborate structure similar to (13-14). As a result of these considerations, we shall see that (a) the analysis of the previous section is justified at leading order in $1/c$, (b) the determinant does contribute to subleading corrections in $1/c$ (though such contributions are not crucial for the practical purposes of this paper). More precisely, when the determinant contribution and other higher order corrections are taken into account, after an appropriate ‘linked cluster’ resummation (the technical details are given in Appendix C), the exponent of the last factor in (30) turns into a series in powers of $1/c$. However, the powers of $1/c$ already written in (30) remain unchanged by these contributions, which only supply higher (and hence more strongly suppressed) powers of $1/c$.

To proceed with our derivations, we shall represent the determinant factor $\sqrt{\det' L}$ in terms of Gaussian integrals involving both ordinary (commuting) and Grassmanian (anti-

commuting) variables. This approach has often been taken up in statistical physics considerations related to disordered systems [65–67]. To implement this approach, we rewrite

$$\sqrt{\det' L} = \frac{1}{\sqrt{\det' L}} \det' L. \quad (36)$$

The first factor can be represented as an ordinary Gaussian integral

$$\frac{1}{\sqrt{\det' L}} = \pi^{-\frac{N-1}{2}} \int d\chi \delta(\sum_i \chi_i) e^{-\sum_{ij} L_{ij} \chi_i \chi_j}. \quad (37)$$

For the second factor, a convenient representation is given in terms of anticommuting (Grassmanian) fields [65] denoted θ and η that satisfy

$$\theta_i \theta_j = -\theta_j \theta_i \quad (\text{so that } \theta_i^2 = 0), \quad \int d\theta_i \theta_j = \delta_{ij}, \quad \int d\theta_i = 0, \quad (38)$$

and similar relations for the components of η , while the components of θ and η anticommute with each other. With the *Berezin integration rules* given above (a more detailed explanation can be found in [65]), we express the determinant $\det' L$ in terms of these new fields as

$$\det' L = \int d\theta d\eta \delta(\sum_i \theta_i) \delta(\sum_i \eta_i) e^{-\sum_{ij} L_{ij} \theta_i \eta_j}. \quad (39)$$

Observing that $\sum_{ij} L_{ij} \chi_i \chi_j = \sum_{i<j} A_{ij} (\chi_i - \chi_j)^2$ and $L_{ij} \theta_i \eta_j = \sum_{i<j} A_{ij} (\theta_i - \theta_j)(\eta_i - \eta_j)$, and inserting (36) represented using (37) and (39) into (9), one can perform the summation over $\mathbf{A}$ in a manner identical to (11). We then arrive at the following representation of (9), with the determinant factor now fully taken into account:

$$P(\rho) \propto \int d\xi \xi^{(N-1)/2} e^{-i\xi\rho} \int d\phi d\chi d\theta d\eta \delta(\sum_i \phi_i) \delta(\sum_i \chi_i) \delta(\sum_i \theta_i) \delta(\sum_i \eta_i) e^{H'}. \quad (40)$$

Here, H' is defined by

$$H' = \frac{c}{N} \sum_{i<j} \exp \left[-i\xi(\phi_i - \phi_j)^2 - (\chi_i - \chi_j)^2 - (\theta_i - \theta_j)(\eta_i - \eta_j) \right] + 2i\xi(\phi_1 - \phi_2). \quad (41)$$

We have applied the $N \rightarrow \infty$ limit, as in (12), to simplify the above expression.

The new representation (40-41) can be treated by saddle point methods in direct parallel to our previous treatment of (13-14). The saddle point configuration for ϕ is exactly the same as in section 3.2 and is still defined by the saddle point equation (19). All the remaining fields in (41) vanish in the saddle point configuration. As a result, the different fields involved completely decouple from each other in the expansion of (41) up to quadratic order in the fields about the saddle point:

$$H' = H(\phi_0) + \frac{1}{2} \sum_{ij} M_{ij} \tilde{\phi}_i \tilde{\phi}_j - \frac{c}{N} \sum_{i<j} [(\chi_i - \chi_j)^2 + (\theta_i - \theta_j)(\eta_i - \eta_j)], \quad (42)$$

where the first two terms are defined by (23) and (25). Since the quadratic forms of χ , θ and η in the above expression decouple from the quadratic form of ϕ , and are furthermore

ξ -independent, the leading order saddle point estimate remains completely identical to section 3. Thus, our neglect of the $\sqrt{\det' L}$ factor in the derivations of section 3 is *a posteriori* justified.

At the same time, the different fields involved no longer decouple if one expands (41) beyond the quadratic order in the fields, and thus extra corrections at higher orders in $1/c$ will be produced by the additional fields in (41) that capture the fluctuations of $\sqrt{\det' L}$. We provide an analysis of these corrections in Appendix C, though they do not play a significant practical role in the considerations of this paper.

5 Subleading corrections and subleading peaks

One can systematically compute the higher order corrections to (30) within the $1/c$ expansion using the language of Feynman diagrams. We give a technical demonstration of such computations at first subleading order in Appendix C. In practice, however, these corrections are numerically insignificant at large c , while at small c , as in Fig. 3, they do not improve the agreement with the numerics. In fact, they slightly shift the peak to the right away from the excellent match in Fig. 3, without inducing any other significant changes. A skeptic might then say that the agreement we observe when the leading order estimate is extended to small values of c is coincidental. We, however, feel that it must be systematic as it is observed for different small values of c and there must be an underlying mathematical reason for it, beyond the scope of our current comprehension. Be it as it may, at this stage, we simply report the agreement of the leading order saddle point estimate with the numerical data at low c as an empirical fact.

Figs. 2 and 3 show that the leading order saddle point estimate does a great job at reproducing the main outline of the resistance distance probability distribution $P(\rho)$. Closer examination, however, reveals more detailed features of the empirical probability distributions, in particular, at low values of c . To demonstrate this aspect, we have given in Fig. 4a a more detailed version of the empirical resistance distance histogram at $c = 6$ that we have previously displayed in a more coarse-grained presentation in Fig. 3a, this time switching to a linear rather than log-linear plot. (We ask the reader to ignore the colors in Fig. 4a at this point.) First, there are sharp narrow sub-peaks that decorate the main peak and, in fact, may suggest a lack of smoothness in the true distribution. We do not have a theoretical understanding of these sharp peaks and will briefly comment on them further in the conclusions. Second, there are ‘bumps’ that are seen on top of the tail of the distribution that extends to higher values of ρ . In the rest of this section, we aim to improve our understanding of these ‘bumps.’

To demonstrate the nature of the subleading peaks at large ρ to the right of the main peak, we have made a further refinement in Fig. 4a. Namely, when plotting the empirical histogram of resistance distances Ω_{ij} between vertices i and j in a computer-generated Erdős-Rényi graph, we have colored the contributions according to the minimal degree of the two vertices involved, $\min(d_i, d_j)$. This coloring reveals that the rightmost visible peak around $\rho \in (1.25, 1.75)$ in fact entirely consists of resistance distances between vertices of degree 1 and the rest of the graph. For other peaks at lower values of ρ , the picture is more complicated, but it is still clear that these peaks are made from contributions involving vertices of low degrees. (Qualitatively similar peaks due to vertices of specific low degrees

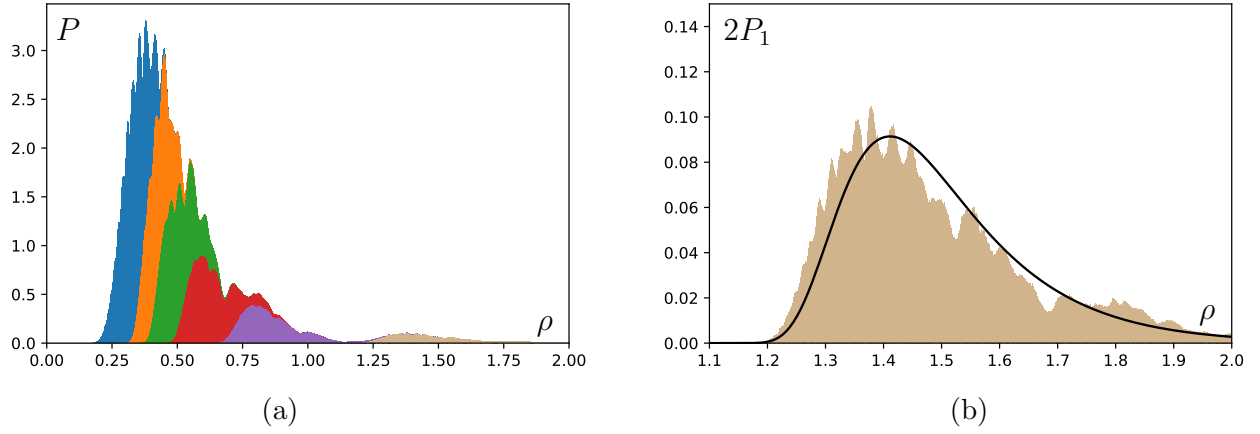


Figure 4: Refined representations for the resistance distance distribution of the largest connected component of an Erdős-Rényi graph with 25000 vertices at mean degree 6: (a) the contributions of the different pairs of vertices to the resistance distance histogram obtained from numerical simulations, colored according to the smaller value of the vertex degrees within each pair; the colors, appearing in sequence from right to left, correspond to the smaller degree values of 1, 2, 3, 4, 5, and all the remaining values lumped together; (b) the (rightmost) contribution to the histogram in (a) involving vertices of degree 1, isolated and compared to the analytic prediction $2P_1(\rho)$ extracted as described under (53) and plotted as a black curve.

have been seen in the analysis of a different characteristic of random graphs in [71].)

The above empirical observations suggest that it is natural to approach the analysis of the subleading peaks by introducing degree-differentiated resistance distance distributions that not only specify the resistance distance between vertices 1 and 2, but also require that the degree of vertex 1 equals d , namely

$$P_d(\Omega_{12} = \rho) = \frac{1}{Z} \sum_{\{\mathbf{A}\}} \delta(L_{11}^{\text{inv}} + L_{22}^{\text{inv}} - 2L_{12}^{\text{inv}} - \rho) \delta\left(d, \sum_j A_{1j}\right) \left(\frac{c}{N}\right)^{\sum_{i<j} A_{ij}}. \quad (43)$$

Evidently,

$$\sum_{d=1}^{\infty} P_d(\rho) = P(\rho), \quad (44)$$

thus one can think of P_d as controlling the different contributions to $P(\rho)$ in a more discerning manner. By writing a Fourier representation for the δ -function involving ρ , as in (6), and a complex plane representation for the Kronecker symbol involving d ,

$$\delta\left(d, \sum_j A_{1j}\right) = \frac{1}{2\pi i} \oint \frac{dz}{z^{d+1}} \prod_j z^{A_{1j}}, \quad (45)$$

one obtains

$$P_d(\rho) \propto \oint \frac{dz}{z^{d+1}} \int d\xi \sum_{\{\mathbf{A}\}} e^{i\xi(L_{11}^{\text{inv}} + L_{22}^{\text{inv}} - 2L_{12}^{\text{inv}} - \rho)} \prod_j z^{A_{1j}} \prod_{i<j} \left(\frac{c}{N}\right)^{A_{ij}}. \quad (46)$$

Then, as in section 2, we can write

$$P_d(\rho) \propto \oint \frac{dz}{z^{d+1}} \int d\xi e^{-i\xi\rho} \xi^{\frac{N-1}{2}} \int d\boldsymbol{\phi} \delta(\sum_i \phi_i) e^{2i\xi(\phi_1-\phi_2)} \quad (47)$$

$$\times \sum_{\{\mathbf{A}\}} \sqrt{\det' \mathbf{L}} e^{-i\xi \sum_{k<l} A_{kl}(\phi_k-\phi_l)^2} \prod_j z^{A_{1j}} \prod_{i<j} \left(\frac{c}{N}\right)^{A_{ij}}.$$

We could have given an accurate treatment of the $\sqrt{\det' \mathbf{L}}$ factor, as in section 4, by introducing extra auxiliary fields, but just like in section 4, the result (at leading order in $1/c$) would be exactly the same as what one gets by treating $\sqrt{\det' \mathbf{L}}$ as constant. We shall therefore simply omit this factor to keep the formulas more compact, without affecting the result. Then, performing the summation over $\mathbf{A}$ as in section 3.1, we get

$$P_d(\rho) \propto \oint \frac{dz}{z^{d+1}} \int d\xi e^{-i\xi\rho} \xi^{\frac{N-1}{2}} \int d\boldsymbol{\phi} \delta(\sum_i \phi_i) e^{2i\xi(\phi_1-\phi_2)} \quad (48)$$

$$\times \prod_{j \geq 2} \left(1 + \frac{cz}{N} e^{-i\xi(\phi_1-\phi_j)^2}\right) \prod_{2 \leq k < l} \left(1 + \frac{c}{N} e^{-i\xi(\phi_k-\phi_l)^2}\right).$$

We now focus on $d = 1$. The contour integral featuring dz/z^2 simply extracts the first z -derivative of the integrand at $z = 0$. This produces a factor of $\frac{c}{N} \sum_{j \geq 2} e^{-i\xi(\phi_1-\phi_j)^2}$. The contribution from $j = 2$ can be ignored as it is suppressed by $1/N$ and vanishes at $N \rightarrow \infty$. The remaining $N-2$ contributions from $j \geq 3$ will all equal each other upon integration since all the other factors in the integrand are invariant under all permutations of $(\phi_3, \phi_4, \dots, \phi_N)$. As a result, $\frac{c}{N} \sum_{j \geq 2} e^{-i\xi(\phi_1-\phi_j)^2}$ can be effectively replaced by $c e^{-i\xi(\phi_1-\phi_3)^2}$ at $N \rightarrow \infty$, leaving

$$P_1(\rho) \propto \int d\xi e^{-i\xi\rho} \xi^{\frac{N-1}{2}} \int d\boldsymbol{\phi} \delta(\sum_i \phi_i) e^{2i\xi(\phi_1-\phi_2)-i\xi(\phi_1-\phi_3)^2} \prod_{2 \leq k < l} \left(1 + \frac{c}{N} e^{-i\xi(\phi_k-\phi_l)^2}\right). \quad (49)$$

We now introduce an $(N-1)$ -dimensional vector $\boldsymbol{\phi}'$ whose components are $\phi'_j = \phi_{j+1} + \phi_1/(N-1)$ so that $\sum_{j=1}^N \phi_j = \sum_{j=1}^{N-1} \phi'_j$. Then,

$$P_1(\rho) \propto \int d\xi e^{-i\xi\rho} \xi^{\frac{N-1}{2}} \int d\boldsymbol{\phi}' \delta(\sum_i \phi'_i) \exp \left[\sum_{k < l} \ln \left(1 + \frac{c}{N} e^{-i\xi(\phi'_k-\phi'_l)^2}\right) \right]$$

$$\times \int d\phi_1 e^{-2i\xi[\phi'_1 - N\phi_1/(N-1)]} e^{-i\xi[\phi'_2 - N\phi_1/(N-1)]^2}. \quad (50)$$

The integral over ϕ_1 is Gaussian and can be straightforwardly evaluated to yield a result proportional to $e^{i\xi} e^{-2i\xi(\phi'_1-\phi'_2)}/\sqrt{\xi}$, so that

$$P_1(\rho) \propto \int d\xi e^{-i\xi(\rho-1)} \xi^{\frac{N-2}{2}} \int d\boldsymbol{\phi}' \delta(\sum_i \phi'_i) \exp \left[\sum_{k < l} \ln \left(1 + \frac{c}{N} e^{-i\xi(\phi'_k-\phi'_l)^2}\right) - 2i\xi(\phi'_1 - \phi'_2) \right]. \quad (51)$$

But after interchanging ϕ'_1 and ϕ'_2 one notices that, with $c \ll N$, this is the same expression as (13-14), except that ρ got replaced by $\rho - 1$ and N got replaced by $N - 1$. Hence, in the $N \rightarrow \infty$ limit,

$$P_1(\rho) \propto P(\rho - 1). \quad (52)$$

To identify the numerical coefficient in this relation, we notice that while $P(\rho)$ is normalized to 1 by (7), the integral of $P_1(\rho)$ over ρ is simply the probability for vertex 1 to have degree 1, as one can see by directly integrating the definition (43). The degree distribution of sparse Erdős-Rényi graphs is known to be Poissonian [59] and hence the probability for vertex 1 to have degree 1 is ce^{-c} . Hence,

$$P_1(\rho) = ce^{-c}P(\rho - 1). \quad (53)$$

In Fig. 4b, we have plotted (53) against the distribution of resistance distances involving vertices of degree 1 from a numerical simulation of an Erdős-Rényi graph at $c = 6$. The saddle point estimate (30), shifted to the right by one unit has been used to approximate $P(\rho - 1)$ on the right-hand side of (53). To get the normalization right, one must remember that in P_1 defined according to (43) the degree of vertex 1 is required to be 1, while in our histogram, either of the two vertices involved in a particular resistance distance value can have degree 1. For that reason, one should compare the histogram to $2P_1(\rho)$. With these specifications, we find convincing agreement between the two curves (remembering the rather low value of c and the asymptotic nature of our saddle point approach).

The result (53) is rather intuitive in the following sense. To reach the rest of the graph from a degree 1 vertex, the electric current must traverse the edge that connects this vertex to its unique nearest neighbor, which adds 1 ohm to the resistance distance, and then reach the target vertex starting from this unique nearest neighbor. It is natural to imagine that the resistance distance contributed by this second stage is distributed in the same way as the resistance distance between two generic vertices. This intuitive picture suggests that the distribution of resistance distances involving vertices of degree 1 should mimic the general distribution of resistance distances, but is shifted by 1 ohm, as in (53). Our analysis provides a mathematical foundation for this intuitive picture.

It is evident due to the suppression by e^{-c} that $P_1(\rho)$, which contributes to $P(\rho)$ according to (44), could not possibly be recovered by computing $1/c$ corrections to the saddle point estimate for $P(\rho)$ developed in section 3. It remains an open question whether subleading saddles exist in (13-14) that would make it possible to recover the ‘bump’ in $P(\rho)$ arising from $P_1(\rho)$ directly from (13-14), without resorting to the decomposition (44).

It should be possible to repeat our derivations for P_d with $d \geq 2$. The construction will still involve applying the Cauchy formula to (48) and then evaluating a Gaussian integral over ϕ_1 . However, the source term in the resulting formula (the term linear in ϕ appearing in the exponent) will no longer exactly match (13-14). So one can still attempt a saddle point analysis at large c as in section 3, but the technical details will change. Intuitively, this complication corresponds to the fact that, at higher d , the current can exit the source vertex toward any of its d nearest neighbors and then has to reach the target vertex from any of these neighbors, which is a more complicated picture than for $d = 1$. We believe it is very likely that an analytic representation can be developed for P_d with $d \geq 2$, but prefer to limit ourselves here to the relatively straightforward case $d = 1$ treated above.

6 Discussion

We have proposed an auxiliary field representation in the spirit of statistical field theory for the resistance distance distribution in large Erdős-Rényi graphs of fixed mean degree c .

Using this representation, a saddle point estimation of this distribution becomes possible at large c in terms of the saddle point equation (19), producing the analytic curve (30), which at progressively larger values of c can be further simplified to (34) or even to the Gaussian form (35). We have furthermore identified the subleading peaks observed at large resistance distances in numerical simulations with contributions of vertices of low degrees, and developed an analytic estimate (53) for the rightmost prominent peak of this sort, coming from vertices of degree 1.

A few asymptotic approximations have been employed in our analysis, and it may be wise to summarize them here. First, starting from (12), we only retained the leading term in c/N , thus focusing on the sparse graph regime $c \ll N$. We have furthermore resorted to an $1/c$ expansion in our analysis of (13-14), making our derivations geared toward the asymptotic regime $1 \ll c \ll N$, where they are validated by Fig. 2. It is a practical observation of Fig. 3 that our asymptotic analysis qualitatively captures the main peak of the distribution even for smaller values of c , strictly speaking, outside the intended validity domain of our approximations.

Our derivations have been heuristic in nature, in the sense that our plausible (but not rigorously justified) assumptions about the saddle point that dominates the quantity of interest have produced results that are convincing in terms of how they compare to the corresponding numerical simulations, even for low values of c such as 6 or 4. One may hope that further mathematical work will provide error bounds for our approximate results, though this would likely require methods beyond what we have used to derive the formulas presented here. Even at a more basic level, within the context of the effective field-theoretic description (13-14) or (40-41), one could hope to obtain a more thorough picture of the full set of saddle points and the explicit integration contour deformation that leads to our estimate, in particular in relation to the multiple branches of the Lambert W function in (20). Our derivation has simply assumed that the relevant saddle point respects the symmetries of the integrand, and that one must use the main branch [68] of the Lambert W function, and the test of this assumption has been in successful comparisons of its outcome with numerical experiments.

Another relevant subtlety is that we have essentially treated the Erdős-Rényi graphs as if they were connected, which gives a negligible mistake in an $1/c$ treatment at large c since the fraction of vertices outside the giant connected component is suppressed as e^{-c} [59]. However, the issue becomes more and more pressing as one decreases c . In particular, if one is interested in the critical ($c \approx 1$) or subcritical ($c < 1$) regime of the Erdős-Rényi graphs (as in related considerations of [44] for shortest path distances), it is certain that the issue can no longer be ignored. It may still be possible to develop an auxiliary field representation along the lines of section 2, but a technical obstruction is that a Gaussian representation is needed for the pseudoinverse of a general matrix, without any prior assumptions about its null eigenvectors.

It must be possible, with relative ease, to adapt our considerations to more sophisticated random graph models that can themselves be effectively treated by statistical field theory methods. One natural starting point is the two-star model [54, 56] and its relatives [57]. Note that, for the Erdős-Rényi case, the averaging over the graphs is very simple and could be performed directly, but we had to employ a statistical field representation for the resistance distance δ -function necessary to obtain the corresponding probability distribution. For more

complicated random graph ensembles, statistical field theory methods may be necessary to effectively handle the average over the graphs as well.

We conclude with some tentative comments on the intriguing small-scale features seen in the numerical simulations, such as the sharp narrow sub-peaks visible in Fig. 4a. These features are not statistical noise (as one might imagine at the first sight) and neither go away with averaging over a large number of runs nor change their positions between different runs. Our analytic results say nothing about these sharp sub-peaks, though they do a great job at capturing the shape of the large-scale peak upon which these small features are superposed. One natural thought is that, since resistance distance values $1/d_i + 1/d_j$ in terms of the vertex degrees d_i and d_j play a significant role in many considerations of resistance distances [10, 11], there could be sharp peaks in the distribution around the rational values $1/n + 1/m$ coming from vertices of degrees n and m . We have not, however, been able to straightforwardly trace the locations of sharp sub-peaks in Fig. 4a back to such simple rational numbers. Be it as it may, the ornate structure of Fig. 4a, which is also replicated in the degree-differentiated resistance distance distributions presented in color in the same figure, makes it very tempting to conjecture that the true resistance distance distribution of an infinitely large Erdős-Rényi graph is not, in fact, a smooth curve. Refuting or supporting this conjecture would evidently require methodology considerably beyond the scope of this article.

Acknowledgments

We thank Eytan Katzav for inciting our interest in properties of resistance distances, and for comments on the manuscript. Research of O.E. has been supported by the CUniverse project (CUAASC) at Chulalongkorn University. Research of T.C. has been supported by the Program Management Unit for Human Resources and Institutional Development, Research and Innovation (grant number B05F630108), and by Thailand Science Research and Innovation Fund Chulalongkorn University [CU-FRB65_ind (5)_110_23_40].

A Pedagogical account of the resistance distance

Consider a graph defined by an adjacency matrix $\mathbf{A}$. Imagine that each edge is made of a conducting wire with resistance 1 ohm, and a current of 1 amp is injected into vertex i and ejected from vertex j . This process induces voltage readings u_k as measured at vertex k . Since 1 amp of current flows between vertices i and j , the resistance measured between these two vertices, which is by definition the resistance distance, is

$$\Omega_{ij} = u_j - u_i. \quad (54)$$

Our purpose in this appendix is to derive the expression (4) for this quantity.

The total current flowing out of vertex k is 1 if $k = i$, -1 if $k = j$ and 0 otherwise. Let l be a nearest neighbor of k (so that $A_{kl} = 1$). Since the edge between k and l is a wire of resistance 1 ohm, this edge is traversed by the current $u_l - u_k$. Summing this expression over all the edges originating from vertex k , we get the total current exiting vertex k , that is

$$\sum_{l: A_{kl}=1} (u_l - u_k) \equiv \sum_{l=1}^N A_{kl}(u_l - u_k) = x_k, \quad (55)$$

where

$$x_k \equiv \begin{cases} 1 & \text{if } k = i, \\ -1 & \text{if } k = j, \\ 0 & \text{otherwise.} \end{cases} \quad (56)$$

Since $\sum_l A_{kl} = d_k$ by (1), we can rewrite this formula in matrix notation as

$$\mathbf{L}\mathbf{u} = -\mathbf{x}, \quad (57)$$

where $\mathbf{L}$ is the Laplacian (3), $\mathbf{u}$ is the vector with components u_k , and $\mathbf{x}$ is a vector with components x_k .

Assume first that the graph described by $\mathbf{A}$ is connected. In that case, $\mathbf{L}$ has one null eigenvector $(1, 1, \dots, 1)^T$. This vector represents a common shift of the voltages u_k by a constant, which evidently does not affect the physics. Because of this null vector, $\mathbf{L}$ is not invertible. We can, however, modify it so as to be invertible by introducing the matrix $\mathbf{1}$ all of whose entries equal 1, and replacing (57) by

$$(\mathbf{L} + \lambda\mathbf{1})\mathbf{u} = \mathbf{x}, \quad (58)$$

where λ is an arbitrary real number. Since $\mathbf{L}\mathbf{1} = \mathbf{1}\mathbf{L} = 0$ and $\mathbf{1}\mathbf{x} = 0$, the above equation is equivalent to (57) together with the extra condition $\mathbf{1}\mathbf{u} = 0$, which is the same as $\sum_k u_k = 0$, thus fixing the voltage shift ambiguity. Then, (58) is immediately solved by

$$\mathbf{u} = -(\mathbf{L} + \lambda\mathbf{1})^{-1}\mathbf{x}, \quad (59)$$

where $(\mathbf{L} + \lambda\mathbf{1})^{-1}$ is the ordinary matrix inverse of $\mathbf{L} + \lambda\mathbf{1}$. Equivalently, in components,

$$u_k = -(\mathbf{L} + \lambda\mathbf{1})_{ki}^{-1} + (\mathbf{L} + \lambda\mathbf{1})_{kj}^{-1}, \quad (60)$$

which is also a solution of (57), for any λ . Finally, from (54),

$$\Omega_{ij} = (\mathbf{L} + \lambda\mathbf{1})_{ii}^{-1} + (\mathbf{L} + \lambda\mathbf{1})_{jj}^{-1} - 2(\mathbf{L} + \lambda\mathbf{1})_{ij}^{-1}. \quad (61)$$

As emphasized already, the inverses are defined for any nonzero λ and the above expression cannot depend on λ due to the relation $\mathbf{L}\mathbf{1} = \mathbf{1}\mathbf{L} = 0$. In particular, taking λ to ∞ converts $(\mathbf{L} + \lambda\mathbf{1})^{-1}$ to the Moore-Penrose pseudoinverse of $\mathbf{L}$, yielding (4) as desired. Note that, for practical computations, (61) with an arbitrarily specified λ is often more convenient [8] than the Moore-Penrose pseudoinverse.

If the graph corresponding to $\mathbf{A}$ is disconnected, it is a matter of convention which value of resistance distance to assign to two vertices that belong to two different connected components, since no current may flow between two such vertices. Furthermore, $\mathbf{A}$ and $\mathbf{L}$ are block-diagonal with different blocks corresponding to different connected components. In this situation, there are more than one null vector (one null vector for each connected component) and the ordinary matrix inverses used in (61) no longer exist. Nonetheless, the definition (4) based on the Moore-Penrose pseudoinverse still works, and it assigns resistance distance zero to any two vertices belonging to two different connected components, while resistance distances within the same connected component are the same as what one would get with (61) by treating this particular component as an isolated connected graph. Thus, (4) provides a convenient prescription for defining resistance distances in disconnected graphs as well, which is commonly adopted in the literature.

B Gaussian integration formulas for matrix inversion

We start with the well-known Gaussian integration formula

$$e^{i \sum_{kl} M_{kl}^{-1} x_k x_l} = \left(\frac{i}{\pi} \right)^{N/2} \sqrt{\det \mathbf{M}} \int d\boldsymbol{\phi} e^{-i \sum_{kl} M_{kl} \phi_k \phi_l + 2i \sum_k \phi_k x_k}, \quad (62)$$

and would like to use it to justify (8). An easy way to verify (62) is by introducing the orthonormal eigenbasis of $\mathbf{M}$ denoted $\mathbf{e}_k$ and the corresponding eigenvalues μ_k . After changing the integration variables from $\boldsymbol{\phi}$ to $y_k \equiv (\boldsymbol{\phi}, \mathbf{e}_k)$ (an orthogonal change of variables with a unit Jacobian), the N integrations decouple into one-dimensional Gaussian (more precisely, Fresnel) integrals, and then (62) is recovered due to $\det \mathbf{M} = \prod_k \mu_k$ and $M_{kl}^{-1} = \sum_n (e_n)_k (e_n)_l / \mu_n$.

We cannot apply (62) for cases when $\mathbf{M}$ is proportional to the graph Laplacian $\mathbf{L}$ defined by (3), since in that case $\mathbf{M}$ possesses a null eigenvector $\mathbf{e}_1 = (1, 1, \dots, 1)^T$ with $\mu_1 = 0$. As a result, neither is the ordinary matrix inverse well-defined, nor is the Gaussian integral on the right-hand side of (62) convergent. This complication can be remedied, however, by considering the following modified formula:

$$e^{i \sum_{kl} M_{kl}^{\text{inv}} x_k x_l} = \left(\frac{i}{\pi} \right)^{\frac{N-1}{2}} \sqrt{\det' \mathbf{M}} \int d\boldsymbol{\phi} \delta(\sum_i \phi_i) e^{-i \sum_{kl} M_{kl} \phi_k \phi_l + 2i \sum_k \phi_k x_k}, \quad (63)$$

where $\det' \mathbf{M} = \prod_{k=2}^N \mu_k$ is the pseudodeterminant and $M_{kl}^{\text{inv}} = \sum_{n=2}^N (e_n)_k (e_n)_l / \mu_n$ is the pseudoinverse. Indeed, changing to the eigenbasis $y_k \equiv (\boldsymbol{\phi}, \mathbf{e}_k)$ as before, one obtains $\int dy_1 \delta(y_1) e^{2iy_1(\mathbf{x}, \mathbf{e}_1)} = 1$ for the y_1 -integration, while the remaining integrations work in exactly the same way as in the derivation of (62).

Finally, substituting $\mathbf{M} = \xi \mathbf{L}$ and $\mathbf{x} = (\xi, -\xi, 0, \dots, 0)^T$ into (63), we arrive at (8).

C Subleading corrections in the $1/c$ expansion

We present here computations for the $1/c$ corrections to the leading order saddle point estimate (30). While this material is of only peripheral importance for the main claim of the paper that the leading saddle point estimate provides a good approximation to the empirical data, it may be useful for future work to have these subleading computations collected here.

To analyze the $1/c$ corrections, we introduce $\varphi_1 = \phi_1 - \phi_0$ and $\varphi_2 = \phi_2 - \phi_0$ and expand (41) around the saddle point configuration $\boldsymbol{\phi}_0 = (\phi_0, -\phi_0, 0, 0, \dots)$, $\boldsymbol{\chi} = \boldsymbol{\theta} = \boldsymbol{\eta} = 0$ up to quartic order in the deviations, with $\tilde{\boldsymbol{\phi}} \equiv \boldsymbol{\phi} - \boldsymbol{\phi}_0 = (\varphi_1, \varphi_2, \phi_3, \dots, \phi_N)$:

$$H' = H(\boldsymbol{\phi}_0) + \frac{1}{2} \sum_{ij} M_{ij} \tilde{\phi}_i \tilde{\phi}_j - \frac{c}{2N} \sum_{ij} [(\chi_i - \chi_j)^2 + (\theta_i - \theta_j)(\eta_i - \eta_j)] + \Delta H + \Delta H', \quad (64)$$

$$\begin{aligned} \Delta H = & -\frac{c\xi^2}{2N} \sum_{3 \leq i < j} (\phi_i - \phi_j)^4 \\ & + \frac{\xi^2}{N\phi_0} \sum_{j \geq 3} \left(-\frac{2}{3} \phi_0 (3 - 2i\xi\phi_0^2) (\varphi_1 - \phi_j)^3 + \frac{1}{6} (-3 + 12i\xi\phi_0^2 + 4\xi^2\phi_0^4) (\varphi_1 - \phi_j)^4 \right) \\ & + \frac{\xi^2}{N\phi_0} \sum_{j \geq 3} \left(\frac{2}{3} \phi_0 (3 - 2i\xi\phi_0^2) (\varphi_2 - \phi_j)^3 + \frac{1}{6} (-3 + 12i\xi\phi_0^2 + 4\xi^2\phi_0^4) (\varphi_2 - \phi_j)^4 \right), \end{aligned} \quad (65)$$

$$\begin{aligned}
\Delta H' = & \frac{i\xi}{N} \frac{1 - 2i\xi\phi_0^2}{\phi_0} \left\{ \sum_{j \geq 3} (\varphi_1 - \phi_j)^2 [(\chi_1 - \chi_j)^2 + (\theta_1 - \theta_j)(\eta_1 - \eta_j)] \right. \\
& \left. + \sum_{j \geq 3} (\varphi_2 - \phi_j)^2 [(\chi_2 - \chi_j)^2 + (\theta_2 - \theta_j)(\eta_2 - \eta_j)] \right\} + \\
& + \frac{i\xi c}{N} \sum_{3 \leq i < j} (\phi_i - \phi_j)^2 [(\chi_i - \chi_j)^2 + (\theta_i - \theta_j)(\eta_i - \eta_j)] + \frac{1}{2} \sum_{i < j} ((\chi_i - \chi_j)^4 - (\theta_i - \theta_j)^2 (\eta_i - \eta_j)^2).
\end{aligned} \tag{66}$$

Here, $H(\phi_0)$ is given by (23), M_{ij} is given by (25), and the saddle point equation (19) has been used to eliminate the terms linear in $\tilde{\phi}$ and to simplify the other terms. To organize the expressions above, we have separated the summation over k and l in (41) into four regions, which we demonstrate as follows for the case of (65):

- $k = 1, l = 2$ – this term can be discarded at $N \rightarrow \infty$ as it is explicitly suppressed by $1/N$; it is not included in (65),
- $k = 1, l \geq 3$ – these terms give rise to the second line in (65),
- $k = 2, l \geq 3$ – these terms give rise to the third line in (65),
- $k, l \geq 3$ – these terms give rise to the first line in (65).

A similar treatment has been applied in (66). One should keep in mind that $\mathbf{M}$ is of order c , and hence $\mathbf{M}^{-1}$ is of order $1/c$, and ϕ_0 is of order $1/c$ according to (22). In all the computations, two kind of terms should be systematically ignored: the terms that vanish at $N \rightarrow \infty$ and the terms that are ξ -independent (the latter can only contribute to the irrelevant normalization factor of $P(\rho)$ as will become apparent below). For example, the last sum in the last line of (66) can only produce ξ -independent contributions in the first order corrections.

In (65), one can see terms cubic in $\tilde{\phi}$. These terms, however, come with an extra suppression in $1/c$ and cannot possibly contribute to the leading order corrections. Indeed, one can convince oneself that all even order terms in (65) come with coefficients $O(c)$, while all odd order terms are $O(1)$; it should be kept in mind that ϕ_0 is $O(1/c)$. This suppression is easily understood because (65) arises from expanding a function even under reflections of ϕ around a small value of ϕ defined by ϕ_0 . As a result, contributions of the cubic terms are suppressed and irrelevant at leading order. The first nonvanishing correction comes from the square of the cubic terms in (65), and it can be seen to contribute at order $1/c^3$. There are similar cubic terms in (66), but they similarly cannot contribute and have already been omitted in (66) in order not to clutter the formulas.

With all of these preliminaries, the relevant corrections can be defined as $\mathcal{F}_{4,\phi}$ and $\mathcal{F}_{4,\chi\theta\eta}$ given by

$$\begin{aligned}
& \int d\phi d\chi d\theta d\eta \delta(\sum_i \phi_i) \delta(\sum_i \chi_i) \delta(\sum_i \theta_i) \delta(\sum_i \eta_i) \\
& \times e^{\frac{1}{2} \sum M_{ij} \tilde{\phi}_i \tilde{\phi}_j - \frac{c}{2N} \sum [(\chi_i - \chi_j)^2 + (\theta_i - \theta_j)(\eta_i - \eta_j)]} (1 + \Delta H + \Delta H') \propto \frac{(-2\pi)^{\frac{N-1}{2}}}{\sqrt{\det' \mathbf{M}}} (1 + \mathcal{F}_{4,\phi} + \mathcal{F}_{4,\chi\theta\eta}),
\end{aligned} \tag{67}$$

where $\mathcal{F}_{4,\phi}$ collects all the terms coming from ΔH , and $\mathcal{F}_{4,\chi\theta\eta}$, those from $\Delta H'$. We have ignored the determinant and numerical prefactor arising from the Gaussian integration over χ , θ and η , as they are ξ -independent and common to all contributions.

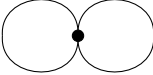
Another relevant point comes from what is known as the ‘linked cluster theorem’ in statistical field theory; for an exposition aimed at a broad audience see chapter 5 of [52]. The essence is that higher powers of the leading order contributions will arise at higher orders (due to disconnected Feynman diagrams consisting of multiple copies of lower-order diagrams). It is possible to resum all of these corrections in a compact form, which amounts to replacing $1 + \mathcal{F}_{4,\phi} + \mathcal{F}_{4,\chi\theta\eta}$ in (67) by $e^{\mathcal{F}_{4,\phi} + \mathcal{F}_{4,\chi\theta\eta}}$. Thus, once $\mathcal{F}_{4,\phi}$ and $\mathcal{F}_{4,\chi\theta\eta}$ have been computed, a wise way to incorporate them into the estimation of (40-41) is by writing

$$P(\rho) \propto \int d\xi \left(\frac{1}{\phi_0} - 2i\xi\phi_0 \right)^{-1} e^{-i\xi\rho} e^{\frac{2}{\phi_0}(1-\frac{1}{2c}) + 4i\xi\phi_0(1+\frac{1}{2c}) + \mathcal{F}_{4,\phi} + \mathcal{F}_{4,\chi\theta\eta}}. \quad (68)$$

This formula provides the $1/c$ -corrected version of the resistance distance distribution. The rest of this appendix simply reports the computation of $\mathcal{F}_{4,\phi}$ and $\mathcal{F}_{4,\chi\theta\eta}$.

Computation of $\mathcal{F}_{4,\phi}$

We now proceed with evaluating $\mathcal{F}_{4,\phi}$, whose Feynman diagram representation is


(69)

We emphasize again that, because of the structure of (68), we can safely ignore any ξ -independent terms in $\mathcal{F}_{4,\phi}$, as they can be merged into the normalization of (13). The quartic terms coming from (65) can be further simplified for the purposes of computing (67). It is important to keep in mind that the Gaussian measure in (67) is invariant under interchange of φ_1 and φ_2 , and under arbitrary permutations of $\phi_3, \dots, \phi_N$. Hence, integrating $(\varphi_1 - \phi_j)^4$ or $(\varphi_2 - \phi_j)^4$ with any $j \geq 3$ gives the same result as integrating $(\varphi_1 - \varphi_3)$, while integrating $(\phi_i - \phi_j)^4$ with $3 \leq i < j$ gives the same result as integrating $(\phi_3 - \phi_4)^4$. Thus, all the sums in (65) can be eliminated at the cost of appending explicit N -dependent factors that count the number of terms, since all the terms in each sum equal each other upon the integration in (67). Once this simplification has been implemented, one can further discard any contributions that vanish at $N \rightarrow \infty$. This leads to the following equivalent replacement for $(\Delta H)_4$, the quartic part of ΔH :

$$(\Delta H)_4 \rightarrow \frac{\xi^2}{3\phi_0} (4\xi^2\phi_0^4 + 12i\xi\phi_0^2 - 3)(\varphi_1 - \varphi_3)^4 - \frac{(N-5)}{4} c \xi^2 (\phi_3 - \phi_4)^4. \quad (70)$$

In general, expanding $(\phi_i - \phi_j)^4 = \phi_i^4 - 4\phi_i^3\phi_j + 6\phi_i^2\phi_j^2 - 4\phi_i\phi_j^3 + \phi_j^4$ and applying the standard Gaussian integration leads to

$$\begin{aligned} \int d\phi \left(\sum_i \phi_i \right) e^{-\frac{1}{2}(-M_{ij})\phi_i\phi_j} (\phi_i^4 - 4\phi_i^3\phi_j + 6\phi_i^2\phi_j^2 - 4\phi_i\phi_j^3 + \phi_j^4) \\ = \frac{(-2\pi)^{\frac{N-1}{2}}}{\sqrt{\det' \mathbf{M}}} [3(M_{ii}^{-1})^2 - 4M_{ii}^{-1}M_{ij}^{-1} + 6M_{ii}^{-1}M_{jj}^{-1} - 4M_{ij}^{-1}M_{jj}^{-1} + 3(M_{jj}^{-1})^2], \end{aligned} \quad (71)$$

where $\mathbf{M}^{-1}$ evidently means the inverse of $\mathbf{M}$ in the subspace orthogonal to the null vector $(1, 1, 1, \dots)^T$ such that $\mathbf{M}^{-1}\mathbf{M} = \mathbf{I} - \mathbf{1}/N$. Evaluating $\mathcal{F}_{4,\phi}$ amounts to applying this integration formula to (70), and subsequently omitting all terms that are either ξ -independent or vanish at $N \rightarrow \infty$. While integrating the last term in (70), it is useful to remember that $M_{33}^{-1} = M_{44}^{-1}$ while M_{34}^{-1} is of order $1/N$ and furthermore its leading $1/N$ part is ξ -independent. As a result, one obtains

$$\mathcal{F}_{4,\phi} = \frac{\xi^2}{\phi_0} (4\xi^2\phi_0^4 + 12i\xi\phi_0^2 - 3)(M_{11}^{-1} + M_{33}^{-1})^2 + 3(N-5)c\xi^2(M_{33}^{-1})^2. \quad (72)$$

Given the complete description of the eigenvectors and eigenvalues of $\mathbf{M}$ under (27), one can straightforwardly construct $\mathbf{M}^{-1}$ as

$$M_{ij}^{-1} = \frac{1}{a-g} e_i^1 e_j^1 - \frac{1}{Nf} e_i^2 e_j^2 + \frac{1}{b-d} (\delta_{ij} - e_i^0 e_j^0 - e_i^1 e_j^1 - e_i^2 e_j^2) \quad (73)$$

with $\mathbf{e}^0 = (1, 1, 1, \dots)^T / \sqrt{N}$, $\mathbf{e}^1 = (1, -1, 0, 0, \dots)^T / \sqrt{2}$, $\mathbf{e}^2 = \sqrt{\frac{N-2}{2N}} (1, 1, -\frac{2}{N-2}, -\frac{2}{N-2}, \dots)^T$. From this expression, one directly recovers

$$M_{11}^{-1} = \frac{1}{a-g} + O(1/N), \quad M_{33}^{-1} = \frac{1-1/N}{b-d} + O(1/N^2). \quad (74)$$

Furthermore, $a = a_0 + O(1/N)$ and $b = b_0 + b_1/N + O(1/N^2)$ with

$$a_0 = -4\xi^2\phi_0 - \frac{2i\xi}{\phi_0}, \quad b_0 = -2i\xi c, \quad b_1 = -8\xi^2\phi_0 - \frac{4i\xi}{\phi_0}, \quad (75)$$

while d and g are $O(1/N)$. We can keep only the contributions of order 1 in the first term of (72), meaning that $M_{11}^{-1} \approx 1/a_0$ and $M_{33}^{-1} \approx 1/b_0$ for the purposes of evaluating this term, while the last term contains an explicit factor of N and we must be careful to retain the contributions of order $1/N$. Note that the contribution of order N in the last term of (72) is ξ -independent and can thus be ignored in the context of (68), while the remaining subleading contribution is of order 1 and ξ -dependent. Explicitly,

$$(M_{33}^{-1})^2 = \frac{1}{b_0^2} - \frac{2}{Nb_0^2} - \frac{2}{N} \frac{b_1 - dN}{b_0^3} + O(1/N^2). \quad (76)$$

Putting everything together and discarding all terms that vanish at $N \rightarrow \infty$ or are ξ -independent results in the following evaluation:

$$\mathcal{F}_{4,\phi} \rightarrow \frac{\xi^2}{\phi_0} (4\xi^2\phi_0^4 + 12i\xi\phi_0^2 - 3) \left(\frac{1}{a_0} + \frac{1}{b_0} \right)^2 + 3c\xi^2 \left(-\frac{2b_1}{b_0^3} \right), \quad (77)$$

or explicitly:

$$\mathcal{F}_{4,\phi} = -\frac{1}{\phi_0 c^2} \left\{ 3 - 6i\xi\phi_0^2 - \frac{1}{4} (3 - 12i\xi\phi_0^2 - 4\xi^2\phi_0^4) \left(1 + \frac{c\phi_0}{1 - 2i\xi\phi_0^2} \right)^2 \right\}. \quad (78)$$

Computation of $\mathcal{F}_{4,\chi\theta\eta}$

As explained above, only terms in $\Delta H'$ quadratic in $\tilde{\phi}$ and either quadratic in χ or bilinear in θ and η contribute nontrivially to $\mathcal{F}_{4,\chi\theta\eta}$ at order $1/c$. In terms of Feynman diagrams, these terms are visualized as


(79)

where the dashed line denotes the χ -propagator, and the line with arrows, the $\theta\eta$ -propagator.

In a manner similar to the treatment of $\mathcal{F}_{4,\phi}$, permutation symmetries of the Gaussian measure and the $N \rightarrow \infty$ limit let one simplify $\Delta H'$ to obtain

$$\begin{aligned} \Delta H' \rightarrow & \frac{2i\xi}{\phi_0} (1 - 2i\xi\phi_0^2)(\varphi_1 - \phi_3)^2 [(\chi_1 - \chi_3)^2 + (\theta_1 - \theta_3)(\eta_1 - \eta_3)] \\ & + i\xi c(N - 5)(\phi_3 - \phi_4)^2 [(\chi_3 - \chi_4)^2 + (\theta_3 - \theta_4)(\eta_3 - \eta_4)]. \end{aligned} \quad (80)$$

Furthermore, the integration measure of χ , θ and η is fully symmetric under all permutations and decoupled from ϕ . Hence, another equivalent replacement is

$$\Delta H' \rightarrow \left(\frac{2i\xi}{\phi_0} (1 - 2i\xi\phi_0^2)(\varphi_1 - \phi_3)^2 + i\xi c(N - 5)(\phi_3 - \phi_4)^2 \right) [(\chi_1 - \chi_2)^2 + (\theta_1 - \theta_2)(\eta_1 - \eta_2)]. \quad (81)$$

Since the Gaussian measure in (68) factorizes, the content of the square brackets in the above expression can be treated separately. We first write

$$\frac{\int d\chi \delta(\sum_i \chi_i) e^{-\frac{1}{2} \sum_{ij} \Pi_{ij} \chi_i \chi_j} (\chi_1 - \chi_2)^2}{\int d\chi \delta(\sum_i \chi_i) e^{-\frac{1}{2} \sum_{ij} \Pi_{ij} \chi_i \chi_j}} = \Pi_{11}^{-1} + \Pi_{22}^{-1} + O(1/N), \quad (82)$$

where $\Pi \equiv 2c(\mathbf{I} - \mathbf{1}/N)$ is read off (64). Similarly,

$$\frac{\int d\theta d\eta \delta(\sum_i \theta_i) \delta(\sum_i \eta_i) e^{-\frac{1}{2} \sum_{ij} \Pi_{ij} \theta_i \eta_j} (\theta_1 - \theta_2)(\eta_1 - \eta_2)}{\int d\theta d\eta \delta(\sum_i \theta_i) \delta(\sum_i \eta_i) e^{-\frac{1}{2} \sum_{ij} \Pi_{ij} \theta_i \eta_j}} = -2 (\Pi_{11}^{-1} + \Pi_{22}^{-1} + O(1/N)). \quad (83)$$

Then, computing the Gaussian average over $\tilde{\phi}$ in a manner directly analogous to the evaluation of $\mathcal{F}_{4,\phi}$, we get:

$$\mathcal{F}_{4,\xi\theta\eta} = -(\Pi_{11}^{-1} + \Pi_{22}^{-1}) \left(\frac{2i\xi}{\phi_0} (1 - 2i\xi\phi_0^2) (M_{11}^{-1} + M_{33}^{-1}) + iNc\xi (M_{33}^{-1} + M_{44}^{-1}) \right). \quad (84)$$

We then recall (74-75) and approximate $M_{11}^{-1} \approx 1/a_0$ and $M_{33}^{-1} \approx 1/b_0$ while being careful about contributions of order $1/N$ in the term that explicitly involves N . We thus arrive at

$$\mathcal{F}_{4,\xi\theta\eta} = -\frac{1}{c} \left(\frac{2i\xi}{\phi_0} (1 - 2i\xi\phi_0^2) \left(\frac{1}{a_0} + \frac{1}{b_0} \right) + ic\xi \left(\frac{-2b_1}{b_0^2} \right) \right), \quad (85)$$

which can be simplified to

$$\mathcal{F}_{4,\xi\theta\eta} = \frac{2i\phi_0^2\xi + c\phi_0 - 1}{c^2\phi_0}. \quad (86)$$

References

- [1] F. Harary, *Graph theory* (CRC Press, 1994).
- [2] P. Miasnikof, A. Y. Shestopaloff, L. Pitsoulis, A. Ponomarenko and Yu. Lawryshyn, *Distances on a graph*, in *Complex Networks & Their Applications IX* (Springer, 2021).
- [3] J. Bouttier, P. Di Francesco and E. Guitter, *Geodesic distance in planar graphs*, Nucl. Phys. B **663** (2003) 535 [arXiv:cond-mat/0303272](#) [cond-mat.stat-mech].
- [4] P. G. Doyle and J. L. Snell, *Random walks and electric networks* (Mathematical Association of America, 1984) [arXiv:math/0001057](#) [math.PR].
- [5] D. J. Klein and M. Randić, *Resistance distance*, J. Math. Chem. **12**(1993) 81.
- [6] D. J. Klein, *Resistance-distance sum rules*, Croat. Chem. Acta **75** (2002) 633.
- [7] W. Xiao and I. Gutman, *Resistance distance and Laplacian spectrum*, Theor. Chem. Acc. **110** (2003) 284.
- [8] E. W. Weisstein, *Resistance distance* at MathWorld
<https://mathworld.wolfram.com/ResistanceDistance.html>
- [9] P. Tetali, *Random walks and the effective resistance of networks*, J. Th. Prob. **4** (1991) 101.
- [10] L. Lovász, *Random walks on graphs: a survey*, in *Combinatorics: Paul Erdős is Eighty* v. 2 (1993) <https://web.cs.elte.hu/~lovasz/erdos.pdf>.
- [11] U. von Luxburg, A. Radl and M. Hein, *Getting lost in space: large sample analysis of the resistance distance*, Adv. Neur. Inf. Proc. Sys. **23** (2010) 2622; *Hitting and commute times in large random neighborhood graphs*, J. Mach. Learn. Res. **15** (2014) 1751.
- [12] F. Y. Wu, *Theory of resistor networks: the two-point resistance*, J. Phys. A **37** (2004) 6653 [arXiv:math-ph/0402038](#).
- [13] P. Van Mieghem, K. Devriendt and H. Cetinay, *Pseudoinverse of the Laplacian and best spreader node in a network*, Phys. Rev. E **96** (2017) 032311.
- [14] N. Trinajstić, D. Babić, S. Nikolić, D. Plavšić, D. Amić and Z. Mihalić, *The Laplacian matrix in chemistry*, J. Chem. Inf. Comp. Sci. **34** (1994) 368.
- [15] O. Ivanciuc, *QSAR and QSPR molecular descriptors computed from the resistance distance and electrical conductance matrices*, ACH Model. Chem. **137** (2000) 607.
- [16] D. Babić, D. J. Klein, I. Lukovits, S. Nikolić and N. Trinajstić, *Resistance-distance matrix: a computational algorithm and its application*, Int. J. Quant. Chem. **90** (2001) 166.
- [17] P. W. Fowler, *Resistance distances in fullerene graphs*, Croat. Chem. Acta, **75** (2002) 401.

- [18] K. Roy, *Topological descriptors in drug design and modeling studies*, Molec. Diversity **8** (2004) 321.
- [19] G. Guillot, R. Leblois, A. Coulon and A. C. Frantz, *Statistical methods in spatial genetics*, Molec. Ecology **18** (2009) 4734.
- [20] Y. Yang and D. J. Klein, *Resistance distance-based graph invariants of subdivisions and triangulations of graphs*, Discr. App. Math. **181** (2015) 260.
- [21] J. Du, G. Su, J. Tu and I. Gutman, *The degree resistance distance of cacti*, Discr. App. Math. **188** (2015) 16.
- [22] J.-B. Liu, W.-R. Wang, Y.-M. Zhang and X.-F. Pan, *On degree resistance distance of cacti*, Discr. App. Math. **203** (2016) 217.
- [23] D. Liben-Nowell and J. Kleinberg, *The link prediction problem for social networks*, in CIKM'2003 Proceedings.
- [24] D. Zhou and B. Schölkopf, *Learning from labeled and unlabeled data using random walks*, in DAGM'2004 Proceedings.
- [25] M. Saerens, F. Fouss, L. Yen and P. Dupont *The principal components analysis of a graph, and its relationships to spectral clustering*, in ECML'2004 Proceedings.
- [26] L. Yen, D. Vanvyve, F. Wouters, F. Fouss, M. Verleysen and M. Saerens, *Clustering using a random walk based distance measure*, in ESANN'2005 Proceedings.
- [27] M. Herbster and M. Pontil, *Prediction on a graph with a perceptron*, in NIPS'2006 Proceedings.
- [28] C. Avin and G. Ercal, *On the cover time and mixing time of random geometric graphs*, Theor. Comput. Sci, **380** (2007) 2.
- [29] D. Spielman and N. Srivastava, *Graph sparsification by effective resistances*, in STOC'2008 Proceedings.
- [30] L. Yen, M. Saerens, A. Mantrach and M. Shimbo, *A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances*, in SIGKDD'2008 Proceedings.
- [31] N. Cesa-Bianchi, C. Gentile and F. Vitale, *Fast and optimal prediction on a labeled tree*, in COLT'2009 Proceedings.
- [32] D. M. Wittmann, D. Schmidl, F. Blöchl and F. J. Theis, *Reconstruction of graphs based on random walks*, Theor. Comp. Sci. **410** (2009) 3826.
- [33] C. Cooper and A. Frieze, *The cover time of random geometric graphs*, Rand. Struct. Alg. **38** (2011) 324.
- [34] P. Wills and F. G. Meyer, *Metrics for graph comparison: a practitioner's guide*, PLoS ONE 15 (2020) e0228728 [arXiv:1904.07414](https://arxiv.org/abs/1904.07414) [stat.AP].

- [35] A. Ponomarenko, L. Pitsoulis and M. Shamsheudinov, *Overlapping community detection in networks based on link partitioning and partitioning around medoids*, [arXiv:1907.08731](#) [cs.SI].
- [36] E. Katifori, G. J. Szöllösi and M. O. Magnasco, *Damage and fluctuations induce loops in optimal transport networks*, Phys. Rev. Lett. **104** (2010) 048704 [arXiv:0906.0006](#) [physics.bio-ph]
- [37] T. Gavrilenko and E. Katifori, *Distribution networks achieve uniform perfusion through geometric self-organization*, [arXiv:2009.04375](#) [physics.bio-ph].
- [38] S. Fancher and E. Katifori, *Tradeoffs between energy efficiency and mechanical response in fluid flow networks*, [arXiv:2102.13197](#) [physics.bio-ph].
- [39] Y. Luo, Ch.-L. Ho, B. R. Helliker and E. Katifori, *Leaf water storage and robustness to intermittent drought: a spatially explicit capacitive model for leaf hydraulics*, [arXiv:2106.08939](#) [physics.bio-ph].
- [40] T. Łuczak, *Random trees and random graphs*, Rand. Struct. Alg. **13** (1998) 485.
- [41] A. K. Hartmann and M. Mézard, *Distribution of diameters for Erdős-Rényi random graphs*, Phys. Rev. E **97** (2018) 032128 [arXiv:1710.05680](#) [cond-mat.dis-nn].
- [42] E. Katzav, M. Nitzan, D. ben-Avraham, P. L. Krapivsky, R. Kühn, N. Ross and O. Biham, *Analytical results for the distribution of shortest path lengths in random networks*, EPL **111** (2015) 26006 [arXiv:1504.00754](#) [cond-mat.dis-nn].
- [43] M. Nitzan, E. Katzav, R. Kühn and O. Biham, *Distance distribution in configuration model networks*, Phys. Rev. E **93** (2016) 062309 [arXiv:1603.04473](#) [cond-mat.dis-nn].
- [44] E. Katzav, O. Biham and A. K. Hartmann, *The distribution of shortest path lengths in subcritical Erdős-Rényi networks*, Phys. Rev. E **98** (2018) 012301 [arXiv:1806.05743](#) [cond-mat.dis-nn].
- [45] A. D. Jackson and S. P. Patil, *Phases of small worlds: a mean field formulation*, [arXiv:2103.14001](#) [cond-mat.stat-mech].
- [46] N. Boumal and X. Cheng, *Concentration of the Kirchhoff index for Erdős-Rényi graphs*, Sys. Contr. Lett. **74** (2014) 74 [arXiv:1307.6398](#) [cs.IT].
- [47] J. Sylvester, *Random walk hitting times and effective resistance in sparsely connected Erdős-Rényi random graphs*, J. Graph Th. **96** (2021) 44 [arXiv:1612.00731](#) [math.CO].
- [48] M. Alamgir and U. von Luxburg, *Phase transition in the family of p -resistances*, in *Proceedings of the 24th International Conference on Neural Information Processing Systems* (2011) p. 379.
- [49] I. Benjamini and R. Rossignol, *Submean variance bound for effective resistance of random electric networks*, Comm. Math. Phys. **280** (2008) 445 [arXiv:math/0610393](#) [math.PR].

- [50] A. Gosh, S. Boyd and A. Saberi, *Minimizing effective resistance of a graph*, SIAM Rev. **50** (2008) 37.
- [51] E. Brézin, *Introduction to statistical field theory* (Cambridge, 2010).
- [52] M. Helias and D. Dahmen, *Statistical field theory for neural networks* (Springer, 2020).
- [53] M. Moshe and J. Zinn-Justin, *Quantum field theory in the large N limit: a review*, Phys. Rept. **385** (2003) 69 [arXiv:hep-th/0306133](#).
- [54] J. Park and M. E. J. Newman, *Solution of the two-star model of a network*, Phys. Rev. E **70** (2004) 066146 [arXiv:cond-mat/0405457](#).
- [55] F. L. Metz, G. Parisi and L. Leuzzi, *Finite size correction to the spectrum of regular random graphs: an analytical solution*, Phys. Rev. E **90** (2014) 052109 [arXiv:1403.2582](#) [cond-mat.dis-nn].
- [56] A. Annibale and O. T. Courtney, *The two-star model: exact solution in the sparse regime and condensation transition*, J. Phys. A **48** (2015) 365001 [arXiv:1504.06458](#) [cond-mat.dis-nn].
- [57] M. Bolfe, F. L. Metz, E. Guzmán-González and I. Pérez Castillo, *Analytic solution of the two-star model with correlated degrees*, Phys. Rev. E **104** (2021) 014147 [arXiv:2102.09629](#) [cond-mat.stat-mech].
- [58] M. Bauer and O. Golinelli, *Random incidence matrices: moments of the spectral density*, J. Stat. Phys. **103** (2001) 301 [arXiv:cond-mat/0007127](#).
- [59] M. Newman, *Networks: an introduction* (Oxford, 2010).
- [60] A. D. Mirlin and Y. V. Fyodorov, *Universality of level correlation function of sparse random matrices*, J. Phys. A **24** (1991) 2273.
- [61] M. Mézard, G. Parisi and A. Zee, *Spectra of Euclidean random matrices*, Nucl. Phys. B **559** (1999) 689 [arXiv:cond-mat/9906135](#).
- [62] Y. Krumbeck, Q. Yang, G. W. A. Constable and T. Rogers, *Fluctuation spectra of large random dynamical systems reveal hidden structure in ecological networks*, Nature Comm. **12** (2021) 3625 [arXiv:2011.05140](#) [q-bio.PE].
- [63] R. Lyons, *Asymptotic enumeration of spanning trees*, Comb. Prob. Comp. **14** (2005) 491 [arXiv:math/0212165](#) [math.CO].
- [64] R. Lyons, R. Peled and O. Schramm, *Growth of the number of spanning trees of the Erdős-Rényi giant component*, Comb. Prob. Comp. **17** (2008) 711 [arXiv:0711.1893](#) [math.PR].
- [65] K. Efetov, *Supersymmetry in disorder and chaos* (Cambridge, 1996).

- [66] Y. V. Fyodorov, *Complexity of random energy landscapes, glass transition and absolute value of spectral determinant of random matrices*, Phys. Rev. Lett. **92** (2004) 240601 [arXiv:cond-mat/0401287](#) [cond-mat.dis-nn].
- [67] Y. V. Fyodorov and A. Nock, *On random matrix averages involving half-integer powers of GOE characteristic polynomials*, J. Stat. Phys. **159** (2015) 731 [arXiv:1410.5645](#) [math-ph].
- [68] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth, *On the Lambert W function*, Adv. Comp. Math. **5** (1996) 329.
- [69] R. Wong, *Asymptotic approximations of integrals* (SIAM, 2001), section VII.4.
- [70] H. F. Inman and E. L. Bradley, *The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities*, Comm. Stat. Th. Meth. **18** (1989) 385.
- [71] R. Kuehn and T. Rogers, *Heterogeneous micro-structure of percolation in sparse networks*, Eur. Phys. Lett. **118** (2017) 68003 [arXiv:1703.06740](#) [cond-mat.stat-mech].